Universitat de les Illes Balears
Departament de Ciències Matemàtiques i Informàtica

New Segmentation and Edge Detection Methods Using Physics-Based Models of Image Formation

Alberto Ortiz Rodríguez

Thesis supervisor: Gabriel Oliver Codina

Dissertation submitted in partial fulfillment
of the requirements for the degree of

DOCTOR EN INFORMÁTICA

Palma de Mallorca, September 2005

Para María,
por tantos años de paciente espera.
Para Lucas,
aunque es mucho menos paciente.
Para Alex,
que acaba de llegar.

Agradecimientos

Aunque ciertas cosas parecen nunca llegar a su fin, de una forma u otra, todo concluye. Esto es (espero) el estado de esta tesis. Ahora que el último error ya ha sido corregido, es el momento de agradecer a todas aquellas personas que, ya sea de forma puntual o bien a lo largo de todo el periodo de realización, han contribuido a que esta tesis terminara. Espero no olvidar a nadie, pero en caso de que lo haga, ruego se acepten mis más sinceras disculpas.

Para empezar, me gustaría agradecer de forma especial:

A mi director de tesis, Dr. Gabriel Oliver, por su perpetua confianza en mi trabajo, sus esfuerzos por proporcionarme los mejores medios para la realización de esta tesis, y por siempre estar dispuesto a escucharme.

Julián Proenza, por su continuo apoyo y por la multitud de discusiones filosóficas acerca de qué debería ser una tesis. Al final, aquel "dicho" que cuelga de la pared en el despacho y que dice *"para trabajar aquí no necesitas estar loco, pero ayuda si lo estás"* ha resultado ser realmente cierto ...

Dra. Yolanda González, por su interés en los diferentes tópicos que esta tesis considera, y por sus útiles comentarios sobre uno de los métodos de estimación de los parametros de iluminación.

El resto de miembros del grupo de Sistemas, Robótica y Visión (Javier Antich, Toni Burguera, José Guerrero y Guillermo Rodríguez-Navas) por su continuo apoyo moral.

Tobías Costa, por la implementación de los algoritmos de Klinker et al. [129] y de Gevers [68], los cuales han sido utilizados en las comparaciones con algunos de los algoritmos de segmentación presentados en esta tesis.

Por otro lado, me gustaría expresar mi agradecimiento general a todos los miembros del Departamento de Matemáticas e Informática por su interés en mi trabajo y por animarme a alcanzar este punto. Además, querría agradecer particularmente a los diferentes equipos directivos que han gestionado el departamento durante el periodo de realización de esta tesis, por sus esfuerzos dedicados a proporcionar las mejores condiciones para la terminación de las tesis.

También me gustaría expresar mi más profundo agradecimiento a todos los miembros de mis dos familias, por hacer mi vida más fácil durante todo el tiempo que he dedicado a esta tesis, no solamente por sus constantes ofrecimientos de ayuda, sino también por ocuparse de tantas (demasiadas) tareas para permitirme disfrutar del mayor tiempo posible para trabajar. Además, estoy particularmente agradecido a Emiliano García, Antonio García y mi padre por su ayuda durante ciertos experimentos preliminares de esta tesis.

Finalmente, el más importante de todos, quiero expresar mi más profundo agradecimiento a las personas que más directamente han sufrido los efectos colaterales de la realización de esta tesis, mi esposa María y nuestros hijos Lucas y Alex. Espero ser capaz de devolveros todo el tiempo que os merecéis y que no he podido dedicaros durante todos estos años ...

Acknowledgements

Although some things seem to never come to an end, one way or another, everything reaches a conclusion. This is (I hope) the state of this thesis. Now that the last error has been corrected, it is time to thank all the people who, punctually or throughout the whole period, have contributed to successfully attain this achievement. I expect not to forget anybody, but in case I do, please accept my most sincere apologies.

For a start, I would like to thank in a special way:

My thesis supervisor, Dr. Gabriel Oliver, for his everlasting confidence on my work, his efforts for providing me with the best means for the realization of this thesis, and for always being willing to listen to me.

Julián Proenza, for his continuous encouragement and for so many philosophical discussions regarding what a thesis should be like. In the end, that saying of yours hanging from the wall in the office and reading *"you don't have to be crazy to work here, but if you are it helps"* has been certainly true ...

Dr. Yolanda González, for her interest in the different topics this thesis addresses and her useful comments on one of the lighting estimation methods.

The remaining members of the Systems, Robotics and Vision group (Javier Antich, Toni Burguera, José Guerrero and Guillermo Rodríguez-Navas) for their moral support.

Tobías Costa, for the implementations of the algorithms by Klinker et al. [129] and by Gevers [68], which have been used to compare with some of the algorithms proposed in this thesis.

On the other hand, I would like to express my general gratitude to all the members of the Department of Mathematics and Computer Science for their interest in my work and for encouraging me to reach this point. Furthermore, I would, particularly, like to thank the different boards of directors who have been in charge of the management of the department during the period of realization of this thesis, for their efforts devoted to always arranging the best possible conditions for the completion of PhD theses.

I would also like to express my most profound appreciation to all the members of my two families for making my life easier during all the time I have spent on this thesis, not only for their continuous offers to lend a helping hand, but also for taking care of (too) many tasks to allow me to have most of the time available for working. Besides, I am, in particular, fully indebted with Emiliano García, Antonio García and my father for their help during some preliminary experiments of this thesis.

Finally, last but not least, my deepest gratitude goes to the people who have more directly suffered the collateral effects of the realization of this thesis, my wife María and our sons Lucas and Alex. I hope to be able to give you back all the time you deserve and that I have not been able to devote to you during the past years ...

Contents

List of Figures **XV**

List of Tables **XIX**

List of Algorithms **XXI**

Symbols and Notation **XXIII**

Acronyms **XXV**

1 Introduction and Objectives **1**
1.1 The Computer Vision Problem . . . 1
1.2 Problem Statement and Objectives . . . 2
1.2.1 General discussion . . . 2
1.2.2 Objectives of this thesis . . . 5
1.3 Thesis Outline . . . 7

2 The Image Formation Process **9**
2.1 Radiometry . . . 9
2.1.1 Radiometrical concepts . . . 9
2.1.2 The Bidirectional Reflectance Distribution Function . . . 11
2.2 Interaction of Light with Matter . . . 11
2.2.1 Introductory concepts . . . 11
2.2.2 Interaction of light with metals . . . 12
2.2.3 Interaction of light with dielectrics . . . 13
2.3 Reflection Models . . . 14
2.3.1 The Reflectance Map . . . 14
2.3.2 Body reflection models . . . 14
2.3.3 Interface reflection models . . . 17
2.3.4 Hybrid models . . . 19
2.3.5 Inter-reflections . . . 21
2.4 Optical Image Sensors . . . 22
2.4.1 General image sensing . . . 22
2.4.2 CCD basics . . . 24
2.4.3 Colour image formation in CCD cameras . . . 26
2.4.4 CCD cameras operation and noise modelling . . . 28
2.4.5 Other aspects of CCD camera operation . . . 31

3 Review of Image Segmentation Approaches **35**
3.1 Brief Review on General Image Segmentation . . . 35
3.2 The Roots of Physics-Based Segmentation . . . 37
3.3 State of the Art in Physics-Based Segmentation . . . 38
3.3.1 Reflectance estimation . . . 39
3.3.2 Cluster analysis . . . 40
3.3.3 Continuity of the image surface . . . 45
3.3.4 Photometric invariants and specialized colour spaces . . . 46
3.4 Summary and Conclusions . . . 55

4 General Framework for Image Segmentation and Edge Detection **59**
4.1 Axis Notation . . . 59
4.2 Expanded Model of Image Formation (without noise) . . . 59
4.3 Use of a Physics-Based Model for Segmentation and Edge Detection . . . 62
4.4 Discussion about the Experimental Setup . . . 63
4.4.1 Camera configuration and input data . . . 63
4.4.2 Performance evaluation of segmentation and edge detection algorithms . . . 65
General discussion . . . 65
Image segmentation evaluation and the confusion matrix . . . 66
A discrepancy-based evaluation strategy for image segmentation . . . 68
Discussion over an illustrative example . . . 74
Edge detection evaluation . . . 77
4.4.3 Performance evaluation setup . . . 78

5 Estimation of Lighting Distribution Parameters **79**
5.1 Estimation from Single Band Images . . . 79
5.1.1 Estimation of surface normal vectors . . . 80
5.1.2 Estimation of the lighting direction . . . 81
5.1.3 Final considerations . . . 87
5.2 Estimation from Colour Images . . . 88
5.3 Error of Assuming Orthographic Projection . . . 88
5.4 Extension to Perspective Projection . . . 91
5.4.1 Projection of the sphere . . . 91
5.4.2 Estimation of surface normal vectors . . . 93
5.4.3 Estimation of the lighting direction . . . 96
5.4.4 Summary of the extended estimation method . . . 97
5.5 Previous Work . . . 97
5.5.1 Illuminant orientation determination . . . 97
5.5.2 Estimation of the illuminant chromaticity . . . 98
5.5.3 General comment on the use of a calibration object . . . 99
5.6 Experimental Results . . . 100
5.6.1 Example of lighting parameters estimation . . . 100
5.6.2 Orthographic projection *versus* perspective projection . . . 103
5.6.3 Robustness measures . . . 106
Sensitivity regarding the initial contour . . . 107
Robustness against Gaussian noise . . . 108
Robustness of s estimations . . . 109
Robustness of L_a and L_d estimations . . . 110
5.6.4 Comparison with other lighting direction estimation methods . . . 112
5.7 Conclusions . . . 115

6 Radiometric Calibration of CCD Cameras 119
6.1 Estimation of the Camera Noise Model Parameters . . . 119
6.1.1 Estimation of the PRNU and DCNU distribution parameters . . . 120
6.1.2 Estimation of camera gain and charge-independent non-spatial noise . . . 121
6.2 Applications of Radiometric Camera Calibration . . . 122
6.2.1 Incorporation of realistic noise into synthetic images . . . 122
6.2.2 Determination of uncertainties in the measurement of image irradiance . . . 123
6.3 Previous Work . . . 125
6.4 Experimental Results . . . 126
6.4.1 Estimation of noise distribution parameters . . . 126
6.4.2 Generation of noisy synthetic images . . . 129
6.4.3 Determination of irradiance measurement uncertainties . . . 129
6.4.4 An example on the use of intensity uncertainties . . . 131
6.5 Conclusions . . . 133

7 Image Segmentation by Scene Reconstruction 137
7.1 Model of Image Formation . . . 137
7.2 Overview of IS^{2}R . . . 139
7.2.1 Basic strategy . . . 139
7.2.2 Shape from X methods . . . 139
7.2.3 Shape from Shading . . . 140
7.2.4 Algorithmic description of IS^{2}R . . . 140
7.3 Initialization of Regions . . . 143
7.4 Region Shape Estimation . . . 145
7.5 Shape Extrapolation . . . 147
7.6 Consistency Test . . . 148
7.7 Discussion on Segmentation Anomalies . . . 150
7.8 Extension of the Model of Image Formation . . . 151
7.8.1 Segmentation of colour images . . . 151
7.8.2 Tolerance to shadows . . . 152
7.8.3 About other extensions . . . 152
7.9 Experimental Results . . . 153
7.9.1 A detailed example . . . 154
7.9.2 Segmentation results for noiseless synthetic images . . . 154
7.9.3 Experiments with real images . . . 158
7.9.4 Comparison with other segmentation algorithms . . . 170
7.10 Conclusions . . . 174

8 Analysis of Colour Channels Coupling and Edge Detection 181
8.1 General Properties of Uniformly Coloured Image Areas . . . 181
8.2 Particularization of the General Properties . . . 183
8.2.1 Case 1: non-glossy pixels/without ambient lighting . . . 183
8.2.2 Case 2: non-glossy pixels/with ambient lighting . . . 184
8.2.3 Case 3: glossy pixels/without ambient lighting . . . 186
8.2.4 Case 4: glossy pixels/with ambient lighting . . . 187
8.3 Usefulness of the Properties . . . 188
8.4 Cases which Fail to Satisfy the Properties . . . 189
8.5 Formulation of a Compatibility Relationship between Pixels . . . 193
8.6 Coverage of Reflectance Transitions . . . 195
8.7 Colour Channel Coupling-based Edge Detection (C^{3}E) . . . 198
8.7.1 Edge map computation . . . 198
8.7.2 Implementation issues . . . 198
8.7.3 Examples of edge maps and discussion . . . 200
8.8 Edge Thinning Strategy . . . 201

8.9 Experimental Results . . . 202
8.9.1 Experiments with synthetic images . . . 203
8.9.2 Experiments with real images . . . 205
8.10 Conclusions . . . 210

9 Image Segmentation by Analysis of Colour Channels Coupling 213
9.1 Outline of the Segmentation Algorithm . . . 213
9.2 Region Growing . . . 214
9.3 Region Merging . . . 215
9.3.1 Merging of non-glossy regions . . . 215
9.3.2 Merging of specularities . . . 219
General description . . . 219
Implementation details . . . 220
9.3.3 An example of merging . . . 223
9.4 Experimental Results . . . 224
9.4.1 Experiments with synthetic images . . . 224
9.4.2 Experiments with real images . . . 228
9.5 Conclusions . . . 235

10 Conclusions and Future Work 243
10.1 Conclusions . . . 243
10.1.1 Global conclusions . . . 244
10.1.2 Conclusions for chapter 4 . . . 245
10.1.3 Conclusions for chapter 5 . . . 245
10.1.4 Conclusions for chapter 6 . . . 246
10.1.5 Conclusions for chapter 7 . . . 246
10.1.6 Conclusions for chapter 8 . . . 247
10.1.7 Conclusions for chapter 9 . . . 247
10.2 Future Work . . . 248

Appendices

A Multi-line Least Squares Fitting 251
A.1 Common-Origin Multi-Line Least Squares Fitting . . . 251
A.2 Common-Slope Multi-Line Least Squares Fitting . . . 252

B Proofs of chapter 8 255

C Reflectance Transitions Covered by the $\mathcal{C}$ Compatibility Relation 265

D The Possibilistic C-means Algorithm (PCM) 267
D.1 General Formulation of PCM . . . 267
D.2 Meaning and Importance of PCM Parameters . . . 268
D.3 PCM as a Mode-Seeking Algorithm . . . 268

References 271

Index 283

List of Figures *

1.1 [C] Example of segmentation where the meaningful structures segmented coincide with scene objects (algorithm C^3S, chapter 9) . . . 3
1.2 [C] Intensity profiles of the *house* image . . . 4
1.3 [C] Intensity profiles of a general image . . . 4
1.4 [C] Segmentation results for the *house* image by the algorithm published in [29] . . . 5
1.5 [C] Segmentation results for images of curved objects by the algorithm published in [29] 5
1.6 [C] Segmentation results for the physics-based segmentation algorithm C^3S (chapter 9) 6

2.1 [M] Radiometrical terms and notation . . . 10
2.2 [M] Angles notation for the bidirectional reflectance distribution function (BRDF) . . . 11
2.3 [M] Body reflection *versus* interface reflection over a dielectric . . . 13
2.4 [M] Examples of reflectance maps . . . 15
2.5 [M] Derivation of total radiance from a Lambertian reflector . . . 15
2.6 [M] The dichromatic reflection model in RGB colour space . . . 20
2.7 [M] Examples of filter transmitance and spectral responsivity curves . . . 22
2.8 [M] Light collected by the lens from a surface patch . . . 23
2.9 [M] CCD structure and operation . . . 24
2.10 [M] Full-Frame CCD architecture . . . 26
2.11 [C] Examples of colour filter arrays . . . 27
2.12 [C] Example of demosaicking . . . 27
2.13 [M] Noise sources affecting CCD operation . . . 29
2.14 [M] Example of pre-kneeing . . . 32

4.1 [M] Axis notation . . . 60
4.2 [M] Diagram of the experimental setup for physics-based image segmentation and edge detection . . . 64
4.3 [M] Confusion matrix for a multi-class classification problem . . . 68
4.4 [M] Illustration of the missing and false alarm rates in a particular case . . . 73
4.5 [M] Examples of segmentations and their discrepancy measures (I) . . . 74
4.6 [M] Examples of segmentations and their discrepancy measures (II) . . . 75
4.7 [M] Diagram of the performance evaluation setup for physics-based image segmentation and edge detection . . . 78

5.1 [M] Example of lighting orientation estimation (I) . . . 82
5.2 [M] Example of lighting orientation estimation (II) . . . 82
5.3 [M] Straight line fitting using orthogonal distances . . . 85
5.4 [M] Example of lighting orientation estimation (IV) . . . 87
5.5 [M] Example of lighting orientation estimation (V) . . . 87

* [M] indicates a monochrome figure, [C] indicates a colour figure

5.6 [M] Projection of sphere points under perspective . . . 89
5.7 [M] Angle δ between true and approximated surface normal vectors . . . 90
5.8 [M] Percentage of sphere points for which the relative error $\frac{|\Delta \cos\theta|}{\cos\theta} \times 100$ is below 5% . 90
5.9 [M] Estimation errors for the suggested calibration setup under orthographic projection 91
5.10 [M] Projection of a sphere over the image plane under perspective projection . . . 92
5.11 [M] Appearance of the projections of the sphere under different positions with regard to the camera optical axis . . . 94
5.12 [M] Symbols used in the derivation of $\frac{z_0}{R}$, $\frac{x_0}{R}$ and $\frac{y_0}{R}$ as a function of f . . . 95
5.13 [C] Example of estimation for CPAM . . . 101
5.14 [M] Summary of results for *orthographic versus perspective* projection experiments . . . 108
5.15 [M] Results for robustness experiment 1 . . . 109
5.16 [M] Examples of noisy images used in robustness experiment 2 . . . 109
5.17 [M] Results for robustness experiment 2 and OPAM . . . 110
5.18 [M] Results for robustness experiment 2 and CPAM . . . 111
5.19 [M] Results for robustness experiment 3 . . . 112
5.20 [C] Results for robustness experiment 4-1 and OPAM . . . 113
5.21 [C] Results for robustness experiment 4-1 and CPAM . . . 114

6.1 [M] Geometric relationships among camera noise model parameters . . . 120
6.2 [M] Confidence intervals for D^R . . . 124
6.3 [M] Discrepancy between $D^R - D^R_a$ and $t\sigma^R$, and between $D^R_b - D^R$ and $t\sigma^R$ . . . 125
6.4 [M] Estimation of camera gain and charge-independent non-spatial noise . . . 127
6.5 [M] Estimation of the PRNU and DCNU distribution parameters . . . 127
6.6 [C] Examples of synthetic noisy images generated by pseudocode 6.3 . . . 129
6.7 [M] Estimation of intensity uncertainties for every colour channel . . . 130
6.8 [C] Predicted *versus* real uncertainties for all colour channels . . . 132
6.9 [C] Illustration of UTLOG . . . 132
6.10 [M] Quantitative results of the comparison between UTLOG and NATLOG. . . . 134

7.1 [M] Step (k.2) of IS^{2}R region growing procedure . . . 142
7.2 [M] Sub-steps (k.4.1) and (k.4.2) of IS^{2}R region growing procedure . . . 143
7.3 [M] 1D neighbourhoods used in the consistency test . . . 149
7.4 [M] 1D example of a segmentation where a region can violate the reflectance-uniformity requirement . . . 151
7.5 [M] Illustration of the concept of intensity descending path in colour images . . . 152
7.6 [C] A detailed example of how IS^{2}R works . . . 155
7.7 [M] Synthetic scenes used to test IS^{2}R . . . 155
7.8 [C] Segmentation of the gray-level SPHERE images and shape reconstruction by means of IS^{2}R . . . 157
7.9 [C] Segmentation of the gray-level PARABOLOIDS images and shape reconstruction by means of IS^{2}R . . . 159
7.10 [C] Segmentation of the gray-level VASE images and shape reconstruction by means of IS^{2}R . . . 160
7.11 [C] Segmentation of the SPHERE, PARABOLOIDS and VASE colour images by IS^{2}R . 161
7.12 [M] Lighting conditions for IS^{2}R corresponding to the first set of experiments with real images. . . . 162
7.13 [C] Results of IS^{2}R for the first scene of the first set of images . . . 163
7.14 [C] Results of IS^{2}R for the second scene of the first set of images . . . 164
7.15 [C] Results of IS^{2}R for the third scene of the first set of images . . . 165
7.16 [C] Lighting conditions for IS^{2}R corresponding to the colour images of the first set of images . . . 165
7.17 [C] Results of IS^{2}R for colour images of the three scenes of the first set of images . . . 166
7.18 [M] Lighting conditions for IS^{2}R corresponding to the second set of experiments with real images. . . . 167

7.19 [C] Results of IS^{2}R for the gray-level images of the TOYS scenes of the second set of images . . . 167
7.20 [M] Results of IS^{2}R for the gray-level images of the CLOTHING scenes of the second set of images . . . 168
7.21 [M] Lighting conditions for IS^{2}R corresponding to the second set of experiments with real images. . . . 168
7.22 [C] Results of IS^{2}R for colour images of the TOYS and CLOTHING scenes of the second set of images . . . 169
7.23 [C] Results of IS^{2}R for the third set of images . . . 170
7.24 [C] Examples of noisy synthetic images used for comparing IS^{2}R, KLN and GEV . . . 171
7.25 [C] Comparison between IS^{2}R, KLN and GEV under increasing noise conditions by means of synthetic images . . . 173
7.26 [M] Ground truths for the real images used in the comparison between IS^{2}R, KLN, GEV and C&M . . . 175
7.27 [C] Segmentation results for IS^{2}R and C&M for the images of figures 7.8 and 7.15 . . . 176
7.28 [C] Segmentation results for IS^{2}R, KLN, GEV and C&M for the images of figure 7.11(a) 177
7.29 [C] Segmentation results for IS^{2}R, KLN, GEV and C&M for the images of figure 7.17(c) 178

8.1 [M] Four possible configurations of the straight line $\Delta D(i,j) = \Delta C_a + m_b(i,j)\Delta C_b$. . . 185
8.2 [C] A synthetic image without ambient lighting . . . 188
8.3 [M] Proposition 8.12 is not a double implication . . . 192
8.4 [M] A CHC and an NCD transition taking place simultaneously . . . 192
8.5 [M] Proposition 8.13 is not a double implication . . . 193
8.6 [M] Geometrical interpretation of properties 1-3 . . . 194
8.7 [M] Summary of how the body reflectance transitions detected manifest and corresponding edges . . . 195
8.8 [M] Second-order derivative output for a step edge . . . 196
8.9 [C] Relationship between LOG zero crossings and reflectance transitions . . . 197
8.10 [C] Examples of reflectance transitions in a real image . . . 199
8.11 [M] Redefinition of the $\mathcal{C}$ compatibility relation . . . 199
8.12 [C] Example of edge map (I) . . . 200
8.13 [C] Example of edge map (II) . . . 201
8.14 [M] Example of thinning of C^{3}E . . . 202
8.15 [M] Experimental results for C^{3}E and noisy synthetic images: Baddeley's measure for $t = 1..15$ . . . 204
8.16 [M] Comparison between C^{3}E and S&G under normal noise conditions by means of synthetic images . . . 205
8.17 [C] Examples of noisy synthetic images used for comparing C^{3}E and S&G . . . 206
8.18 [M] Comparison between C^{3}E and S&G under increasing noise conditions by means of synthetic images and Baddeley's measure . . . 207
8.19 [C] Real test images and their manually-generated ground truths . . . 208
8.20 [C] Experimental results for C^{3}E and real image 8.19(a) . . . 208
8.21 [M] Edge maps resulting for C^{3}E ($t = 4$), S&G and M&G for images in figure 8.19 . . . 210
8.22 [C] Edge maps resulting for C^{3}E ($t = 4$), S&G and M&G for more real images . . . 211

9.1 [M] Outline of C^{3}S . . . 214
9.2 [C] Example of output of the region growing stage of C^{3}S . . . 215
9.3 [C] Local analysis *versus* global analysis for region merging . . . 216
9.4 [C] A synthetic image of a scene containing two materials whose reflectances are parallel to one another . . . 217
9.5 [M] Strips used to compute the statistical measures involved in ζ_{con} . . . 218
9.6 [M] Detection of specular pixels . . . 220
9.7 [M] Plots of $\Delta\theta$ for different values of ΔD^{c*}, θ and ρ_b^c . . . 221
9.8 [M] Relationship between $\Delta\theta$ and surface normals of pixels involved . . . 221

9.9 [C] Examples of detection of specularities 223
9.10 [C] Example of output of the region merging stage 224
9.11 [M] Experimental results for C^3S and noisy synthetic images 226
9.12 [M] Comparison between C^3S, KLN and GEV under normal noise conditions by means of synthetic images 227
9.13 [C] Comparison between C^3S, KLN and GEV under increasing noise conditions by means of synthetic images 230
9.14 [C] Real test images and their manually-generated ground truths 231
9.15 [C] Experimental results for C^3S and real image 9.14(a) 231
9.16 [C] Comparison between C^3S, KLN, GEV and C&M by means of real images 233
9.17 [C] Best segmentation results for C^3S, KLN, GEV and C&M for images in figure 9.14(a-c) 236
9.18 [C] Best segmentation results for C^3S, KLN, GEV and C&M for images in figure 9.14(d-f) 237
9.19 [C] Real images used in the third experiment to compare the performance of C^3S, KLN, GEV and C&M 238
9.20 [C] Best segmentation results for C^3S corresponding to the third experiment with real images 238
9.21 [C] Best segmentation results for KLN corresponding to the third experiment with real images 239
9.22 [C] Best segmentation results for GEV corresponding to the third experiment with real images 240
9.23 [C] Best segmentation results for C&M corresponding to the third experiment with real images 241

10.1 [C] Diagram of the performance evaluation setup proposed in chapter 4 labelled with the contributions of this thesis 243

C.1 [C] Space of reflectance triplets for a 3-colour-channel imaging sensor 265

D.1 [M] Decay of u_{ij} with regard to the fuzzifier q 269

List of Tables

3.1 Summary of physics-based segmentation approaches 56

4.1 Scene attributes encoded in the image formation model 62
4.2 Percentage of non-spatial noise retained by frame averaging 64
4.3 Confusion matrix for a two-class classification problem 67
4.4 Discrepancy measures formulae 70
4.5 Other evaluation measures based on the overlapping area matrix (I) 71
4.6 Other evaluation measures based on the overlapping area matrix (II) 72
4.7 Examples of segmentations and their discrepancy measures (III) 76
4.8 Some measures for evaluating the performance of edge detection algorithms 77

5.1 Expression corresponding to a non-centered ellipse 84
5.2 Example of lighting orientation estimation (III) 86
5.3 Illustrative data for the lighting estimation example 102
5.4 Lighting parameters estimates using synthetic images generated under orthographic projection 103
5.5 Lighting parameters estimates using synthetic images generated under perspective projection, varying the tilt of the illumination 104
5.6 Lighting parameters estimates using synthetic images generated under central projection, varying z_0 105
5.7 Lighting parameters estimates using synthetic images generated under central projection, varying x_0 and y_0 so that $x_0^2 + y_0^2 = 0.01^2$ 106
5.8 Lighting parameters estimates using synthetic images generated under central projection, varying x_0 and y_0 so that $x_0^2 + y_0^2 = 0.15^2$ 107
5.9 Results for robustness experiment 4-2 and OPAM 115
5.10 Results for robustness experiment 4-2 and CPAM 116
5.11 Comparison results for synthetic spheres under orthographic projection 117
5.12 Comparison results for synthetic spheres under perspective projection 118

6.1 Contribution of the different noise sources to the uncertainty in digital intensity levels . 123
6.2 Comparison of the estimates resulting from the radiometric calibration. 128
6.3 Excerpt of the uncertainty table 131

7.1 Summary of parameters of IS^{2}R 154
7.2 Parameters used in the SPHERE, PARABOLOIDS and VASE scenes 156
7.3 Results for the SPHERE, PARABOLOIDS and VASE scene cases 156
7.4 Parameters used by IS^{2}R in the three scenes of the first set of real images. 162
7.5 Reflectances recovered by IS^{2}R for the three scenes of the first set of real images 162
7.6 Parameters used by IS^{2}R in the different scenes of the second set of real images. 164
7.7 Parameters for IS^{2}R during the comparison under increasing noise conditions 171

7.8 Parameters for KLN during the comparison with IS^{2}R under increasing noise conditions 171
7.9 Parameters for GEV during the comparison with IS^{2}R under increasing noise conditions 171
7.10 Comparison between IS^{2}R, KLN and GEV under increasing noise conditions by means of synthetic images . . . 172
7.11 Parameters for KLN during the comparison with IS^{2}R by means of gray-level and colour synthetic and real images . . . 173
7.12 Parameters for GEV during the comparison with IS^{2}R by means of gray-level and colour synthetic and real images . . . 174
7.13 Comparison between IS^{2}R, KLN, GEV and C&M by means of gray-level synthetic and real images . . . 174
7.14 Comparison between IS^{2}R, KLN, GEV and C&M by means of colour synthetic and real images . . . 175

8.1 Fulfillment of properties 1-3 . . . 189
8.2 Reflectance transitions expressed by propositions 8.11-8.13. . . . 191
8.3 Comparison between C^{3}E, S&G and M&G by means of real images . . . 209

9.1 Parameter values used for the images of figure 9.9 . . . 222
9.2 Parameter values used by C^{3}S in the experiments with synthetic images . . . 225
9.3 Parameters for KLN during the comparison with C^{3}S by means of synthetic images . . . 227
9.4 Parameters for GEV during the comparison with C^{3}S by means of synthetic images . . . 228
9.5 Comparison between C^{3}S, KLN and GEV under increasing noise conditions by means of synthetic images . . . 229
9.6 Comparison between C^{3}S, KLN, GEV and C&M by means of real images . . . 232
9.7 Parameter values for KLN during the experiments with the real images of figure 9.14 . . . 234
9.8 Parameter values for GEV during the experiments with the real images of figure 9.14 . . . 234
9.9 Estimates of directional illumination strength, L_d^c, for images of figure 9.14 . . . 234
9.10 Parameter values and performance measures for C^{3}S best segmentations of images of figure 9.14 . . . 234

10.1 Incorporation of IS^{2}R, C^{3}E and C^{3}S to the summary of physics-based segmentation approaches of table 3.1 . . . 244

List of Algorithms

5.1 Lighting parameters estimation pseudocode for the perspective projection case 97

6.1 Determination of the PRNU and DCNU distribution parameters 121
6.2 Determination of camera gain and non-spatial noise 121
6.3 Procedure for generating synthetic images with realistic noise 122

7.1 Iterative growth of a region $\mathcal{R}$ of uniform reflectance in IS^2R 141

9.1 Algorithmic formulation of the merging stage of C^3S 219
9.2 Test used by C^3S to detect a specularity peak 222

D.1 Algorithmic formulation of PCM 268

Symbols and Notation

This is the list of symbols most often used throughout the document with their default interpretation:

a	scalar a
$\boldsymbol{a}$	vector a
$\widehat{x}$	estimate of x
μ_x	image consisting in the pixel-by-pixel (temporal) average of several images x_t of the same scene taken at different time instants
$E_I[x]$	(spatial) average of the pixel values of image x
σ_x^2	image consisting in the pixel-by-pixel (temporal) variance of several images x_t of the same scene taken at different time instants
$\mathrm{Var}_I[x]$	(spatial) variance of the pixel values of image x
θ, σ	polar/slant angle
ϕ, τ	azimuth/tilt angle
(x, y, z)	3D co-ordinates of a scene point
(u, v)	co-ordinates of the projection of a scene point over the image plane (standard length units)
(i, j)	image co-ordinates (in pixels)
ω	solid angle
Ω	projected solid angle
E	irradiance
L	radiance
λ	wavelength
Λ	set of wavelengths within the visible spectrum (approximately between 380 (blue) and 760 (red) nanometers)
$\boldsymbol{N}$	surface normal vector
$\boldsymbol{S}$	direction towards the only directional light source
$\boldsymbol{n}$	unit surface normal vector

$\boldsymbol{s}$	unit direction towards the only directional light source
$\boldsymbol{v}$	unit direction towards the viewer
$\boldsymbol{h}$	unit angular bisector of $\boldsymbol{v}$ and $\boldsymbol{s}$
L_a	ambient radiance
L_d	radiance of a directional light source
$\rho_a(i,j)$	ambient reflectance at image cell (i,j)
$\rho_b(i,j)$	body reflectance at image cell (i,j)
$\rho_i(i,j)$	interface reflectance at image cell (i,j)
$C_a^c(i,j), (L_a\rho_a(i,j))^c$	ambient composite reflectance for colour channel c at image cell (i,j)
$C_b^c(i,j), (L_d\rho_b(i,j))^c$	body composite reflectance for colour channel c at image cell (i,j)
$C_i^c(i,j), (L_d\rho_i(i,j))^c$	interface composite reflectance for colour channel c at image cell (i,j)
$m_b(i,j)$	geometric factor for body reflectance at image cell (i,j)
$m_i(i,j)$	geometric factor for interface reflectance at image cell (i,j)
$D_a^c(i,j)$	ambient illumination reflection term for colour channel c at image cell (i,j)
$D_b^c(i,j)$	body reflection term for colour channel c at image cell (i,j)
$D_i^c(i,j)$	interface reflection term for colour channel c at image cell (i,j)
$I^c(i,j)$	number of electrons collected at image cell (i,j) for colour channel c
$\tau^c(\lambda)$	filter transmittance for colour channel c
$s(\lambda)$	spectral responsivity of the sensor
$K(i,j)$	photo-response for image cell (i,j)
$\mu_{dc}(i,j)$	average dark charge for image cell (i,j)
$N_S^c(i,j)$	zero-mean random variable accounting for photon noise at image cell (i,j) for colour channel c
$N_{dc}(i,j)$	zero-mean random variable accounting for dark current shot noise at image cell (i,j)
N_R	camera read-out noise
N_Q	quantization noise
A^c	camera gain for colour channel c
$\mu^c(i,j)$	average intensity value for image cell (i,j) and colour channel c
$N^c(i,j)$	noise for image cell (i,j) and colour channel c
$N_e^c(i,j)$	noise for image cell (i,j) and colour channel c dependent on the number of collected electrons
$N_f^c(i,j)$	noise for image cell (i,j) and colour channel c independent of the number of collected electrons
$D^c(i,j)$	output of the camera for colour channel c and image cell (i,j)
f	camera focal distance
f/d	camera F-number
R_i	region i in the reference segmentation
$\widehat{R}_j$	region j in the segmentation output

Acronyms

This is the list of acronyms introduced throughout this documentation to name the different algorithms presented:

C^3E	*Colour Channel Coupling-based Edge* detection algorithm	(chapter 8)
C^3S	*Colour Channel Coupling-based Segmentation* algorithm	(chapter 9)
CPAM	*Central Projection Assuming Method* of lighting parameters estimation	(chapter 5)
IS^2R	*Image Segmentation by Scene Reconstruction*	(chapter 7)
OPAM	*Orthographic Projection Assuming Method* of lighting parameters estimation	(chapter 5)
R^2CIU	*Robust Radiometric Calibration for Intensity Uncertainty* estimation algorithm	(chapter 6)
TEAV	lighting *Tilt Estimation by AVeraging*	(chapter 5)
TECA	lighting *Tilt Estimation by Centers Alignment*	(chapter 5)

Other acronyms corresponding to algorithms by other authors also appearing along the documentation are the following:

B&P	Bichsel and Pentland Shape from Shading algorithm [17]	(chapter 7)
C&M	Comaniciou and Meer image segmentation algorithm [29, 30]	(chapters 7,9)
GEV	Gevers image segmentation algorithm [80]	(chapters 7,9)
H&K	Healey and Kondepudy camera radiometric calibration algorithm [93]	(chapter 6)
KLN	Klinker et al. image segmentation algorithm [129]	(chapters 7,9)
L&R	Lee and Rosenfeld lighting estimation method [141]	(chapter 5)
M&G	Meer and Georgescu edge detection algorithm [168]	(chapter 8)
S&G	Stokman and Gevers edge detection algorithm [79, 262]	(chapter 8)
Z&C	Zheng and Chellapa lighting estimation method [311]	(chapter 5)

Finally, other acronyms often used throughout this document are:

ACRM	*Approximate Color Reflectance Model*
BRDF	*Bidirectional Reflectance Distribution Function*
CCD	*Charge Coupled Device*
CG	percentage of *Correctly Grouped* pixels
DCNU	*Dark Current Non-Uniformity*
DRM	*Dichromatic Reflection Model*
FOM	(Pratt's) *Figure of Merit*
MVE	*Minimum Volume Ellipsoid* estimator
NIR	*Neutral Interface Reflection* model

OS	percentage of *Over-Segmentation*
PCA	*Principal Component Analysis*
PCM	*Possibilistic C-means* algorithm
PRNU	*Photo-Response Non-Uniformity*
SPD	*Spectral Power Distribution*
URM	*Unichromatic Reflection Model* for metals
US	percentage of *Under-Segmentation*

1

Introduction and Objectives

This chapter defines the problem which is addressed in this thesis. To this end, section 1.1 introduces briefly the field of computer vision, section 1.2 states the problem for which is intended to provide a solution and sets the goals to attain, and, finally, section 1.3 outlines this documentation.

1.1 The Computer Vision Problem

Almost any biological form is endowed with more or less sophisticated sensorial structures evolved to provide information to the organism about the surrounding environment so as to be able to interact with it. Without the information furnished by these sensors, it would be impossible to carry out survival-related tasks such as recognizing the different entities populating the environment (species- and group-mates, enemies, food, obstacles, etc.) or just navigating safely through the environment.

A large amount of sensorial structures can be found among the immense variety of known biological forms. Many of them appear in almost every species, differing amongst them in structure, in capabilities and in the sort of information which is processed and finally got from the environment. In all cases, however, information is recovered transforming first and processing next energy that reaches the organism in a certain form —electromagnetic radiation, mechanical energy (i.e. acoustic waves), etc.— which depends on the particular sensorial structure. The fact that a certain "sensor" has been evolved by a species will depend, thus, on how likely is that the required energy form travels through the environment where that species lives.

Vision is one of these structures. Probably, it is the most frequent in nature among the different evolved animal forms. The fact that human beings depend so much on vision for their daily life and that a huge amount of information can be extracted from the environment through vision have made it as the object of study of lots of researchers throughout the world, coming from the most varied disciplines for decades. The effort has been set off for not only the study of the structures that make vision possible *per se*, but also regarding how to get the information about the environment which is coded in the electromagnetic radiation in which it propagates. Throughout the years, terms such as *computer vision*, *computational vision* or *image understanding* have been coined for referring to somewhat overlapped aspects of this work. Aloimonos and Rosenfeld in [3] set for computer vision the goal of determining both the visual information necessary for a given task and the procedures to obtain such information from the data provided by the optical sensors utilized in the particular application considered.

The set of processes involved in vision are often classified within a hierarchy which goes from what is known as *early* or *low-level vision* through *intermediate vision* to *high-level vision* [286]. These notions of low- and high-level vision are used as a matter of routine although there is not a definitive definition for each. However, it is accepted that:

- Low-level vision is usually associated with the extraction of certain physical properties of the environment, such as depth, 3D shape, object edges, surface material properties, etc.

- High-level vision, however, has to do with problems such as the computation of shape properties and spatial relationships, as well as with object recognition and classification. It is, therefore, more related with the interpretation of the image and the use of the information recovered rather than with the determination of the physical properties of scene objects. These processes typically depend on the goals of the application together with the knowledge available about the objects expected in the scene.
- Finally, processes within the domain of high-level vision that do not depend on knowledge about the scene are typically classified as intermediate vision. Segmentation is one of the processes which is sometimes located at this level.

Nevertheless, although it is usual thinking of vision as a sort of pipeline, where low-level processes are in one end and high-level processes are in the other end, it is not clear that low-level processes must end before the start of higher-level processes. It is rather a sort of distinction as for the type of information entering and exiting every particular process.

Among the different research results obtained in computer vision throughout the last 30-40 years, the most sound one is that emulating human vision is not so trivial as it seems because of the tremendous apparent simplicity how humans get information from the outer world. In fact, computer vision can be one of the most complex problems within the general goal of understanding and replicating intelligence, as indicates the fact that more than 50% of the neurons of the brain are devoted to vision [2]. Besides, most problems in early vision related with recovering scene properties admit, a priori, several solutions, because of the lack of enough constraints, or because the problem is just badly defined [115]. On the other hand, the knowledge about biological vision systems accumulated throughout the last thirty years is fragmented and is mostly confined in early processing stages, directly related with sensor signals. It is not surprising, thus, that some ambitious attempts to provide visual capabilities to machines have not produced the expected results.

However, potential applications of computer vision are so important that even partial and short-term solutions have been, and in fact are, useful. Besides, since the first steps in image processing, significative advances have taken place, leading to the development of systems able to carry out some visual functions, sometimes in structured environments, but also in less controlled conditions in many other times.

1.2 Problem Statement and Objectives

1.2.1 General discussion

To humans, an image is not just a random collection of pixels, but a meaningful arrangement of regions. Despite the large diversity of shapes, colours and textures which can be contained in images, humans have no difficulty interpreting them. We even agree about the different regions in the image and are able to label them semantically in a hierarchy of different levels of detail, to distinguish an object to one another but also to identify parts of the same object.

In computer vision, grouping parts of a general image into units that are semantically meaningful is known as the process of *image segmentation*. This idea of segmentation has its roots in the research carried out by the Gestalt psychologists in the early part of the twentieth century, who extensively studied the preferences exhibited by human beings in grouping or organizing sets of shapes appearing in the visual field [11]. They identified several factors that lead to human perceptual grouping: similarity, proximity, continuity, symmetry, parallelism, closure and familiarity [36]. As a consequence, a strong belief in these grouping mechanisms as organizers of the scene in meaningful units from which the image could be understood arose and spread among the researchers. Since then, some of these clues have been used as guidelines for many grouping algorithms in computer vision [155].

From a computational point of view, any process of grouping is inevitably associated with a criterion of homogeneity $\mathcal{H}(\cdot)$, which determines the properties of the wanted *segments*. The segmentation problem has traditionally been formulated in a formal way in these terms, as a function of $\mathcal{H}(\cdot)$, as follows [26, 32, 108, 153, 220]:

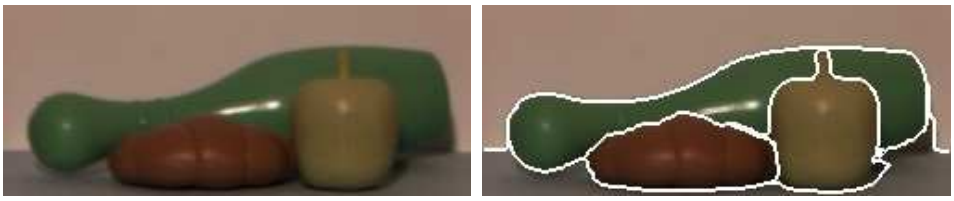


Fig. 1.1. Example of segmentation where the meaningful structures segmented coincide with scene objects (algorithm C^{3}S, chapter 9).

Given an image $\mathcal{I}$ and a homogeneity criterion $\mathcal{H}(\cdot)$, find a partition $\mathcal{P}$ of $\mathcal{I}$ into a set of regions $\widehat{R}_1$, $\widehat{R}_2$, ..., $\widehat{R}_n$ such that (by definition of partition, $\bigcup_{i=1}^{n} \widehat{R}_i \subseteq \mathcal{I}$ and $\widehat{R}_i \bigcap \widehat{R}_j = \emptyset$, $\forall i \neq j$):

(a) each $\widehat{R}_i$ is connected (between each pair of pixels $p, q \in \widehat{R}_i$ there is a path $p = p_1, p_2, ..., p_t = q$ of adjacent pixels $p_k \in \widehat{R}_i$, $k = 1, \ldots, t$, either in the sense of 4-connectivity or 8-connectivity),

(b) $\mathcal{H}(\widehat{R}_i) = \mathit{true}$, $\forall i = 1, \ldots, n$, and

(c) $\mathcal{H}(\widehat{R}_i \cup \widehat{R}_j) = \mathit{false}$, $\forall i \neq j$ and $\widehat{R}_i$ and $\widehat{R}_j$ have a common boundary.

Broadly speaking, $\mathcal{H}(\cdot)$ should aim at identifying meaningful structures in the image. From the most general interpretation of the segmentation problem (even beyond the previous definition), such adjective, meaningful, can refer to:

(1) what higher-level vision processes intend to recognize, which makes segmentation an application dependent task within the vision pipeline, or else

(2) perceptually coherent units assimilable to scene objects, which, to a certain extent, leads to thinking of segmentation as a domain independent process.

While the former meaning alludes to vision tasks such as, for instance, the segmentation of the straight lines of an image, the latter connotation is inherently more ambitious, since identifies meaningful structures with objects and defines the goal of the segmentation task as the generation of data structures which allow isolating the different objects present in the scene. Without doubt, this second meaning corresponds to what comes readily to mind when thinking of image segmentation. By way of illustration of a segmentation according to this last meaning, figure 1.1 provides the output of C^{3}S, one of the segmentation algorithms proposed in this thesis (chapter 9).

If $\mathcal{H}(\cdot)$ is specified in the previous terms, however, the segmentation task becomes, at first sight, an *ill-posed problem* because, to put $\mathcal{H}(\cdot)$ in a formal way, the question *"what is the image of an object?"* should be first answered, and it is not likely to find a unique answer. Intuition would probably say the image of an object consists of a region of nearly constant colour, or, from the point of view of the whole image, the intensity of any colour channel is a (perhaps noisy) piecewise constant bi-dimensional function whose plateaus correspond to scene objects. Nevertheless, what is known as an image is actually the result of sensing the electromagnetic radiation coming from the scene and, as will be seen later, some images correspond to the previous model, as it is the case of the image of figure 1.2, but, unfortunately, it is not the general case. Generally speaking, the intensity values provided at pixel locations by a camera are a combination of:

- the illumination distribution,
- the reflection properties of scene objects,
- the geometry of the scene,
- the propagation medium, and
- the performance of the imaging sensor.

The physical laws that rule the processes of light propagation, reflection and image acquisition determine, thus, the final pixel value. By way of example, figure 1.3 shows what should be expected in a general image. As can be observed in the profiles of the scanline, the image intensity is not piecewise constant if curved objects appear in the scene or if glossiness can be appreciated.

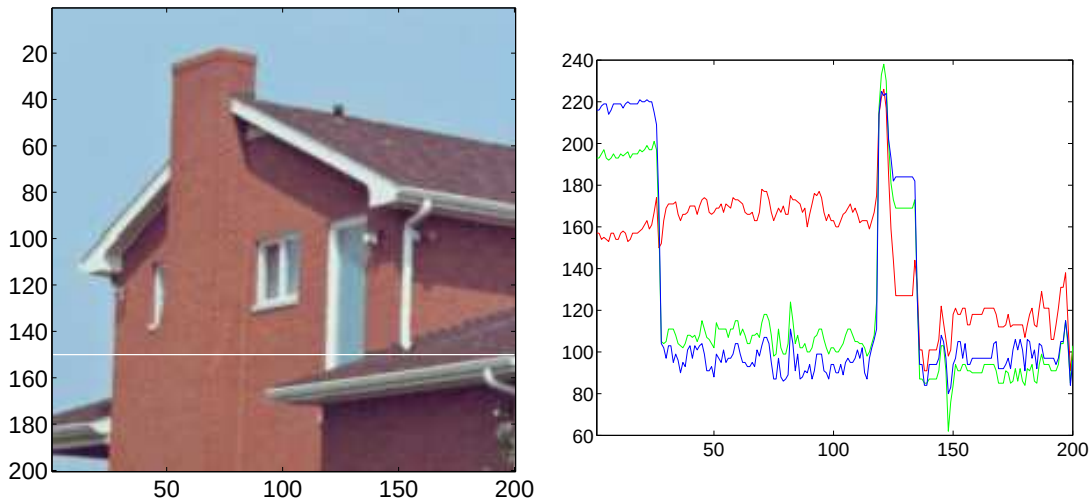


Fig. 1.2. Intensity profiles of the *house* image: [left] *house*; [right] plot of the intensity profiles for the red, green and blue colour channels at scanline 150 (in white in the image).

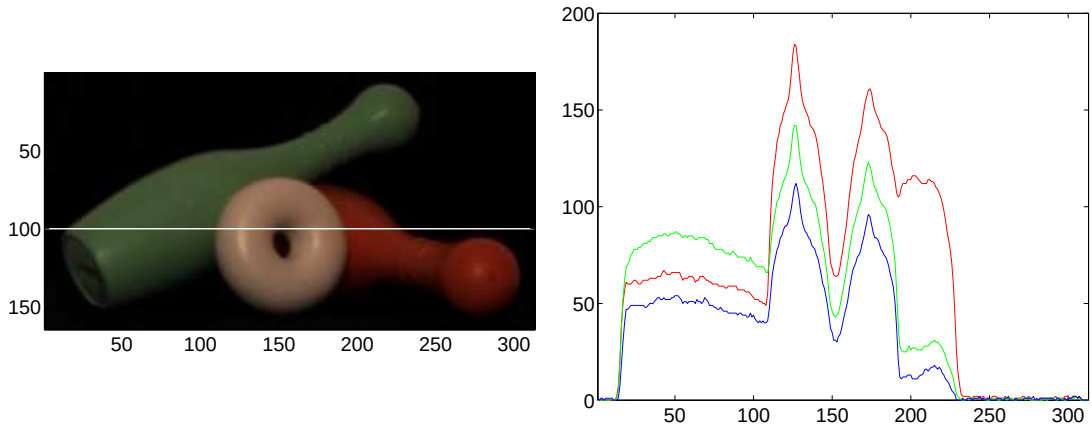


Fig. 1.3. Intensity profiles of a general image: [left] a general image; [right] plot of the intensity profiles for the red, green and blue colour channels at scanline 100 (in white in the image).

If the scene effectively consists of quasi-planar surfaces and they do not show glossiness, effects such as shading and specularities are hardly noticeable in the image, and areas of the scene of uniform perceptual colour appear as regions of more or less constant colour intensity provided that shadows and inter-reflections are avoided or are negligible. In this case, the model of a noisy piecewise constant function is, thus, valid. Throughout the years, many segmentation algorithms have exploited this assumption for both uniformly coloured and textured objects[1] with pretty nice results for that kind of images, as it is shown in figure 1.4 for the standard image *house* and the mean-shift-based algorithm by Comaniciu and Meer [30]. However, when the effects of scene curvature are noticeable in the image, one can obtain what is shown in figure 1.5 from the same algorithm.

In general, objects curvature and objects glossiness, together with other optical phenomena such as shadows and inter-reflections, give rise to changes in image intensity not necessarily related to object boundaries, what makes methods not taking into account them perform poorly with this sort of images, typically producing over-segmentation. However, embedding a physics-based model of image

[1] See [26] among many other existing surveys on the subject.

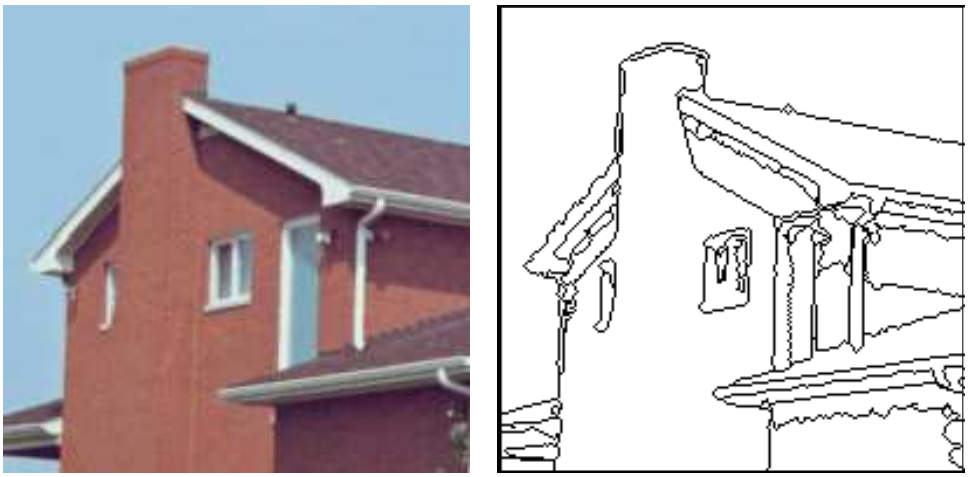


Fig. 1.4. Segmentation results for the *house* image by the algorithm published in [29]: [left] original image; [right] region contours.

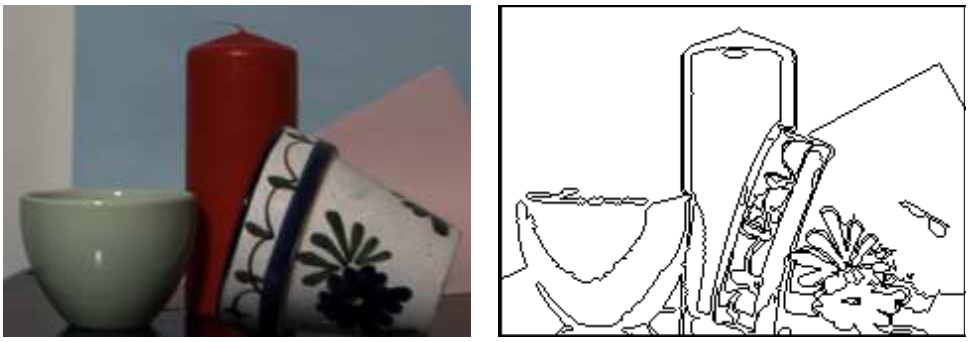


Fig. 1.5. Segmentation results for images of curved objects by the algorithm published in [29]: [left] original image; [right] region contours.

formation into the segmentation algorithm allows coping with those effects in a more suitable way. By way of illustration, figure 1.6 provides the results produced for the two previously shown images by the physics-based segmentation algorithm C^3S (chapter 9).

Irrespectively of the physics- or non-physics-based approach taken, the segments resulting from the segmentation of an image can be described in two different ways, contours and regions, what leads to the two broad families of segmentation approaches: region-based and contour-based. These two approaches need not be that different from each other: on the one hand, boundaries of regions can be defined to be contours, and, on the other hand, if one enforces closure in a contour-based framework, then one can get regions from a contour-based approach. The difference is more on whether the emphasis is put on similarity or on dissimilarity. Nevertheless, vision researchers usually distinguish in the titles of their papers between both approaches with terms such as *edge detector*, *segmentation algorithm* or related variants. This is the reason why the title of this thesis refers to both broad families, although, as has been said, from a theoretical point of view, one is the complementary of the other and, therefore, the term *image segmentation* should be enough.

1.2.2 Objectives of this thesis

In order to contribute to the solution of the problem of image segmentation, this thesis addresses the following set of goals:

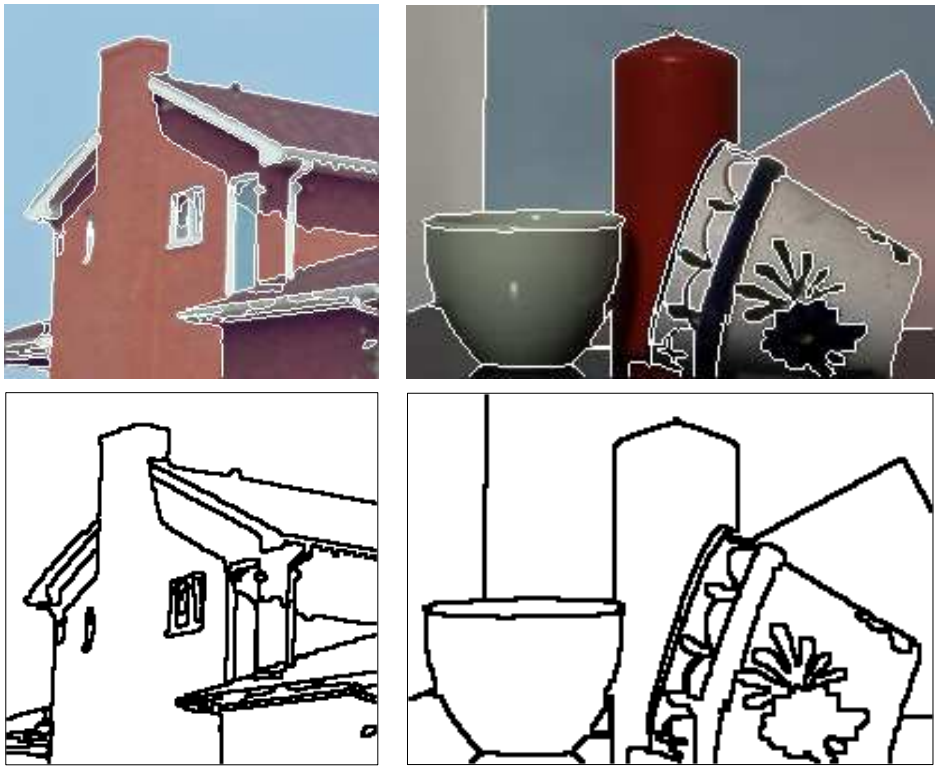


Fig. 1.6. Segmentation results for the physics-based segmentation algorithm C^3S (chapter 9). In the upper row, region contours are given superimposed in white over the original image.

(1) *study the segmentation problem within the physics-based vision paradigm*, taking into account all the elements involved in the process of image formation, from the reflection of light in the scene to the capture of the image by means of an optical sensing device;
(2) *explore several unsupervised approaches* and provide algorithms for each;
(3) consider physics-based models of image formation including *ambient illumination*;
(4) ensure the *coverage of object transitions* of the algorithms is not cut off from their theoretical inception;
(5) *work at the signal level*, or, in other words, minimize the use of domain information within the algorithm;
(6) *take into account image noise models* reasonably accurate in the derivation of the algorithm in order to be able to set the corresponding parameters in an adaptive way; and
(7) *compare the different algorithms developed* with recognized segmentation and edge detection algorithms, both physics- and non-physics-based.

In this list, goals (1) and (2) could be classified as the general objectives of the work which is going to be presented throughout the following chapters. Goals (4) and (7), on the other hand, correspond to deficiencies which have been found in many physics-based algorithms and in the pieces of work where they were presented, and, for this reason, have been considered fundamental throughout the realization of this thesis. Finally, goals (3), (5) and (6) intend to constitute added values of the algorithms to be proposed.

Furthermore, this set of goals is intended to be achieved under the following constraints:

- the scenes are assumed to contain only uniformly coloured objects, not textured objects, except for fine-grained textures which unavoidably can be present in the scene;
- the segmentation task will have to be achieved on the basis of an only image; and
- images have to be assumed to consist of one or several bands (i.e. gray-level and colour images).

1.3 Thesis Outline

This documentation is organized into ten chapters and several appendices providing additional information. Chapter 1 has introduced briefly the field of computer vision, has stated the problem for which is intended to provide a solution, has set the goals to attain, and, finally, in this section, the structure of this documentation is shown.

Chapter 2 is devoted to introducing the several aspects of the physical processes involved in the formation of images in real cameras. The different sections provide a summary of the radiometrical concepts used throughout the chapter, explain light-matter interaction and physics-based reflection models and describe the model of operation of current optical imaging sensors. This chapter is intended to cover all the aspects of the formation of the image with full detail, what means that, perhaps, in some points exceeds what is strictly needed to understand the rest of this documentation.

Chapter 3 summarizes the most significant work related with physics-based segmentation and edge detection, starting with a brief review on general approaches for image segmentation. A taxonomy for physics-based segmentation algorithms is also introduced.

Chapter 4 focuses on the development of a general framework for image segmentation and edge detection on the basis of physics-based image formation models. Besides, an experimental setup for these sort of algorithms is presented and discussed, together with some new measures for characterizing the performance of image segmentation algorithms.

Chapter 5 presents several algorithms for estimating the lighting distribution parameters of a scene, what is needed by some segmentation algorithms relying on the physics of image formation. The model of illumination assumed by these lighting estimation algorithms, although limited to ambient lighting and an only point light source, is in accordance with the requirements of the segmentation algorithm of chapter 7.

Chapter 6 introduces an algorithm called R^2CIU for characterizing the noise sources affecting the performance of current CCD cameras. The main purpose of this characterization is to estimate intensity uncertainties to be incorporated in the segmentation algorithms of the following chapters.

Chapter 7 proposes a segmentation algorithm named IS^2R based on estimating the reflectance of objects surface and using it together with information of the shape of the scene to join pixels to the currently grown region. To this end, the scene is partially reconstructed from the image data using a shape from shading technique.

Chapter 8 studies the use of the coupling between colour channels in uniform reflectance areas to find material changes in the image and presents an edge detection algorithm called C^3E making use of the results of the study.

Chapter 9 takes advantage of the complementarity between edge detection and image segmentation to devise a segmentation algorithm called C^3S using the edge maps produced by C^3E.

Finally, chapter 10 concludes this documentation, enumerating the different conclusions which have been drawn during the realization of this thesis. Future lines of research are also proposed and discussed.

Preliminary and final versions of the different contributions of this thesis have been published in the form of technical reports [199,200,202,203,208,209,215] and conference papers peer-reviewed [201,204–207,210–214]. Furthermore, [207] was considered as the 2nd best paper of the conference and granted with a mention of honor.

2

The Image Formation Process

Generally speaking, the digital values provided by a camera are the result of several scene factors which interact with each other: the illumination distribution, the reflection properties of scene objects, the objects geometry, the propagation medium and the performance of the imaging sensor. Throughout the years, a lot of knowledge has been gained about the interaction of light with matter, with a great deal of help from the computer graphics community in their attempts to produce every time more and more realistic images [83]. Unfortunately, selecting appropriate image formation models for machine vision can be quite difficult. If one desires the most accurate available model, then light-matter interaction can be described in terms of the interaction of photons with atoms or molecules. It would, however, be a formidable task to derive computer vision algorithms from a physical model at such a low level. To avoid this situation, several models of reflection and light propagation manageable enough so as to be effectively incorporated in machine vision algorithms at the cost of some loss of physical accuracy have been developed. A sophistication of this extent has allowed, in particular, to handle images more complex than those which have been modeled through statistical techniques for many years. The implementation of tasks such as highlight identification and extraction, material classification or colour constancy has become feasible as well.

This chapter is devoted to introducing the main aspects of the physical processes involved in the formation of images in real cameras, together with the models generally accepted by the physics-based vision research community. First of all, section 2.1 provides a brief introduction to radiometrical concepts. Next, light-matter interaction and physics-based reflection models are the object of study of sections 2.2 and 2.3. Finally, optical imaging sensors and their models of operation are described in section 2.4. Other factors involved in the formation of the image, such as refraction processes leading to transparency, the influence of the propagation medium or the interaction of light with light-emitting surfaces, by either means fluorescence, incandescence or fosforescence, are not covered by this work and are not included in the following. In particular, [83] unveils part of the physics underlying transparent materials and emissivity of surfaces, while the attenuation of light as it travels through the propagation medium has been covered by several studies and surveys, such as [157,185,243], concerning atmospheric phenomena, or [181, 198, 221], for the case of underwater environments.

2.1 Radiometry

2.1.1 Radiometrical concepts

The most comprehensive and least controversial vocabulary about radiometry is the one proposed by the Commission Internationale de l'Eclairage (CIE) [31], although very good descriptions of radiometric terms can be found in several papers and books [104, 145, 184]. In the following, a brief summary of terms is included:

1. *Solid angle*: The solid angle ω of a cone of directions is defined as the area cut by the cone on the unit sphere. In the case of a sphere with radius r:

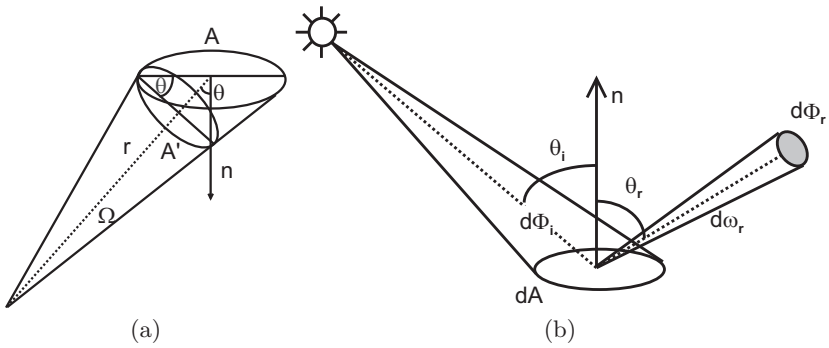


Fig. 2.1. Radiometrical terms and notation: (a) solid angle subtended by a patch of area A at a distance r from the origin and surface normal $\boldsymbol{n}$; (b) symbols for flux-related radiometrical terms.

$$\omega = \frac{A_{sphere}}{r^2} \text{ (steradians)} . \tag{2.1}$$

A small planar patch of area A at distance r from the origin and surface normal $\boldsymbol{n}$ subtends a solid angle Ω such as:

$$\Omega = \frac{A'}{r^2} = \frac{A\cos\theta}{r^2} = \omega\cos\theta , \tag{2.2}$$

where θ is the angle between the surface normal $\boldsymbol{n}$ and a line connecting the patch to the origin (see figure 2.1(a)). That is to say, A' is the surface A as seen from the origin. This effect is known as the *foreshortening* of surface A.

2. *Radiant Flux*(Φ): Power emitted, propagated or received as optical electromagnetic radiation (W).
3. *Radiant Intensity*(I): The radiant intensity of a source is the exitant flux per unit solid angle ($W \cdot sr^{-1}$); that is to say, the outgoing flux in a given direction:

$$I = \frac{\mathrm{d}\Phi_r}{\mathrm{d}\omega_r} . \tag{2.3}$$

The total flux emitted by a source is the integral of radiant intensity over the full sphere of possible directions (4π sr).

4. *Irradiance*(E): Amount of light falling on a surface. It is the flux per unit area ($W \cdot m^{-2}$) incident on the surface. That is, the incident flux density:

$$E = \frac{\mathrm{d}\Phi_i}{\mathrm{d}A} . \tag{2.4}$$

5. *Radiant Exitance*(M): It is the exitant flux density, measured in $W \cdot m^{-2}$:

$$M = \frac{\mathrm{d}\Phi_r}{\mathrm{d}A} . \tag{2.5}$$

6. *Radiance*(L): Amount of light radiated from a surface in a given direction. It is the power per unit area per unit solid angle ($W \cdot m^{-2} \cdot sr^{-1}$) emitted from the surface. The term solid angle appears because a surface can radiate energy in a hemisphere of directions, and different amounts of energy in different directions:

$$L = \frac{\mathrm{d}^2\Phi_r}{(\mathrm{d}A\cos\theta_r)\mathrm{d}\omega_r} . \tag{2.6}$$

The unit area mentioned in the definition of radiance is the *foreshortened* area: the surface area multiplied by the cosine of the angle between a perpendicular to the surface and the specified direction. Hence, radiance can be equivalently defined as the flux emitted per unit surface area per unit projected solid angle.

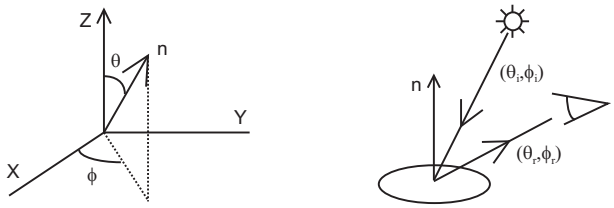


Fig. 2.2. Angles notation for the bidirectional reflectance distribution function (BRDF).

Visible light is normally a mixture of energy at many wavelengths and is characterized by a *spectral power distribution* (SPD) to indicate how much energy is present at each wavelength. Many times the SPD is given relative to the highest value and is called a *relative SPD*. Due to this wavelength dependence, many times it is necessary to specify the aforementioned radiometric quantities and other magnitudes as a function of wavelength. In such cases, the adjective *spectral* is added in front of the term, and its units become per unit wavelength interval (nanometer). For instance, *spectral radiance* is $L(\lambda) = \mathrm{d}^3\Phi/(\mathrm{d}\omega\mathrm{d}\lambda\mathrm{d}A\cos\theta)$.

2.1.2 The Bidirectional Reflectance Distribution Function

The *Bidirectional Reflectance Distribution Function* (BRDF) $f(\theta_i, \phi_i; \theta_r, \phi_r)$ relates the radiance exitant from a surface in a given direction (θ_r, ϕ_r) with light irradiance falling on it from another (θ_i, ϕ_i), being θ the so-called polar angle and ϕ the azimuth angle[1]. This function was originally introduced by Nicodemus et al. [184] as a unified notation for the specification of reflectance in terms of both incident- and reflected-beam geometry:

$$f(\theta_i, \phi_i; \theta_r, \phi_r) = \frac{\mathrm{d}L(\theta_r, \phi_r)}{\mathrm{d}E(\theta_i, \phi_i)} \,. \tag{2.7}$$

For many surfaces, the radiance is not altered if the surface is rotated about the surface normal. In this case, the BRDF depends only on the difference $\phi_r - \phi_i$. This is certainly true of matte surfaces and specularly reflecting surfaces. It is not true of surfaces as for example the mineral called tiger's eye or the iridiscent feathers of some birds (these all have a distinct directionality in their surface microstructure).

In any case, the exact form of a BRDF depends on the surface material. It can be determined experimentally by illuminating a flat sample of the material of interest with a lamp mounted on a goniometer and measuring its irradiance using a sensor mounted on another goniometer. Another way to obtain the BRDF is to model how light is reflected from a surface and to find the corresponding reflectance properties analytically or by numerical simulation. Some studies have been published about this subject, like e.g. [34].

The original definition of the BRDF given above can be extended to include the light wavelength as a variable. The resulting function is called the *bidirectional spectral-reflectance distribution function* (BSRDF) which is defined as:

$$f(\theta_i, \phi_i; \theta_r, \phi_r; \lambda) = \frac{\mathrm{d}L(\theta_r, \phi_r; \lambda)}{\mathrm{d}E(\theta_i, \phi_i; \lambda)} \,. \tag{2.8}$$

2.2 Interaction of Light with Matter

2.2.1 Introductory concepts

When a ray of light strikes the surface of an object it may be absorbed, transmitted or reflected depending on its optical properties. These optical properties can be related to the electromagnetic

[1] The polar and the azimuth angles are also known as, respectively, the slant and the tilt angles.

quantities ϵ (electrical permitivity, $C^2/(N \cdot m^2)$), μ (magnetic permeability, $N \cdot s^2/C^2$) and σ (conductivity, $(\Omega \cdot m)^{-1}$) and are summarized by the *complex index of refraction* $M = n - K_0 i,\ n, K_0 \in R$. The dimensionless quantity n is called the *refractive part* of M. In a nonattenuating medium (i.e. the amplitude of an electromagnetic wave does not decrease as the wave propagates in the material), n is the ratio of the speed of light in vacuum to the speed of light in the material ($n = c/v_{material}$) and $K_0 = 0$.

For absorbing media, K_0 takes values above 0 and the interpretation of n is more complicated. Because of this, K_0 is called the *absorptive part* of M. By way of example, in case the absorption is homogeneous, a plane wave of angular frequency ω traveling in the x direction that encounters an interface at $x = 0$ suffers an exponential attenuation which can be written as:

$$E(x) = E(0)e^{\frac{-2\omega K_0 x}{c}}, \tag{2.9}$$

where $E(x)$ represents the irradiance at a depth x in the material. The distance $x = c/(2\omega K_0)$ is often called the *skin depth* of the material.

It is possible to solve for n and K_0 in terms of the fundamental electromagnetic properties of the material (ϵ, μ, σ) by requiring that the incident light wave satisfy Maxwell's equations. The resulting expressions for n and K_0 are given by:

$$n^2(\lambda) = \frac{\mu\epsilon c^2}{2}\left[1 + \left[1 + \left(\frac{\lambda\sigma}{2\pi c\epsilon}\right)^2\right]^{1/2}\right], \tag{2.10a}$$

$$K_0^2(\lambda) = \frac{\mu\epsilon c^2}{2}\left[-1 + \left[1 + \left(\frac{\lambda\sigma}{2\pi c\epsilon}\right)^2\right]^{1/2}\right]. \tag{2.10b}$$

It is important to note that, for a given material, ϵ and σ depend on wavelength, λ.

Further, it is common to divide materials into two classes on the basis of their optical properties [252]:

1. *optically homogeneous* materials, which exhibit a constant index of refraction throughout the material, and include metal, glass and crystals as the most common examples; and
2. *optically inhomogeneous* materials, which, on the other hand, consist of a substrate with many embedded colorant particles that differ optically from the substrate, and comprise plastics, paper, textiles and paints among many others.

It is also usual to distinguish between *metals* and *dielectrics* because of their different behaviour with regard to light reflection. The most prominent difference between metals and other materials is the large number of electrons that are free to move throughout a metal, causing them to have large values of conductivity σ. By contrast, electric charges are not free to move in dielectrics, and thus these materials result to be poor conductors. Another important property of metals is that they are optically homogeneous, while there exists homogenous and inhomogeneous dielectrics.

Real reflection tends to be more complex than described in the rest of this section. However, the given explanations are very useful approximations [252] (see figure 2.3 for a graphical description). In the following, the terms *interface* and *body reflection* will be used with preference against the more popular terms *specular* and *diffuse reflection*, because the latter are used for different concepts in different fields.

2.2.2 Interaction of light with metals

The optical properties of a metal are largely determined by how light interacts with its electrons. The various possible interactions can be divided into two classes.

In the first class of interaction, light is *scattered* by free conduction electrons. This class of interaction turns out to be the most common in metals because of the large number of free electrons. These free electrons scatter light efficaciously, causing metals to have a small skin depth and large

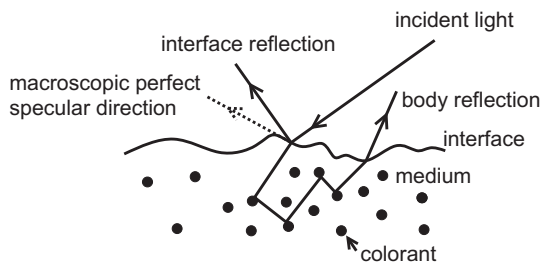


Fig. 2.3. Body reflection *versus* interface reflection over a dielectric.

reflectance values. Since these free electrons have no natural frequencies, they scatter light of all wavelengths equally well. This is why most metals have spectral reflectance functions that do not vary with wavelength in the visible range and thus appear silver or gray in color.

In addition to being scattered by free electrons, light can also be *absorbed* during interactions with electrons that are confined to exist in certain energy zones (Brillouin zones). Each zone is separated from adjacent zones by a gap of disallowed energies. For a certain range of incident light wavelengths, the energy of the incident light can match the energy gap between zones. In this situation, an atom can absorb the light, allowing an electron to jump to a higher energy zone. This absorbed energy is typically converted into heat and, thus, does not appear in the reflected wave. The important point is that there is a critical wavelength λ_0 such that incident light energy is not absorbed for $\lambda > \lambda_0$.

For most metals, there is no absorption of light in the visible wavelength range. The most notable exceptions to this rule are copper and gold. The spectral reflectance functions for both copper and gold are near 1 for long visible wavelengths (towards red) and decrease sharply below their critical wavelengths λ_0. Copper and gold absorb blue and green light strongly while reflecting red light. This gives these metals their reddish colors. On the contrary, silver begins absorbing in the ultraviolet at approximately $\lambda_0 = 310$ nm [88].

2.2.3 Interaction of light with dielectrics

Dielectrics differ significantly from metals in that they are poor conductors of electricity. Physically this is because all Brillouin zones that contain electrons are full and there exist large energy gaps separating these full zones from adjacent empty zones. Hence, an applied electric field does not produce a current, and the material behaves as an insulator. An important consequence of having a small conductivity is that K_0 is small and incident light is able to penetrate a significant distance into the body of the dielectric material. In some cases, a fraction of the incident light that penetrates reaches the other end of the material, producing transparency; in other cases, all the light is absorbed or scattered inside the material, producing opacity.

In general, when light strikes a surface, it must first pass through the interface between the air and the surface medium. Because the medium refraction index differs from that of the air, some of the light is reflected at the interface producing *interface reflection*. In general, such reflection is assumed to be in the "perfect specular direction" relative to the local surface normal. Most materials are, however, "optically rough", with local surface normals that differ from the macroscopic surface normal, so that the interface reflection is somewhat scattered at the macroscopic level, giving rise to a high brightness peak rather than an ideal impulse.

The amount of light reflected at the interface is governed by Fresnel's laws, which relate interface reflectance to the angle of incidence (relative to the local surface normal), the index of refraction of the material, and the polarization of the incoming illumination. The *interface reflectance* is a function of wavelength since the index of refraction generally depends on it, although the amount of reflection is typically constant to within a few percent across the visible spectrum; acrylic plastic, for example, has an index of refraction of 1.485 at 400 nm and 1.505 at 700 nm [113]. Hence, interface reflection is frequently assumed to be constant with respect to wavelength, and is thus said to have the same color

as the illuminant. By contrast, the effect of polarization is more severe, and interface reflection tends to be highly polarized, especially for large angles of incidence.

The light that penetrates through the interface passes through the medium, where it undergoes scattering from the colorant (*internal scattering*), and eventually is either transmitted through the material (if it is not opaque), absorbed by the colorant, or re-emitted through the same interface by which it entered, producing *body reflection*. The geometric distribution of this body reflection is sometimes assumed to be isotropic (i.e. independent of the viewing direction). The colour of the body reflection is generally different from that of the illumination, since interactions with colorant particles result in absorption with a probability dependent on wavelength. Body reflection is usually considered to be unpolarized.

2.3 Reflection Models

While the *Lambertian* model is accepted as a sufficient approximation of body reflection most times, the complexity of the interface reflection has given rise to the development of several models, differing among them in the physical approach followed or in the assumptions about the surface geometry at the microscopic level. In the following sections, mathematical models for both cases of reflection are revised, first in isolation and, next, as a part of general models of light-matter interaction.

2.3.1 The Reflectance Map

The apparent brightness of a surface patch depends on the orientation of the patch relative to the viewer and the light sources. Different surface elements of a non-planar object reflect different amounts of light toward an observer as a consequence of their differing attitude in space. A smooth opaque object gives rise to a shaded image, one in which brightness varies spatially, even though the object is evenly illuminated and covered by a uniform surface layer.

A convenient representation for this spatial variation is the so-called *reflectance map*, $R(p, q)$, a concept introduced by Horn in [101] which models scene radiance as a function of the surface normal $\boldsymbol{N} = (p, q, -1)$. In this representation, the normal to a surface $\boldsymbol{z}(x, y)$ is calculated through the cross-product of the tangents at (x, y) on the X and Y directions: $(\Delta x, 0, Z_x \Delta x) \equiv (1, 0, Z_x)$ and $(0, \Delta y, Z_y \Delta y) \equiv (0, 1, Z_y)$, respectively. Then, $(1, 0, Z_x) \times (0, 1, Z_y) = (p = Z_x, q = Z_y, -1)$ is the surface normal at (x, y, z). The $P - Q$ plane is called the *gradient space*.

Usually, the reflectance map is depicted as a series of contours of constant scene radiance, as indicated in figure 2.4. In the figure, the left map corresponds to a surface showing only body reflection, while the right map corresponds to a surface showing both body and surface reflection. In the right map, two peaks can be appreciated, corresponding to the surface orientations that maximize each of the two different types of reflection.

There are different ways to measure the reflectance map: by using a goniometer, from the image of an object of known shape, or it may be calculated if properties of the surface material and the distribution of light sources are given.

2.3.2 Body reflection models

Lambert was the first to investigate the mechanism underlying diffuse reflection [137]. A surface obeying Lambert's law, an *ideal Lambertian surface*, appears equally bright from all viewing directions, regardless of how it is irradiated, and reflects all incident light, absorbing none. From this definition, several facts can be deduced: (1) the reflected radiance is isotropic, that is, L is constant, with the same value for all directions (θ_r, ϕ_r); (2) the integral of reflected radiance over the hemisphere above the surface, i.e. the radiant exitance M, must equal the total irradiance E falling over the surface, since no light is absorbed; and (3) the BRDF $f(\theta_i, \phi_i; \theta_r, \phi_r)$ must be a constant for such a surface. The following reproduces the derivation of the reflection model for an ideal Lambertian surface by Horn [102].

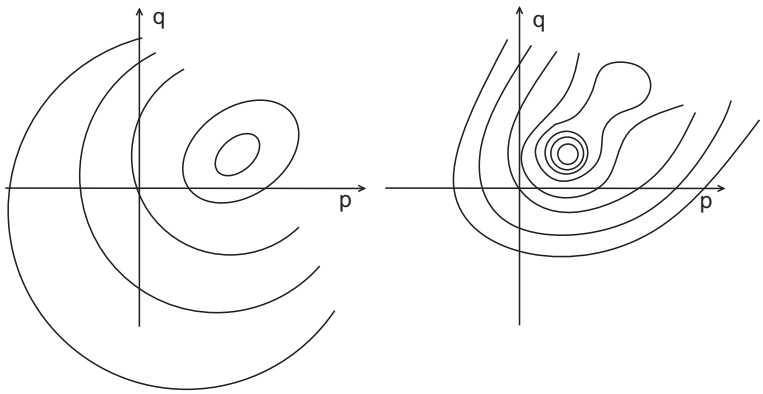


Fig. 2.4. Examples of reflectance maps.

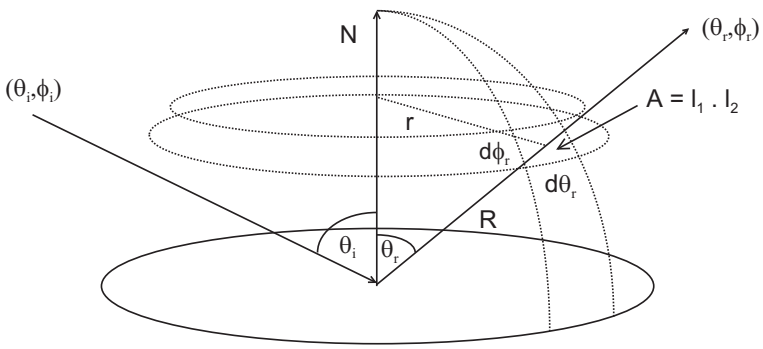


Fig. 2.5. Derivation of total radiance from a Lambertian reflector.

On the one hand, the radiant exitance M can be found by integration:

$$M = \int_{\Omega_r} L \mathrm{d}\Omega_r \,, \tag{2.11}$$

where Ω denotes projected solid angle. Since $\mathrm{d}\Omega_r = \cos\theta_r \mathrm{d}\omega_r$:

$$M = \int_{\omega_r} L \cos\theta_r \mathrm{d}\omega_r \,, \tag{2.12}$$

Now, from figure 2.5, $\mathrm{d}\omega_r = A/R^2 = l_1 \cdot l_2/R^2 = (r\mathrm{d}\theta_r)(R\mathrm{d}\phi_r)/R^2 = (R\sin\theta_r \mathrm{d}\theta_r)(R\mathrm{d}\phi_r)\ /R^2 = \sin\theta_r \mathrm{d}\theta_r \mathrm{d}\phi_r$. Hence:

$$M = \int_{-\pi}^{\pi} \int_{0}^{\pi/2} L \cos\theta_r \sin\theta_r \mathrm{d}\theta_r \mathrm{d}\phi_r = L\pi \,. \tag{2.13}$$

From equation 2.13, f is given as:

$$f(\theta_i, \phi_i; \theta_r, \phi_r) = \frac{L}{E} = \frac{L}{M} = \frac{1}{\pi} \,. \tag{2.14}$$

In this way, if a Lambertian surface is irradiated from different directions (θ_i, ϕ_i) the resultant radiance L_r is given by:

$$L_r = \int_{\omega_i} f \mathrm{d}E(\theta_i, \phi_i) = \frac{1}{\pi} \int_{\omega_i} L(\theta_i, \phi_i) \cos\theta_i \mathrm{d}\omega_i \,, \tag{2.15}$$

where $\cos\theta_i$ accounts for the foreshortening of the surface. Here, the whole set of values of L defines what is usually referred to as an *extended light source*. Equation 2.15 is a general form of *Lambert's Cosine Law*.

If, instead of an extended light source, the scene illumination reduces to a point light source at infinity, the whole scene can be assumed to be uniformly irradiated from just one direction, say (θ, ϕ), with radiance L_d. In such a case, from equation 2.15:

$$L_r = \frac{1}{\pi} L_d \cos\theta \,. \tag{2.16}$$

Equation 2.16 is a more popular form of the Lambert's law which is used in most of the physics-based vision methods including diffuse reflection in their image formation models.

If, in this latter case, vectors $\boldsymbol{N}$ and $\boldsymbol{S}$ are used for representing, respectively, local surface normal and direction (θ, ϕ):

$$L_r = \frac{1}{\pi} L_d \frac{\boldsymbol{N} \cdot \boldsymbol{S}}{\|\boldsymbol{N}\| \cdot \|\boldsymbol{S}\|} \,. \tag{2.17}$$

The term $\frac{\boldsymbol{N} \cdot \boldsymbol{S}}{\|\boldsymbol{N}\| \cdot \|\boldsymbol{S}\|}$ represents the dependence of scene radiance on surface orientation. As mentioned above, a usual notation for $\boldsymbol{N}$ is $(p = Z_x, q = Z_y, -1)$, where Z is the surface. Similarly, $\boldsymbol{S} = (p_s, q_s, -1)$ is another representation for direction (θ, ϕ). Then:

$$R(p, q) = \frac{\boldsymbol{N} \cdot \boldsymbol{S}}{\|\boldsymbol{N}\| \cdot \|\boldsymbol{S}\|} = \frac{p p_s + q q_s + 1}{\sqrt{p^2 + q^2 + 1}\sqrt{p_s^2 + q_s^2 + 1}} \,. \tag{2.18}$$

This function $R(p, q)$ is the aforementioned reflectance map and, as can be seen, makes explicit the relationship between surface orientation and brightness. It can be easily shown that the maximum of $R(p, q)$ in this case is at $(p = p_s, q = q_s)$; that is, when the surface normal aims at the light source. The minimum $(R(p, q) = 0)$ is attained when $(p, q, -1) \perp (p_s, q_s, -1)$; that is, when the surface normal lies perpendicular to the light source.

Real matte surfaces rarely behave as perfect diffusers in the sense that part of the incident light is absorbed and does not emerge through the interface by which it entered. In those cases, Lambert's law is complemented with an additional factor called the *body reflectance* or *diffuse albedo*, $\rho \in [0, 1]$, representing the amount of incident light which is not re-emitted. The incorporation of ρ to the original Lambert's law, together with the dependence of the different terms on wavelength λ can be found in equation 2.19:

$$L_r(\lambda) = \cos\theta L_d(\lambda) \rho(\lambda) \,. \tag{2.19}$$

Here, ρ has absorbed the term $1/\pi$. In this way, if two matte objects of the same shape and illuminated by the same light source appear one brighter than the other, it is because the darker possesses a lower diffuse albedo. The diffuse albedo becomes, thus, a property of the surface material.

Unfortunately, even for objects exhibiting an apparently matte surface finish, certain limitations exist for using the Lambertian model. In particular, for some surfaces, the diffuse reflection is not independent of the viewer direction (i.e. sand, rough plaster or rough sandpaper). Several improvements have been suggested to overcome these limitations (see [295] for a concise overview). Among all of them, two models have been separately developed recently. The diffuse reflection model proposed by Oren and Nayar [197] applies to rough surfaces whose local geometry is modeled as micro-facets distributed over various orientations and arranged in V-grooves. Assuming the diffuse reflection of each micro-facet to be Lambertian, the model consists of the combined reflection of these micro-facets subject to masking, shadowing and inter-reflection effects. An important consequence of the work of Oren and Nayar is that for very rough surfaces, the dominant factors that determine radiance are the geometrical effects caused by surface roughness and not the precise local diffuse reflectance characteristics. However, when the surface is relatively smooth, the geometrical effects turn out to be negligible and the exact model of local diffuse reflectance plays a critical role. This problem of accurately modeling local diffuse reflection from smooth surfaces is addressed by Wolff in [294]. His model consists in the combination of an isotropical sub-surface light scattering distribution produced from internal inhomogeneities coupled with the refraction of externally incident and internally scattered light at the air-surface dielectric boundary.

In a common paper [295], Wolff, Nayar and Oren, apart from suggesting a unification of both models in order to accurately explain diffuse reflection with intermediate roughness, detail important deviations

from Lambertian behaviour in matte rough and smooth surfaces. Unfortunately, both models imply the introduction of additional parameters which are difficult to estimate for objects of unknown geometry. Because of this, they are seldom adopted into machine vision algorithms. The Lambertian model, however, can be found almost within every physics-based vision algorithm because of its simplicity. The knowledge of what physical conditions allow applying it to good approximation plays, thus, a central role. In this sense, the authors determine that it is accurate within a margin of about 5% when θ_r and θ_i are simultaneously less than $50°$.[2]

2.3.3 Interface reflection models

The other extreme of material reflectance properties is illustrated by an *ideal specular reflector*, which reflects all of the light arriving from direction (θ_i, ϕ_i) into direction $(\theta_i, \phi_i + \pi)$. From the definition of the BRDF:

$$\begin{aligned} L(\theta_r, \phi_r) &= \int_{\omega_i} f \mathrm{d}E(\theta_i, \phi_i) = \int_{\omega_i} f L(\theta_i, \phi_i) \cos\theta_i \mathrm{d}\omega_i \\ &= \int_{-\pi}^{\pi} \int_{0}^{\pi/2} f L(\theta_i, \phi_i) \cos\theta_i \sin\theta_i \mathrm{d}\theta_i \mathrm{d}\phi_i \,. \end{aligned} \tag{2.20}$$

Now, since $L(\theta_r, \phi_r) = L(\theta_r, \phi_r + \pi)$, f can be deduced to be as follows:

$$f = \frac{\delta(\theta_i - \theta_r)\delta(\phi_i - \phi_r - \pi)}{\sin\theta_i \cos\theta_i} \,. \tag{2.21}$$

This model assumes that the highlight is only a single point, but this is a highly idealized behaviour of specular reflection which only holds for a highly polished surface, such as a perfect mirror. In general, it is recognized that specular reflection consists of two components: the specular spike and the specular lobe. The specular spike is zero in all directions except for a very narrow range around the direction of macroscopic specular reflection. The specular lobe spreads around the direction of specular reflection and is caused by the roughness of the surface, which makes the microscopic specular direction differ from point to point. With the aim of capturing this complex behaviour as accurately as possible into one mathematical formulation, several models more or less based on the underlying optics principles have been developed. Two different ways have been followed so far for describing this type of reflection using a sound physical basis: physical or wave optics, which makes use of electromagnetic theory and Maxwell equations; and geometrical optics, characterized by the basic assumption of incoming light whose wavelength is small compared with the roughness of the material, which greatly simplifies the derivation and final formulation of the models, but also reduces its scope of application. Two general representatives of these approaches which fit experimental data well are the Beckmann-Spizzichino model [13], for physical optics, and the Torrance-Sparrow model [281], for geometrical optics. Simplifications of both reflection models are used in computer vision and computer graphics to describe the specular component.

On the one hand, the model by Torrance and Sparrow only describes the lobe component and assumes that a surface is composed of randomly oriented mirror-like micro-facets, each one comprising one side of a symmetric V-groove cavity, whose size is considered much larger than the wavelength of incident light. Only those micro-facets whose normal is half-way between the incident and the reflected beams contribute to the surface reflectance. Besides, the model incorporates the masking effect (i.e. adjacent facets obstruct flux incident upon a given facet) and the shadowing effect (i.e. adjacent facets obstruct the flux reflected by a given facet).

On the other hand, the Beckmann-Spizzichino model describes the reflection of plane electromagnetic waves from smooth and rough surfaces using basic concepts of electromagnetic wave theory. The model, which consists of the aforementioned specular spike and lobe components, results in a fairly

[2] None of the algorithms proposed in this thesis make use of the models by Oren/Nayar and Wolff in an explicit way. In this sense, it is considered that including their derivation and formulation is out of the scope of this section.

complicated function for surface radiance involving the angles of incidence and reflection and the surface roughness parameters. Among the different assumptions of the model, it is noteworthy the fact that masking and shadowing of surface points by adjacent points are not included in the formulation.

The Torrance-Sparrow model provides a good approximation of the specular lobe component of the Beckmann-Spizzichino model, although it is not capable of describing the mirror-like behaviour of smooth surfaces and, hence, should not be used to represent specular spikes. However, in view of its simple mathematical form and taking into account that both components are simultaneously significant for only a small range of roughness levels, the Torrance-Sparrow model results to be a good choice in most cases [180]. In this sense, by way of illustration, a modified version of the original model by Healey [88] for the case of a distant light source incident from direction (θ_i, ϕ_i) with radiance L_d is given in equation 2.22:

$$L_r(\lambda) = L_d(\lambda)\frac{F(\gamma,\lambda)P(\alpha)G(\theta_i,\theta_r,\alpha,\beta)}{\cos\theta_r}\,, \tag{2.22}$$

where:

- α is the angle between the unit surface normal $\boldsymbol{n}$ and the unit angular bisector $\boldsymbol{h}$ of (θ_i, ϕ_i) and (θ_r, ϕ_r) ($\boldsymbol{h} = (\boldsymbol{v}+\boldsymbol{s})/\|\boldsymbol{v}+\boldsymbol{s}\|$, $\boldsymbol{s}$ and $\boldsymbol{v}$ are, respectively, the unit vectors corresponding to (θ_i, ϕ_i) and (θ_r, ϕ_r));
- β is the angle between direction (θ_r, ϕ_r) and $\boldsymbol{h}$;
- γ is the angle between direction (θ_i, ϕ_i) and $\boldsymbol{h}$;
- $F(\gamma, \lambda)$ is the Fresnel term, representing the reflectance of a smooth surface, given by:

$$F(\gamma,\lambda) = 0.5(R_\perp + R_\parallel)\,, \tag{2.23}$$

  where:

$$R_\perp = \frac{a^2+b^2-2a\cos\gamma+\cos^2\gamma}{a^2+b^2+2a\cos\gamma+\cos^2\gamma}\,, \tag{2.23a}$$

$$R_\parallel = \frac{a^2+b^2-2a\cos\gamma\tan\gamma+\sin^2\gamma\tan^2\gamma}{a^2+b^2+2a\cos\gamma\tan\gamma+\sin^2\gamma\tan^2\gamma} \tag{2.23b}$$

  and

$$a = \sqrt{\frac{\sqrt{c^2+d^2}+c}{2}} \qquad b = \sqrt{\frac{\sqrt{c^2+d^2}-c}{2}} \tag{2.23c}$$

$$c = n^2(\lambda) - K_0^2(\lambda) - \sin^2\gamma \qquad d = 2n(\lambda)K_0(\lambda) \tag{2.23d}$$

  if unpolarized light is incident at an angle γ ($R_\perp$ is the component polarized perpendicular to the plane of incidence, while $R_\parallel$ is the component polarized parallel to the plane of incidence, the plane of incidence is the one containing the surface normal and the incident beam, and $(n - K_0 i)(\lambda)$ is the complex index of refraction);
- P is a probability distribution function for the orientation of micro-facet normals, which accounts for the number of micro-facets effectively contributing to the reflected light and is given by:

$$P(\alpha) = \frac{1}{\overline{m}^2\cos^4\alpha}\mathrm{e}^{-\frac{\tan^2\alpha}{\overline{m}^2}}\,, \tag{2.24}$$

  where $\overline{m}$ is the *surface roughness* parameter, defined to be the *root-mean-square* slope of micro-facet normals;
- G is a geometrical attenuation factor that incorporates the shadowing and masking of surface micro-facets by adjacent facets:

$$G(\theta_i,\theta_r,\alpha,\beta) = \min\left(1, \frac{2\cos\alpha\cos\theta_r}{\cos\beta}, \frac{2\cos\alpha\cos\theta_i}{\cos\beta}\right)\,. \tag{2.25}$$

In this model, the Fresnel term is analogous to ρ in equation 2.19; that is to say, it expresses the amount of incident light which is not absorbed by the surface, although it is referred to the ideal case of smooth surfaces.

A further simplification of the Torrance-Sparrow model can be obtained taking into account that for most surfaces the Fresnel term is almost 1 [280] and that the effect of masking and shadowing is almost negligible most times, so that factors F and G are removed from equation 2.22 to obtain a simpler model for specularities [92].

To finish this section, it is worth mentioning the specular part of the Phong model, which became quite popular in the computer graphics research community because of its simplicity and surprisingly remarkable realism [228]. This model predicts specularities which extend beyond an only point by means of an exponential expression involving the angle between the perfect specular direction and the viewing direction, φ:

$$L_r(\lambda) = L_d(\lambda) \cos^m \varphi \,, \tag{2.26}$$

where m determines the roughness of the surface, so that the larger m, the smoother the surface. Despite its popularity, this model is founded on its aesthetic results rather than on a sound physical basis.

2.3.4 Hybrid models

Real surfaces rarely are purely Lambertian or purely specular. They rather reflect light as a combination of both body and interface reflection by means of an additive composition. One straightforward equation for representing the final radiance is the following:

$$L(p;\lambda) = K_b(p)L_b(p;\lambda) + K_i(p)L_i(p;\lambda) \,, \tag{2.27}$$

where p is a scene point, L_b and L_i are, respectively, the body and interface reflection components, and K_b and K_i are parameters of the surface material which determine the relevance of the corresponding component in the final radiance. Using this general structure for final radiance, several models have been proposed particularizing mainly on the expressions for L_b and L_i (see [180,228,281] among others).

Nevertheless, the by far most referenced and used hybrid reflection model is the *Dichromatic Reflection Model* (DRM) introduced by Shafer [252]. The DRM was originally conceived for inhomogeneous dielectric materials and matches the structure of equation 2.27. In fact, the original formulation of the DRM does not specify the diffuse and the specular reflection components explicitly, and most of the work done with it neither makes them explicit. The DRM may be summarized by equation 2.28 [3]:

$$L(p;\lambda) = \overbrace{m_b(p)\left[L_d(\lambda)\rho_b(p;\lambda)\right]}^{L_b(p;\lambda)} + \overbrace{m_i(p)\left[L_d(\lambda)\rho_i(p;\lambda)\right]}^{L_i(p;\lambda)} \,. \tag{2.28}$$

With a reasonable degree of accuracy, each component L_j ($j \in \{b, i\}$) can be modeled as the product of two terms: $C_j(p;\lambda) = L_d(\lambda)\rho_j(p;\lambda)$, dependent only on wavelength and not on local geometry, and expressing the fraction of the incoming (directional) light $L_d(\lambda)$ which is conveyed by that reflection component due to the material reflectance $0 \leq \rho_j(p;\lambda) \leq 1$; and $m_j(p)$, which is a geometrical factor depending on the surface geometry at point p, and independent of wavelength. According to the author, the separation of the interface reflection component into a geometrical and a spectral term has an effect no greater than 2% in pixel-value errors, and this value is attained only if the angle of incidence of the light is larger than 60° and the viewing angle is nearly the perfect specular direction. As for the body reflection component, only if $L_d(p;\lambda)\rho_i(p;\lambda)$ is not constant there is an interdependence between the geometry and the spectral factors; since it generally does not vary much with wavelength, the effect of this interdependence should be negligible.

[3] Sometimes position p in equation 2.28 is replaced by another variable $\theta = (i, e, g)$ referring to the so-called *photometric angles*: i, the angle between the surface normal and the direction towards the light source or incident direction, known as the *angle of incidence*; e, the angle between the surface normal and the direction towards the viewer, known as the *angle of exitance*; and g, the angle between the incident direction and the viewer direction, known as the *phase angle*.

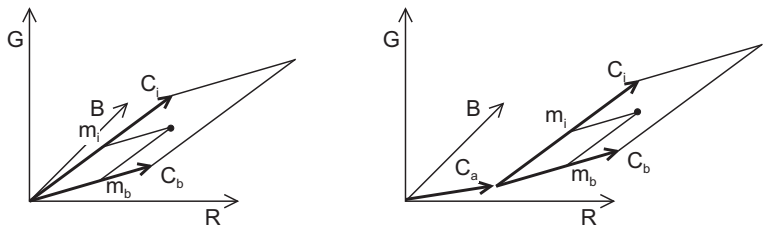


Fig. 2.6. The dichromatic reflection model in RGB colour space.

This model predicts that the pixels corresponding to the same uniformly colored surface will lie on a hyperparallelogram in color space. In effect, because the geometry is different at each point, the geometrical factors m_i and m_b vary from point to point. However, vectors $\boldsymbol{C_i}$ and $\boldsymbol{C_b}$, resultant from the discretization of wavelength in a set of colour bands, are the same at all points for the same surface material, because they do not vary with geometry. In other words, the pixel values are a linear combination of $\boldsymbol{C_i}$ and $\boldsymbol{C_b}$ with the coefficients determined by m_i and m_b at each point (see figure 2.6(left)).

In a subsequent paper [88], Healey shows that:

a) using the Reichman body-scattering model[4], the DRM is a reasonable approximation for a large class of inhomogeneous dielectrics; and
b) from the Torrance-Sparrow surface-reflection model and the Fresnel equations, a simple model of the form

$$L(p;\lambda) = \overbrace{m_i(p)\left[L_d(\lambda)\rho_i(p;\lambda)\right]}^{L_i(p;\lambda)} \tag{2.29}$$

is an accurate color reflection model for metals under a single spectral composition of illumination. This model is called the *Unichromatic Reflection Model for metals* (URM). In this case, since only one spectral component is present in the model, pixels corresponding to the same uniformly coloured metal lie in a straight line (in figure 2.6(left), this line would correspond to the one containing vector $\boldsymbol{C_i}$).

With both models, Healey develops the *Approximate Color Reflectance Model* (ACRM):

$$L(p;\lambda) = \begin{cases} m_i(p)C_i(p;\lambda) & \text{metal} \\ m_i(p)C_i(p;\lambda) + m_b(p)C_b(p;\lambda) & \text{inhomogeneous dielectric} \end{cases} \tag{2.30}$$

In a contemporary paper, Tominaga and Wandell [280] test the DRM under the additional assumption that the interface component of the reflectance $\rho_i(\lambda)$ is a constant. This assumption is derived from the fact that, obtaining $\rho_i(\lambda)$ from Fresnel's law and considering that the refractive index at the surface is constant over the visible spectrum, $\rho_i(\lambda)$ must be independent of wavelength, and thus a constant, under these circumstances. Lee et al. [145] further investigate the constancy of $\rho_i(\lambda)$ and coin the term *Neutral Interface Reflection* model (NIR) for this particularization of the DRM. The measurements presented in this paper show that $\rho_i(\lambda)$ is nearly constant for certain kinds of plastics, leaves, painted surfaces and fruits. Experiments also show, however, that $\rho_i(\lambda)$ is not approximately constant for other inhomogeneous materials such as coloured paper and some ceramics.

Finally, the DRM as presented above depends on the assumption that the illumination at any point comes from a single light source. A further improvement over the model can be made by incorporating an ambient lighting term $L_a(p;\lambda) = L_a(\lambda)\rho_a(p;\lambda)$, representing light coming from all directions in equal amounts, $L_a(\lambda)$, which increases the irradiance at scene surfaces independently of local surface

[4] The Reichman body-scattering model [235] is an extension of the previous Kubelka-Munk theory, that is a general mathematical treatment of scattering and absorption in colorant layers. The extension mainly consists of removing certain non-realistic assumptions. Experiments have shown that Reichman's model accurately predicts the reflecting properties of real materials.

geometry [252]. As for ρ_a, it is generally represented as a linear combination of the body and interface reflectances [58, 252]. After adding the ambient term, the DRM turns out to be:

$$L(p;\lambda) = \overbrace{L_a(\lambda)\rho_a(p;\lambda)}^{L_a(p;\lambda)} + \overbrace{m_b(p)\left[L_d(\lambda)\rho_b(p;\lambda)\right]}^{L_b(p;\lambda)} + \overbrace{m_i(p)\left[L_d(\lambda)\rho_i(p;\lambda)\right]}^{L_i(p;\lambda)} . \tag{2.31}$$

This model is a better approximation of daylight, that contains directional light from a point source (L_d terms), the sun, and ambient light from the sky (L_a term), and of light in a room, which comes from (directional) light fixtures (L_d terms) and from inter-reflections on walls and other objects (L_a term). Since $L_a(p;\lambda)$ does not vary with geometry, the effect of this change is to translate the hyperparallelogram of pixel values from a single surface away from the origin by vector $\boldsymbol{C_a}$ (see figure 2.6(right)). It is interesting to note that a point whose colour lies exactly at $\boldsymbol{C} = \boldsymbol{C_a}$ has $m_i = m_b = 0$, and might therefore be suspected of lying within a shadow.

2.3.5 Inter-reflections

In the previous sections, the image formation process has reduced to the case where scene surfaces received light from some identifiable light sources. This model is, however, inaccurate. Actually, each surface also receives light that is reflected by other objects in the scene. This phenomenon is known as *inter-reflection* and should be recognized and accounted for if a very accurate analysis of the image is needed.

The extension of the DRM to include ambient lighting is in fact a form of taking into account inter-reflections into closed spaces such as a room. In that case, it is assumed that the repeated light reflections within the enclosed space give rise to a constant light level of less power than the light sources illuminating the scene. Such new "illuminant" would be characterized by its own SPD resultant from the combination of the multiple reflections on the objects and walls of the scene.

In a more detailed analysis, inter-reflections can be modeled by extension of the DRM [253]. If light bounces from surface A to surface B, then the illumination on B is determined by the direct lighting and the light produced by both the body reflection and the interface reflection coming from A. These three colours all illuminate B, so the reflected light from B has six components:

1. light from the light source directly to B, reflected as interface reflection from B;
2. light from the light source directly to B, reflected as body reflection from B;
3. light from the light source to A, reflected as interface reflection from A onto B, then reflected as interface reflection from B;
4. light from the light source to A, reflected as body reflection from A onto B, then reflected as interface reflection from B;
5. light from the light source to A, reflected as interface reflection from A onto B, then reflected as body reflection from B; and
6. light from the light source to A, reflected as body reflection from A onto B, then reflected as body reflection from B.

The first two components are the components of the DRM for surface B, while the last four correspond to the inter-reflections. If interface reflection is assumed to have the same colour as the illuminant, the six previous components can be reduced to four colours.

The previous analysis just considers two mutually reflecting surfaces at a time. Dealing with inter-reflections in the general case can clearly be a formidable task. It is because of this that inter-reflections are ignored if their effects are not as relevant so as to prevent machine vision algorithms from obtaining accurate enough results without taking them into account[5]. Otherwise, one can try to remove them from the image. For instance, Funt et al. propose a method capable of producing inter-reflection-free images in [56], restricting the domain to one-bounce inter-reflections between two matte, convex surfaces of uniform colour.

[5] Besides, the effect of mutual illumination can be greatly reduced if the scene is illuminated by a direct light source of enough radiance [56].

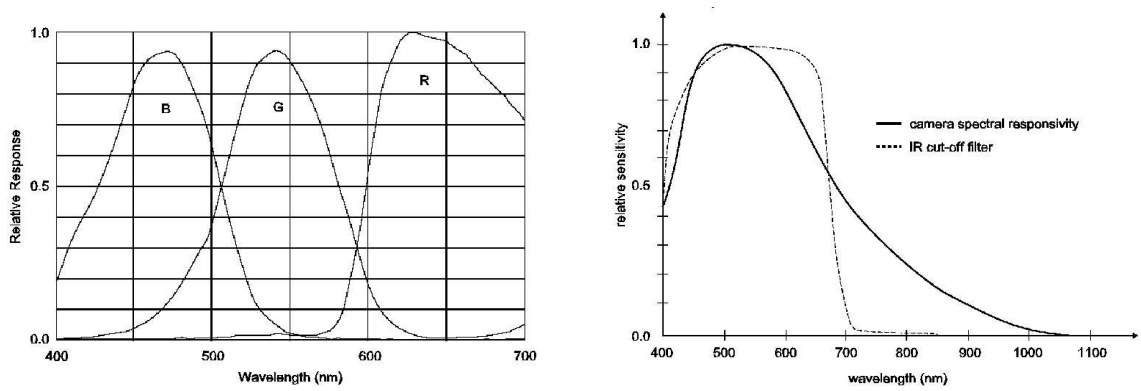


Fig. 2.7. Examples of filter transmittance and spectral responsivity curves: (a) filter transmittance for the red, green and blue channels for the SONY ICX075AK CCD image sensor [258]; (b) spectral responsivity of a typical CCD chip (solid line) and infrared cut filter (dotted line) [274].

2.4 Optical Image Sensors

The operation of image sensors at the radiometric level is reviewed in this section. On the one hand, section 2.4.1 presents a simple camera model independent of the imaging technology and generally accepted by the physics-based vision research community [119, 187]. Next, sections 2.4.2 and 2.4.3 overview the still leading imaging technology, Charge Coupled Devices (CCD), while section 2.4.4 reviews a specific model for the operation of CCD systems, based on a previous model by Healey and Kondepudy [93], to provide a detailed image formation expression including the different sources of noise affecting CCD's performance. Finally, section 2.4.5 briefly surveys other aspects of the operation of CDD cameras not included in the previous models.

2.4.1 General image sensing

The final pixel value given by a camera can be expressed as [187]:

$$D^c(i,j) = Q_0^c \int_{\Lambda} E(i,j;\lambda)\tau^c(\lambda)s(\lambda)\mathrm{d}\lambda\,, \tag{2.32}$$

where: $D^c(i,j)$ is the digital value of the c colour channel at image cell (i,j); Λ represents the set of wavelengths in the visible spectrum (approximately between 380 (blue) and 760 (red) nanometers); Q_0^c is a scaling factor which accounts for several elements influencing the whole image, such as the exposure time and the gain for the c colour channel; $E(i,j;\lambda)$ is the incoming light at image cell (i,j); $\tau^c(\lambda)$ is the filter transmittance for the c colour channel, expressing the fraction of light that the corresponding colour filter allows reaching the image cell at each wavelength (see figure 2.7(a) for an example); and $s(\lambda)$ is the spectral responsivity of the sensor, which represents the degree of sensitiveness of the sensor to the different wavelengths (see figure 2.7(b) for an example). In a typical RGB camera, $c \in \{r,g,b\}$, while in a monochrome camera, $\tau(\lambda) = 1$.

The term $E(i,j;\lambda)$ in equation 2.32 represents the radiation coming from the set of scene points which are optically reachable from image cell (i,j). From a theoretical point of view, the relation between $E(i,j;\lambda)$ and the radiance $L(p;\lambda)$, originating at the scene points p optically reachable from (i,j), involves the interaction of L with the propagation medium together with the effects of the camera optics on the arriving light through its *point-spread function* and the *lens collection* capability [85]. Assuming a non-attenuating propagation medium and ignoring the blurring and low-pass filtering effects of the point-spread function of the optics in a properly focused camera, the required relation appears in a work by Horn and Sjoberg [104]. For the sake of completeness, their analysis is reproduced in the following, removing co-ordinates i, j and p to simplify the notation when referring to E and L.

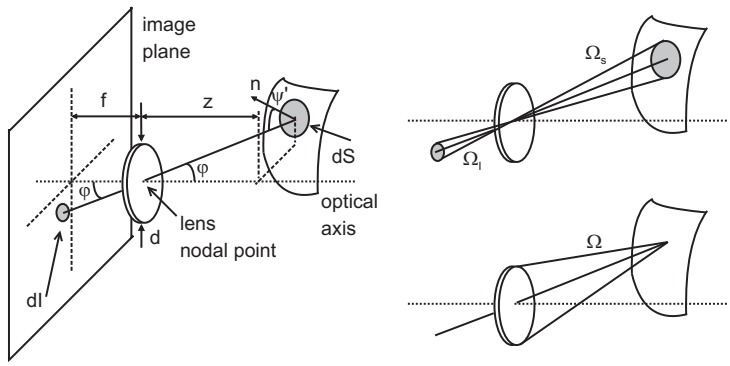


Fig. 2.8. Light collected by the lens from the surface patch of area dS is projected into the image patch of area dI.

Considering a small scene patch dS and the corresponding image patch dI (see figure 2.8), the irradiance E falling over dI is given as the radiant flux dΦ coming from dS and passing through the lens aperture per unit area, $E = \mathrm{d}\Phi/\mathrm{d}I$. On the other hand, the radiance L at dS in a certain cone of directions dΩ is the flux radiated by dS, dΦ, per unit area and projected solid angle, $L = \mathrm{d}^2\Phi/(\mathrm{d}S\mathrm{d}\Omega)$. Therefore, $\mathrm{d}\Phi = \mathrm{d}S \int_\Omega L\mathrm{d}\Omega$, where Ω is the projected solid angle subtended by the lens aperture. Consequently:

$$E = \frac{\mathrm{d}S \int_\Omega L\mathrm{d}\Omega}{\mathrm{d}I} \,. \tag{2.33}$$

Now, if φ is the angle between the optical axis and the line connecting dI and the lens nodal point, and f is the focal distance, then the solid angle Ω_I subtended by dI is given by:

$$\Omega_I = \frac{\mathrm{d}I \cos\varphi}{(f/\cos\varphi)^2} \,. \tag{2.34}$$

If, on the other hand, ψ' is the angle between the surface normal at the scene patch dS and the line to the lens nodal point and z is the distance over the optical axis from the nodal point to dS, then the solid angle Ω_S subtended by dS as it is seen from the lens nodal point is given by:

$$\Omega_S = \frac{\mathrm{d}S \cos\psi'}{(z/\cos\varphi)^2} \,. \tag{2.35}$$

Since both solid angles are equal:

$$\mathrm{d}I = \frac{\cos\psi'}{\cos\varphi}\left(\frac{f}{z}\right)^2 \mathrm{d}S \,, \tag{2.36}$$

and, consequently:

$$E = \left(\frac{z}{f}\right)^2 \cos\varphi \int_\Omega \frac{L}{\cos\psi'}\mathrm{d}\Omega = \left(\frac{z}{f}\right)^2 \cos\varphi \int_\omega L\frac{\cos\psi}{\cos\psi'}\mathrm{d}\omega \,. \tag{2.37}$$

If the lens is small relative to the distance to the scene patch, $\psi' \approx \psi, \forall\psi$. Furthermore, if the radiance L is assumed constant over the set of directions ω, it can be removed from within the integral. Finally, if the diameter of the lens aperture is d, the solid angle occupied by the lens aperture $\int_\omega \mathrm{d}\omega$ is approximately equal to the foreshortened area $\pi(d/2)^2\cos\varphi$ divided by the square of the distance $z/\cos\varphi$. Therefore:

$$E = \left(\frac{\pi}{4}\right)\left(\frac{d}{f}\right)^2 (\cos\varphi)^4 L \,. \tag{2.38}$$

The quantity $\frac{f}{d}$ is the so-called *F-number*, by which optics aperture is configured in current cameras. In particular, notice that the higher the F-number, the more attenuated the irradiance E with respect

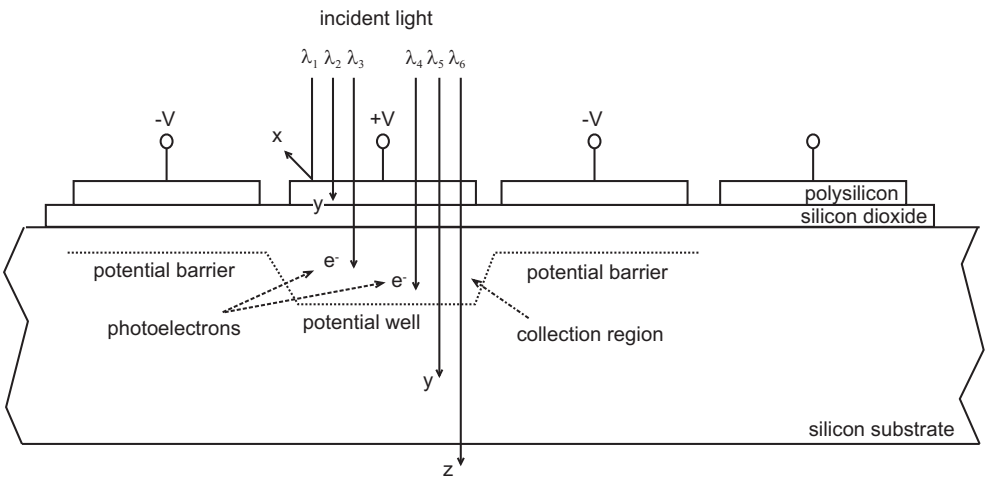


Fig. 2.9. CCD structure and operation: photons at different wavelengths λ_i strike the CCD cell and are reflected (x), recombined (y), pass through the material (z), or are absorbed (e^-) producing photoelectrons which are confined into the potential well created thanks to the voltages applied to the gate structures on top of the collection sites.

to scene radiance L. On the other hand, equation 2.38 reveals that the sensitivity of an imaging system is constant for a particular image point, but *a priori* not uniform throughout the image.

To finish, if $L(p;\lambda)$ is assumed constant for all the scene points p optically reachable from image cell (i,j) and equal to $L(i,j;\lambda)$, then, on the basis of equation 2.38:

$$E(i,j;\lambda) = \left(\frac{\pi}{4}\right)\left(\frac{d}{f}\right)^2 (\cos\varphi)^4 L(i,j;\lambda)\,. \tag{2.39}$$

2.4.2 CCD basics

The Charge Coupled Device (CCD) was invented in the late 1960's by researchers at Bell Labs and experimentally verified in 1970 [5, 21, 109]. Although it was initially conceived as a new memory technology for computers, it soon became apparent that the CCD had many other potential applications, including signal processing and imaging, the latter because of silicon's light sensitivity for wavelengths below 1.1 μm, which includes, in particular, the visible spectrum [39]. Their light sensitive properties were quickly exploited for imaging applications and they produced a major revolution in Astronomy[6]. On the other hand, their high resolution, high and wide spectral responsivity, low noise, non-linearities below 0.1%, fast response, small size, low consumption and durability have made them the most prominent device used in machine vision [93]. In addition, since a CCD is based on fixed sensing elements of equal size, the device provides precise spatial quantization that enables accurate spatial representation of images in a computer.

An analogy devised by Morley Blouke of Tektronix and Jerome Kristian of Mt. Wilson Observatory [109] which is quite useful to understand the operation of a CCD compares it with a mechanism to measure the spatial distribution of rainfall over a field. The mechanism consists of an array of buckets which are placed over a grid of conveyor belts laid on the field, in such a way that, after a storm, the buckets are transferred to a metering station where the amount of water in each bucket is measured. In this way, each measurement represents the amount of rainfall at a particular location on the field.

In a CCD, light sensitive elements etched on a thin wafer of silicon play the role of buckets producing, in this case, a measure of the spatial distribution of incident light. The measurement process relies

[6] They improved the light gathering power of telescopes by almost two orders of magnitude. Nowadays an amateur astronomer with a CCD camera and a 15 cm telescope can collect as much light as an astronomer of the 1960s equipped with a photographic plate and a 1 m telescope [109].

on the photoelectric effect by which the energy associated with photons striking the CCD reacts with the silicon, exciting electrons from the valence band to the conduction band and producing electron-hole pairs. The photon-generated electrons (photoelectrons) are then collected in one of many discrete collection sites.

For every wavelength, the number of photoelectrons collected at each site is linearly dependent on the number of photons per unit time and per unit area striking the silicon [40]. Although, generally speaking, one can say every photon generates an electron-hole pair, there are deviations which degrade the efficiency of the charge generation process such that incomplete conversions of photons into electrons occur. This degradation is characterized by a figure of merit called *quantum efficiency*, η, usually introduced in the equations of image formation in the form of a spectral responsivity, $s(\lambda)$, to reflect its wavelength dependence (see equation 2.32). It is defined as the number of electrons created per incident photon at a given wavelength, and vary depending on the characteristics of the particular material in which the collection sites are formed, through factors such as its absorption coefficient or the recombination lifetime of photoelectrons, and also on the materials overlying the silicon[7]. In some cases it can exceed 100% (i.e. more than one electron created per incident photon), although, for silicon-based CCDs, the quantum efficiency is generally assumed to be limited to 100% at wavelengths above ∼400 nm. At wavelengths below this value (ultraviolet and deep blue bands), the photon energy is sufficiently large to excite multiple electrons and bring them to the conduction band, so that special UV coatings have to be deposited in order to avoid these electrons from contaminating the measurements. As for wavelengths in the near infrared band, commercial CCDs do not show null quantum efficiency, which leads to the incorporation of infrared cut filters in high-quality cameras (see figure 2.7(b)).

An electronic representation of the spatial distribution of the light incident on the CCD is finally obtained by integrating photoelectrons in the individual collection sites over a fixed time interval known as the *integration* or *exposure time*. During a process named *charge coupling*, charge packets at collection sites are next transferred from site to site by manipulating gate potentials by means of a number of charge transfer techniques which preserve the separation of individual charge packets[8]. The fraction of charge which can be effectively transferred between adjacent collection sites is called the *charge transfer efficiency* (CTE) of the device.

Taking as a basis the charge coupling process for moving charge packets between CCD cells, several architectures to read the whole image out of the device have been devised so far. *Full frame* (FF) CCDs have the simplest architecture and are the easiest to fabricate and operate (see figure 2.10). In this type of CCD, the image is read out of the device by transferring in parallel the charge packets at every row to the next row along vertical electron-conducting channels. Every one of these channels is connected to a cell of an output serial shift register, so that, after each parallel row transfer, a complete new row comes off the CCD and is stored into the register. From there, before the next parallel transfer, every charge packet is brought to a charge-to-voltage output amplifier that generates a signal proportional to the amount of charge[9]. The process continues until the entire two-dimensional array of collection sites has been read out. Because the vertical parallel shift register is used for both scene imaging and readout, a mechanical shutter must be used to preserve image integrity during the charge transfer process, which reduces the frame rate achievable by the device. Other types of CCD, such as the

[7] On the one hand, the *absorption coefficient* is related to how far into the material an incident photon will travel before being absorbed. If the absorption coefficient is too high or too low, the electron may be created in a region where it can not be eventually collected as usable signal, reducing thus the quantum efficiency.

On the other hand, the *recombination lifetime* is the time during which photon-induced electrons are mobile within the silicon before they transition back to the valence band. Longer lifetimes increase the probability of photoelectrons to be captured and read out from the CCD, thereby increasing the quantum efficiency.

Finally, the materials overlying the silicon are mainly added on top the CCD to build structures for reading the collected charge out of the device. Ideally, those materials should be transparent, but, depending on wavelength, they absorb or reflect light, decreasing the quantum efficiency.

[8] Namely, four phase (4Φ) CCDs, three phase (3Φ) CCDs, pseudo two phase (P2Φ) CCDs, true two phase (T2Φ) CCDs and virtual phase (VΦ) CCDs. See [39, 98] for a detailed description.

[9] Typically, 50,000 electrons lead to a video signal amplitude of $1V_{pp}$ at the output for an amplifier gain of 1 [274].

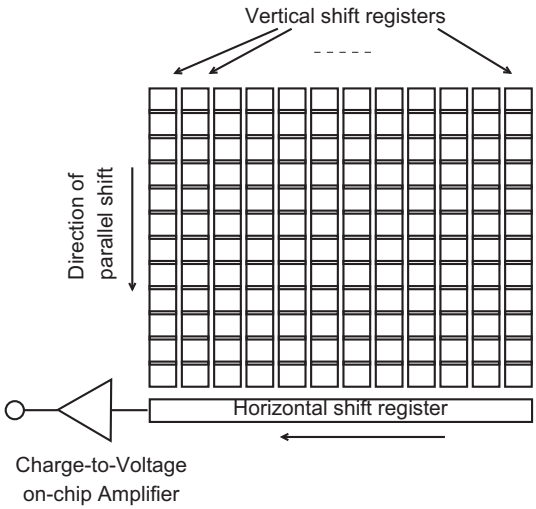


Fig. 2.10. Full-Frame CCD architecture.

frame transfer (FT) CCD, the *interline transfer* (IL) CCD, or the *frame interline transfer* (FIT) CCD overcome this problem by means of a separate array where the charge packets are quickly transferred to be read out in parallel with the next integration. See [39, 98] for a more detailed explanation.

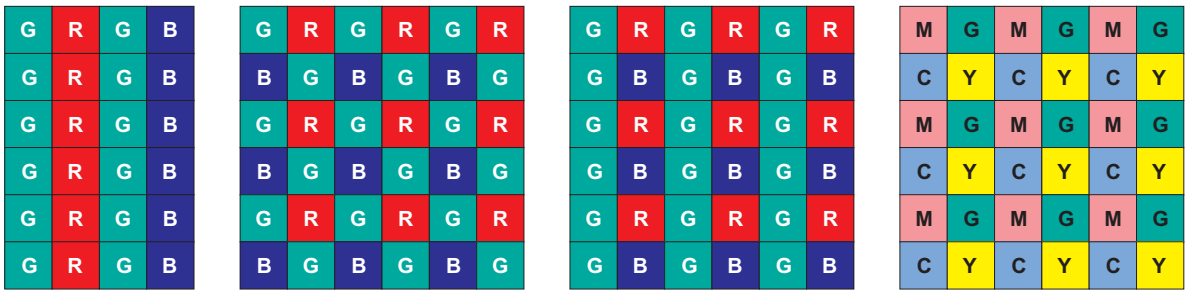
Finally, the signal produced by the output amplifier is transformed into an analog video signal conforming to any of the internationally accepted standards (NTSC, EIA, S-video, RGB, ...), or else into a digital signal suitable to be directly used in a digital system (RS-422, USB, IEEE 1394/Firewire/iLink, ...).

2.4.3 Colour image formation in CCD cameras

Since a light sensitive element just provides a measure of the light incident onto it, CCD sensors can be said to be monochrome in nature, and additional means have to be provided to quantify the spectral composition of the incoming radiation. A straightforward solution consists in placing a "wheel of filters" in front of a monochrome camera and taking several images of the scene exchanging the filters from exposure to exposure. In colour cameras, colour filters are also involved, in one form or another, in the operation of the sensor. Two main techniques are of use nowadays. On the one hand, the multi-CCD approach makes use of a *beam-splitter prism* which separates the incoming light in several colour bands and casts the filtered radiation onto a set of spatially separated CCD sensors, one for every colour band. In contrast, 1-CCD cameras incorporate an *array of colour filters* on top of the CCD, which specializes every collection cell for measuring radiation within a certain wavelength range. Due to its lower complexity, current colour cameras typically implement the 1-CCD approach, contrary to the multi-CCD solution, which has been relegated to high-quality expensive cameras due to the intricacy of their manufacturing.

The arrangement of the colour filters of single CCD systems typically follows a stripe or a mosaic pattern, or a mixture of both; the precise layout, however, varies from manufacturer to manufacturer on the basis of colour reproduction criteria [42, 98]. By way of example, see figures 2.11(a), (b) and (c), where mosaic (b) corresponds to the *Bayer colour filter*[10], implemented in the majority of current 1-CCD colour cameras. As can be observed, the arrays typically contain twice as many green filters than red and blue filters, to take into account the largest sensitivity of the human eye to the green wavelength range (centered around 560 nm). On the other hand, both types of patterns present differences regarding horizontal and vertical resolution for red, green and blue detectors. Further, the

[10] Named after his inventor, the Kodak engineer Bryce E. Bayer. In his original patent (US patent #3971065, 1976), he suggested that either of two color schemes could be employed for capturing multi-color information with a camera's sensor: RGB (Red-Green-Blue) or CMY (Cyan-Magenta-Yellow).



(a) (b) (c) (d)

Fig. 2.11. Examples of colour filter arrays: (a) RGB colour filter array, stripe pattern; (b) RGB colour filter array, mosaic pattern (Bayer); (c) RGB colour filter array, stripe/mosaic pattern; (d) CMYG colour filter array, mosaic pattern.

(a) (b)

Fig. 2.12. Example of demosaicking: (a) raw CCD data; (b) final image. (Reproduced from http://micro.magnet.fsu.edu/primer/digitalimaging/cmosimagesensors.html)

aforementioned configurations correspond to the so-called *primary colour mosaics*, based on primary colours red, green and blue, in front of *complementary colour mosaics*, which use opposite colours cyan, yellow and magenta and are implemented as 3-colour patterns, or else as 4-colour patterns, by adding green filters, as shown in figure 2.11(d). Among the advantages of using complementary mosaics are increased sensitivity resulting in improved light transmission through the filter, and a stronger signal [98]. This occurs because CMY filters exhibit a reduced absorption of light waves in the visible region when compared to the corresponding RGB filters. However, the use of CMY/CMYG mosaics implies adding a further step of conversion of the data collected by the sensor into RGB values.

After the capture of the image, the missing colour information between pixels in the raw image provided by the CCD must be filled in through a process known as *demosaicking* (see figure 2.12 for an example). A variety of sophisticated and well-stablished image processing algorithms are available to perform this task directly on the integrated circuit. In order to determine the correct colour for each pixel in the array, well-known interpolation techniques such as *nearest neighbour*, *linear*, *cubic* or *cubic spline* are applied using selected neighbouring pixels to produce separate estimates of chromaticity and luminosity for every pixel in the array. In this way, luminosity can take advantage of the full resolution of the CCD array and only chromaticity suffers from the loss of resolution due to the specialization of collection cells. The precise interpolation algorithm, however, determines the quality of the final image, reason by which many camera manufacturers consider their best color interpolation algorithms to be trade secrets and do not publish them [42]. In general, these techniques produce excellent results at image places where colour changes slowly over a large number of pixels, but can suffer from artifacts such as *aliasing* at edges and boundary regions where large colour and/or intensity transitions occur.

2.4.4 CCD cameras operation and noise modelling

Ideally, the number of electrons collected at a given cell for colour band c, I^c, is given by [93]:

$$I^c = T \int_{\Lambda} \left(\int_y \int_x E(x, y; \lambda) S_r(x, y) \eta(\lambda) \, \mathrm{d}x \, \mathrm{d}y \right) \tau^c(\lambda) \mathrm{d}\lambda \,, \tag{2.40}$$

where (x, y) are continuous coordinates on the sensor plane, T is the integration or exposure time, $E(x, y; \lambda)$ is the irradiance incident at point (x, y) over the collection site, $S_r(x, y)$ is the *spatial response* of the collection site, $\eta(\lambda)$ is the ratio of electrons collected per incident light energy (i.e. it is a sort of quantum efficiency), and λ, Λ and $\tau^c(\lambda)$ represent the same as in equation 2.32. The spectral responsivity $s(\lambda)$ of equation 2.32 would correspond to $\eta(\lambda)$ in equation 2.40 if $S_r(x, y) = 1$.

As well as any other sensor, CCD arrays are affected by noise, so that a neat measure of the actual irradiance spatial distribution, as indicated by equation 2.40, can rarely be obtained. Many papers and books have been written about the different sources of CCD noise, above all referred to application domains where the irradiance measure is critical, such as Astronomy. For computer vision, the most important consequence is that, apart from the random noise and contrarily to what is typically assumed, the CCD array does not really consist of equivalent imaging cells.

This noise corrupts the signal along the different steps which lead to the formation of the image —charge collection, charge transfer, and the conversion of charge into a measurable voltage— and comes either from processing errors during CCD fabrication and from the behaviour of the underlying electronics. Among the different sources of noise, the *shot noise*, the *dark current shot noise*, the *fixed pattern noise*, and the *photo-response non-uniformity noise* are generated in the CCD itself, while the *reset noise*, the *amplifier noise*, the $1/f$ *noise*, and the *quantization noise* occur during amplification and conversion of the analog signal to a digital output. All these noise terms are revised in the following and graphically illustrated in figure 2.13.

On the one hand, processing variations and photomask alignment errors during CCD fabrication can give rise to differences in detector size, spectral response and thickness in coatings which eventually cause small variations in quantum efficiency and charge collection volume from collection site to collection site [98]. As a consequence, even if a CCD is uniformly illuminated, these variations lead to a site-to-site non-uniformity in collected charge. This spatial irregularity is often referred to as *photo-response non-uniformity* (PRNU), although other authors use the term *fixed pattern noise* [93].

On the other hand, atoms in the silicon lattice vibrate and push each other due to their thermal energy. As the atoms bump into each other, they can excite electrons into the conduction band, even in the absence of light. The result is a pool of electrons known as *dark* or *thermal electrons*, whose rate of generation is termed *dark current*. The expected number of dark electrons generated is proportional to the integration time T, and is highly temperature dependent. Since dark electrons can be stored at collection sites, they become indistinguishable from photoelectrons, increasing, thus, the magnitude of the resulting signal[11]. Dark current is known to vary from site to site and this variation is called *fixed pattern noise* or *dark current non-uniformity* (DCNU). It is due to differences in, again, detector size, doping intensity and foreign matter getting trapped during fabrication [98]. Besides, dark electrons also accumulate at collection sites while the array is being read out, so that, in large arrays, far pixel measures can show a larger contamination by dark current than pixels close to the on-chip amplifier. This effect is generally known as *shading*. Although DCNU and PRNU are different, they are sometimes collectively called *pattern noise*.

A third source of noise stems from both the statistical nature of photon production and the way how incident light is measured. This can be understood more easily going back to the analogy of rain drops falling onto an array of buckets. Both rain drops and photons arrive discretely, independently and randomly and are described by Poissonian statistics. If the buckets are very small, some buckets may collect more drops than others. If the rain was let to fall long enough, all the buckets would measure the same value, but for short measurement times the spread in measured values is non-negligible. This

[11] In low-signal and/or long-exposure-time applications, cooling devices (e.g. thermoelectric coolers [98]) and specific dark-current-reduction strategies (e.g. multi-pinned phase technology [109]) must be employed to avoid dark current from completely contaminating the irradiance measurement.

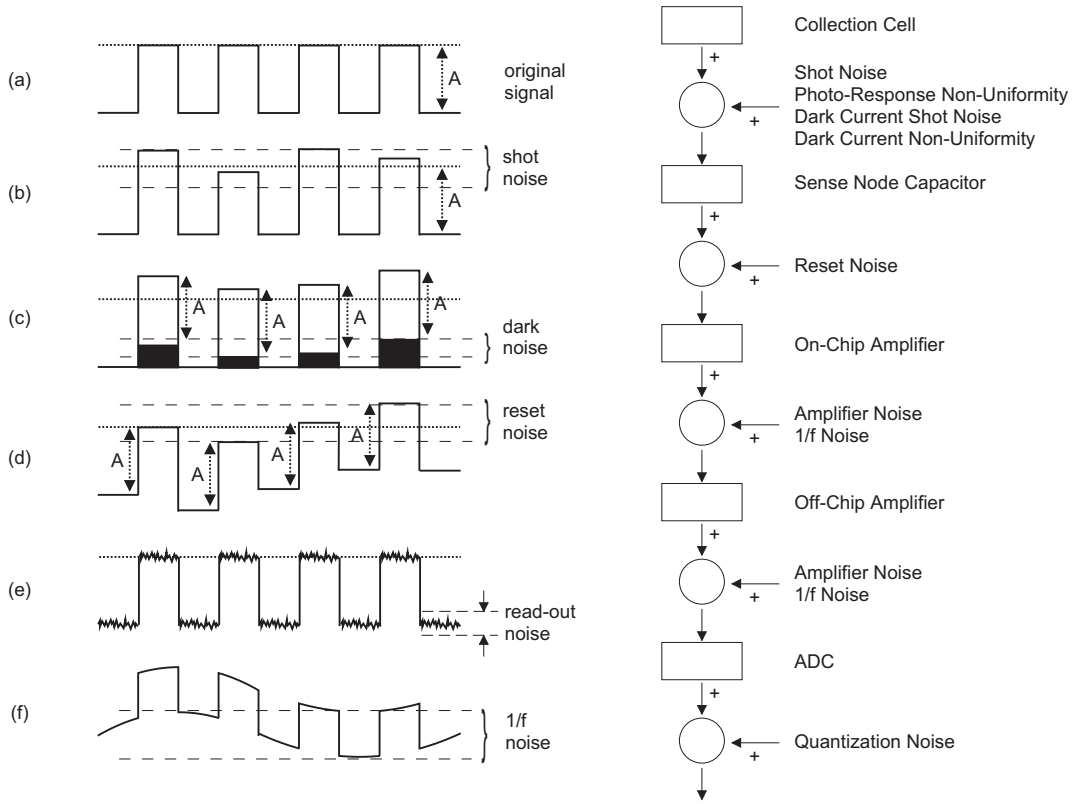


Fig. 2.13. Noise sources affecting CCD operation: [left] the array is viewing a uniform source that produces identical output of amplitude A for every pixel (all these processes occur simultaneously; 1/f noise has been magnified for illustration purposes); [right] noise transfer diagram.

effect also takes place in the temporal domain, so that, rarely the same number of photons is counted for a given pixel and two consecutive but independent observation intervals of the same length [109]. In the end, the probability of p photons incident on a CCD cell along an observation window of length T seconds is given by $P(p|\nu, T) = (\nu T)^p e^{-\nu T}/p!$, where ν is the photon production rate; from this, the variance of the number of incident photons equals the mean, i.e. νT. The generation of photoelectrons is accordingly perturbed, so that, if the signal produced is I^c, the corresponding standard deviation of the signal is given by $\sqrt{I^c}$. This noise is known as *photon* or *shot noise* and is always present in imaging systems. As has been seen, it is, in fact, the uncertainty in the data. Similarly to photon noise, the production of dark electrons also responds to Poissonian statistics, so that a further source of noise arises, the *dark current shot noise*, whose standard deviation equals the square-root of the number of thermal electrons generated within the image exposure time.

Writing all the aforementioned in a formal way, the total charge at a collection site can be expressed as the sum of two Poisson-distributed random variables $\mathcal{P}(K(i,j)I^c(i,j)) + \mathcal{P}(\mu_{dc}(i,j))$, or, separating constant and random terms [93]:

$$(K(i,j)I^c(i,j) + \mu_{dc}(i,j)) + N_S^c(i,j) + N_{dc}(i,j)\,, \tag{2.41}$$

where:

- Avoiding a certain dependence on wavelength, $K(i,j)$ is a constant associated with collection site (i,j) representing the joint effect of the product of $\eta(\lambda)$ and $S_r(x,y)$ (equation 2.40), particularized for cell (i,j). It accounts, thus, for the PRNU. Taking the corresponding scaling of the product of $\eta(\lambda)$ and $S_r(x,y)$ out from within the integral, a random variable K, representing the variation over $K(i,j)$ values, could be defined, being its (spatial) mean and variance $\mathrm{E}_I[K] = 1$ and $\mathrm{Var}_I[K]$, respectively.

- $\mu_{dc}(i,j)$ is the average dark charge for image cell (i,j). Its variation across the array would account for the DCNU.
- $N_S^c(i,j)$ and $N_{dc}(i,j)$ are zero-mean random variables corresponding to, respectively, photon and dark current shot noises. As discussed before, the variances of $N_S^c(i,j)$ and $N_{dc}(i,j)$ are, respectively, the expected number of photoelectrons generated for collection cell (i,j), $K(i,j)I^c(i,j)$, and the expected number of dark electrons also for collection cell (i,j), $\mu_{dc}(i,j)$.

After integration time, the CCD transfers the charge to the output amplifier for readout. As the charge packet moves from storage site to storage site, the larger the *charge transfer efficiency* (CTE) the fewer electrons are left behind. In, say, a 1000×1000 array, the charge packet farthest from the output amplifier must travel 1000 pixels along the corresponding vertical shift register and 1000 pixels more along the horizontal shift register (see figure 2.10). Furthermore, depending on the physical structure of every pixel, a number of transfers will be needed to pass the charge from one pixel to the next; in a four-phase device, the charge circulates through four wells, so that, following the example, a total of 8000 transfers would be involved in the worst case. Arrays for consumer applications typically have CTE greater than 0.9999 while for scientific devices it approaches 0.999999 [98]. Because of these high efficiencies, it is usual to omit the CTE from the noise models, although it should be included for very large arrays (e.g. $0.9999^{8000} = 0.4493$ while $0.999999^{8000} = 0.992$). In particular, in this model, it has been considered reasonable to assume that all the charge collected at each site is transferred to the output amplifier (i.e. CTE ≈ 1).

At the CCD output, the signal from the image sensor is converted from the charge domain to the voltage domain by means of a sense capacitor and a source-follower amplifier [98]. The different noise components involved in this process are briefly reviewed in the following [41]:

- On the one hand, the CCD sense capacitor has to be reset to some reference level before measuring each charge packet. Once more, thermal noise generated by the resistance of the resetting FET transistor makes the voltage in the capacitor uncertain after the reset, being the associated standard deviation given by $\sqrt{kTC}/q$, where k is the Boltzman's constant, T is the absolute temperature, C is the sense node capacitance and q is the electron charge. This noise is known as *reset noise* or, by obvious reasons, *kTC noise*.
- On the other hand, the amplifier introduces mainly two sources of noise, collectively named as *readout noise*. In the first place, the amplifier resistance causes thermal noise whose magnitude is independent of frequency. It is, therefore, white noise and is usually referred to as simply *amplifier noise*. In the second place, the *flicker noise*, also called *1/f noise*, is noise that has an approximately inverse dependence on frequency: the higher the frequency or pixel rate, the lower the noise. Its effects on the input signal are shown in figure 2.13(f). (See [109] for a more detailed discussion about amplifier related noises.)

Among these three sources of noise related with the output stage of the CCD, both the reset and the 1/f noise are drastically reduced by means of a further signal processing step known as *correlated double sampling* [98, 109]. This technique mainly consists in measuring the voltage difference in the signal outputting the on-chip amplifier, instead of using the voltage amplitude directly.

As a consequence of the operation of the pair sense-node/on-chip-amplifier and the previous comments, a further zero-mean noise factor independent of the number of collected electrons N_R, representing the amplifier noise, is included in the outputting signal after equation 2.41. As this signal is later transformed into a video signal suitable to be used by the subsequent processing steps, a combined gain $(A^c)'$, including the amplifier and the camera circuitry, appears as a multiplicative factor in the final expression for the voltage signal leaving the camera:

$$V^c(i,j) = \left(K(i,j)I^c(i,j) + \mu_{dc}(i,j) + N_S^c(i,j) + N_{dc}(i,j) + N_R\right)(A^c)' \,. \tag{2.42}$$

Finally, since the analog video signal must be converted to digital form to be used by a computer, a final noise source N_Q appears in the form of the quantization error. N_Q can be shown to be a zero-mean random variable with a uniform probability distribution over the range $[-\frac{1}{2}, \frac{1}{2}]$ and variance 1/12 [98]. In this way, the final digital value D^c is given by:

$$D^c(i,j) = (K(i,j)I^c(i,j) + \mu_{dc}(i,j) + N_S^c(i,j) + N_{dc}(i,j) + N_R)\, A^c + N_Q\,, \tag{2.43}$$

where A^c is the gain which transforms charge into digital intensity levels (i.e. a scaled version of $(A^c)'$).

Now, the digitized signal corresponding to pixel (i,j) can be stated as a random variable $D^c(i,j) = \mu^c(i,j) + N^c(i,j)$:

$$\begin{aligned} D^c(i,j) = {} & \overbrace{(K(i,j)I^c(i,j) + \mu_{dc}(i,j))\, A^c}^{\mu^c(i,j)} \\ & + \overbrace{\underbrace{(N_S^c(i,j) + N_{dc}(i,j))\, A^c}_{N_e^c(i,j)} + \underbrace{N_R A^c + N_Q}_{N_f^c}}^{N^c(i,j)}\,, \end{aligned} \tag{2.44}$$

where N_e^c depends on the number of electrons effectively stored at image cell (i,j), while N_f^c does not.

2.4.5 Other aspects of CCD camera operation

When converting light into an electrical signal, the performance of real imaging sensors can be altered by several other factors not yet taken into account that can prevent from getting the theoretical digital value. This unexpected behaviour can be either caused by the CCD chip and the associated electronics or due to the camera optics. Although this section deals with the former and the latter is considered out of the scope of this work since it is the subject of study of *geometric camera calibration*, the sensitivity fall-of due to the camera optics does affect the radiometric performance of the whole imaging system. Correspondingly, it will also be discussed.

A number of these anomalous behaviours is reported in the physics-based vision literature [39, 82, 119, 188, 270]. The most representative ones are briefly discussed below:

- As equation 2.38 shows, the camera optics does not strictly give rise to a uniform brightening of the CCD. The maximum irradiance is achieved at the sensor center, while the brightness decrease is circular and function of the distance to this center. It can be easily deduced that the irradiance ratio between a given pixel at (i,j) and the pixel at the image center (i_0, j_0), given the angle φ from equation 2.38 for pixel at (i,j), is:

$$\frac{E(i,j;\lambda)}{E(i_0,j_0;\lambda)} = \cos^4 \varphi(i,j)\,. \tag{2.45}$$

  Of course, this model is valid just for a very simple imaging system with an only thin lens. In the case of a real camera optics, where compound lenses are used, the problem gets worse. It is because, roughly speaking, some of the rays that pass through the first lens may be occluded by portions of the second lens, and so on. This depends on the inclination of the entering ray with respect to the optical axis. Thus, points in the image away from the optical axis benefit less from the light-gathering power of the lens than does the point on the optical axis. This effect is called *vignetting* [102]. As in the case of the PRNU, this brightening variation can be modeled by a multiplicative constant altering the charge produced by the collection cell $I^c(i,j)$, so that, modifying accordingly equation 2.44, the final digital value produced by the camera would be:

$$D^c(i,j) = (K_1(i,j)K_2(i,j)I^c(i,j) + \mu_{dc}(i,j) + N_S^c(i,j) + N_{dc}(i,j) + N_R)\, A^c + N_Q(i,j)\,, \tag{2.46}$$

  where K_1 would correspond to the PRNU, while K_2 would represent the brightening variation due to the camera optics. Clearly, unless one is interested in separating both effects, from a practical point of view, K from equation 2.44 can be redefined as the product of K_1 and K_2, as in [269, 270]. Therefore, any estimation of K will include the joint effect of both sources of sensitivity reduction.
- Although CCD sensors present a high linearity, CCD cameras in general delinearize the output signal for display requirements. This is because the input voltage U of a cathode ray tube and the radiant intensity I emitted by the phosphorus have an exponential relationship $I \propto U^\gamma$, where γ

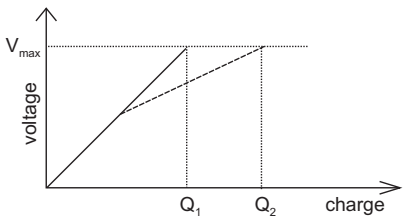


Fig. 2.14. Example of pre-kneeing: before activating the pre-kneeing circuit, a global linear relationship holds between collected charge and output voltage (solid line); after activating the pre-kneeing circuit, a piece-wise linear relationship (*knee-shaped*) is established between charge and voltage (dashed line). As a consequence, the input signal stays longer in the dynamic range of the camera (up to Q_2).

is the so-called *gamma value* and takes value typically between 2 and 3. In order to compensate for this non-linearity, camera makers introduce in commercial cameras a circuit for doing *gamma correction* by which the output voltage U_{out} is given as:

$$U_{out} = U_{max} \left(\frac{U_{in}}{U_{max}} \right)^{\frac{1}{\gamma}} . \tag{2.47}$$

Most scientific and industrial cameras allow switching the gamma value to 1 to produce a linear output. If this is not the case, *gamma re-correction* can be done by inverting equation 2.47 if γ is known. Otherwise, the output signal must be corrected using a calibration image containing a number of gray patches of known reflection properties, to calculate γ by regression or for building a linearization look-up table [188].

- Every collection site of a CCD sensor can store a limited amount of electrons. When this limit is exceeded, the additional charge overflows to neighbouring cells, causing them to report false light levels. This effect is known as *blooming* and makes manifest as a more or less extensive white blob around the affected pixels. Some CCD chips alleviate blooming by means of overflow drain structures built within the device, although at the cost of a lower dynamic range or reduced aperture [98]. Otherwise, special care must be taken when finding blobs of highest intensity in the image, usually corresponding to specularities of the scene due to the huge amount of light they convey.
- Although a collection site did not reach its charge saturation level, the resultant electrical signal can exceed the highest voltage the ADC can process. When this happens, *clipping* of the input signal occurs and important deviations in the hue of the pixel can arise because clipping is not equally applied to all the colour channels, just to the channels surpassing the maximum voltage accepted by the ADC. In order to increase the dynamic range of the ADC and reduce the probability the signal gets clipped, some cameras incorporate a so-called *pre-kneeing circuit* [119]. This circuit changes the slope of the linear relationship between collected charge and output voltage from a certain level of the input signal, with the net effect of a certain attenuation for higher levels of charge (see figure 2.14). As a result, the dynamic range of the camera is enlarged, although at the cost of an output signal which is no longer globally linear.
- It is commonly assumed that the integration performed in equations 2.32 and 2.40 is over the visible spectrum. However, the silicon that is used in CCD chips is particularly sensitive to the near infrared band, as it was already observed in section 2.4.2. In colour images, the additional light that enters the camera in the infrared band makes all the colours lighter and less saturated than humans expect to see them. High quality cameras usually incorporate *infrared cut-off filters* to force null quantum efficiency outside the visible spectrum (see figure 2.7(b)).

Among the different problems outlined above, blooming and clipping can be greatly alleviated, if not completely removed, by an adequate regulation of the camera aperture or the radiance of the light source. The others, as commented above, require more sophisticated methods of solution. Anyway, cameras specially designed to be used as measurement systems tend to incorporate means to avoid

such problems. Other cameras, however, are more oriented towards producing good-looking pictures and special care must be placed on their usage.

3

Review of Image Segmentation Approaches

This chapter summarizes the most significative work related with physics-based segmentation and edge detection. It will however start with a brief review on general approaches for image segmentation and edge detection in order to put physics-based algorithms into proper context. Furthermore, a taxonomy for the physics-based segmentation algorithms reviewed in this chapter will be introduced, to ease a discussion on their strong and weak points.

The rest of the chapter is organized as follows: first, section 3.1 reviews general approaches for image segmentation and edge detection; next, section 3.2 focuses on the roots of physics-based vision, particularizing on image segmentation; a state of the art in physics-based segmentation is presented in section 3.3; finally, a summary of this state of the art, together with a discussion about what has and has not been done in physics-based segmentation, is included in section 3.4.

3.1 Brief Review on General Image Segmentation

The inherent interest on the automatic processing of images has led to multitude of attempts to confer machines with the sense of sight since almost the first digital computers. Although the first theories on automatic processing of images are dated back in the fifties, the thesis by Roberts in 1965 [238] is generally accepted as the first research work related with the understanding of the three-dimensional outer world by means of computers [115]. Throughout this nearly forty years of research, image segmentation and edge detection are probably the vision tasks which have received most attention and effort due to their character of being fundamental preprocessing steps towards scene understanding in computer vision. As a consequence, an unaccountable amount of methods have been proposed during this period, what can be easily checked by the large amount of references (above 150 in many cases) which the numerous surveys published so far contain [26, 54, 84, 153, 183, 220, 237, 240, 244, 257, 260, 314].

Due to the well-known larger computational requirements of colour image processing, during an important part of this forty-year-period, most of the methods were devised for dealing with gray-level images, despite colour information was known to yield more reliable segmentations. Nowadays, however, this is no longer a major problem because of the continuous increase in computers power and speed, and the accompanying decrease in computation cost, together with the availability of relatively inexpensive colour cameras. Accordingly, there has been a remarkable growth in algorithms for colour image segmentation in this last decade, although on many occasions they are just generalizations of methods for gray-level segmentation and only in some cases the physics of interaction of light with coloured materials is taken into account [26, 153, 257].

Many vision textbooks distinguish between two approaches to tackle the segmentation of an image: region- or area-based methods, and contour- or edge-based methods (see [11, 44, 108, 116, 182, 288], among many others). In the region-based approach, all pixels that are homogeneous in some property are grouped together and are labelled to indicate that belong to the same region. Edge-based methods, however, look for pixels that lie on the region boundaries and, therefore, separate groups of pixels showing different values of a certain property. Similarity and continuity are thus the pixel qualities involved in, respectively, the region- and edge-based approaches. Other authors split the region-based

approach on the basis of whether spatial proximity in the image plane is taken into account or not in the process of pixel grouping, leading to a third class called pixel- or feature-space-based segmentation [153, 257, 260]. This third family embraces techniques where the resultant regions are required to correspond to "connected components" in the feature space but not over the image.

Moreover, each of these three categories is in turn subdivided in subclasses according to the computational techniques they usually make use of. More specifically, the region-based family comprises region-growing and split-and-merge strategies, the pixel-based approaches refer mainly to histogram analysis and clustering methods, and, finally, edge-based segmentation is subdivided according to the way how edges are found: either using global or local information, or by means of first-order or second-order derivative operators [257]; active contours and other variants are also considered in a different subfamily [153].

The different segmentation surveys published so far refine the above-mentioned categories, focus on some of them or add new ones according to the use or not of particular (mathematical) techniques. In order to introduce the main approaches utilized in one form or another for image segmentation throughout this forty-year period of computer vision, the following summarizes briefly and in chronological order the different taxonomies presented in some of those surveys:

- Rosenfeld and Davis [240] classify image segmentation techniques on the basis of the image model assumed for the particular technique to be applicable to it. Algorithms are initially catalogued according to the nature of the working model, namely statistical or spatial. Although colour and multi-spectral images are also considered, the study focus on gray-level images. Different statistical models are considered and, therefore, new categories appear: first-order models (i.e. gray-level histograms), nth-order models (i.e. co-occurrence matrices), models for local properties of image pixels (i.e. gray-level differences for pairs of points), random field models and time series models. Referring to the spatial models, Rosenfeld and Davies distinguish between methods based on: foreground/background models, image decomposition models (i.e. how is the image structured in terms of regions), the inter-region transitions and the shape of the regions with regard to either their interiors or their boundaries.
- Fu and Mui [54] discuss segmentation from the viewpoint of cytology image processing. They categorize segmentation techniques into three classes: (1) characteristic feature thresholding or clustering, (2) edge detection and (3) region extraction. The different techniques described are the following:
  - threshold selection schemes for gray-level and structural (texture) histograming;
  - clustering schemes, considered as the multidimensional extension of the concept of thresholding, based on multi-spectral information, mean/variation of gray-level, texture, colour, etc...;
  - edge detection, both parallel (the decision of whether a set of points are on an edge is dependent on the gray-level of the set and its neighbours) and sequential (the decision is taken according to the previously examined image points); a special emphasis is put on frequency filtering, gradient operators, adaptive local operators, functional approximations, heuristic search, dynamic programming, relaxation and line and curve fitting;
  - region merging, region splitting and combinations of both.
- Haralick and Shapiro [84] classify image segmentation methods as: (1) measurement space guided spatial clustering, (2) single linkage region growing, (3) hybrid linkage region growing, (4) centroid linkage region growing, (5) spatial clustering and (6) split and merge.
- Sahoo et al. [244] survey only segmentation algorithms based on thresholding. They categorize global thresholding techniques into two classes: point dependent techniques and region dependent techniques. Multi-thresholding techniques are also considered.
- Spirkovska [260] use the three-class classification mentioned above, pixel-based, edge-based and region-based schemes, and study some common image segmentation approaches such as Otsu's thresholding, Chow-Kaneko's adaptive thresholding, Yanowitz and Bruckstein's adaptive edge-based thresholding, Parker's local intensity gradient and Horowitz and Pavlidi's split, merge and group approaches.
- Pal and Pal [220] initially distinguish between two main approaches: the classical one, which involves histogram thresholding, edge detection, relaxation and semantic and syntactic methods, and the

fuzzy mathematical approach, also instantiated as methods for edge detection, thresholding and relaxation. After this first dichotomy, segmentation techniques are distributed among: (1) gray-level tresholding; (2) iterative pixel classification, in the form of relaxation, Markov Random Fields or Neural Networks; (3) surface-based segmentation, referring to range images; (4) color image segmentation; (5) edge detection; and (6) methods based on fuzzy sets, such as fuzzy thresholding, fuzzy clustering and fuzzy edge detection.

- Skarbek and Koschan [257] first identify four broad types of regions:
  1. *region* as a connected component in a feature space;
  2. *region* as a (maximal) connected set of pixels for which a uniformity condition is satisfied;
  3. *region* as a connected set of pixels bounded by edge pixels; and
  4. *region* as a surface or an object of homogeneous surface material.

  Both region definitions 1 and 2 use some sort of uniformity predicate, but in case 1 it is pixel-based and in case 2 it is area-based. Consequently with this four definitions, four more or less large groups of segmentation methods are distinguished: pixel-based (definition 1), area-based (definition 2), edge-based (definition 3) and physics-based (definition 4). Next, within the pixel-based category, up to three different subcategories are considered: histogram thresholding, clustering in colour space and fuzzy clustering in colour space. Subsequently, region splitting, region merging, region growing and split-and-merge techniques are distinguished among the area-based methods. As for edge-based segmentation, local techniques, using only information of the neighbourhood of pixels, are discussed against global techniques, making use of a sort of global optimization. Finally, physics-based methods are subdivided in approaches dealing exclusively with inhomogenoeus dielectrics and general approaches.
- Cheng et al. [26] distinguish up to six categories of colour image segmentation methods after a brief review of gray-level segmentation. The six categories proposed include histogram thresholding (mode method) and colour clustering, region-based approaches, edge detectors, fuzzy-sets based techniques, physics-based approaches and neural network approaches.
- Lucchese and Mitra [153] make a first distinction between feature-space based techniques, image-domain based techniques and physics-based techniques. Afterwards, on the one hand, the feature-space-based group is decomposed in clustering, adaptive k-means clustering and histogram tresholding. On the other hand, split-and-merge approaches, region-growing approaches, edge-based approaches and neural-network-based approaches are put in the image-domain-based group. Finally, no distinction is made within the physics-based group.

As can be observed in this brief review, the physics-based category appears only in the last four surveys [26,153,220,257] despite the work in physics-based vision started in the mid-1970s, many years before the publication of some of the other five surveys considered in this section.

3.2 The Roots of Physics-Based Segmentation

The image understanding problem has traditionally been solved by means of low-level segmentation or feature extraction processes and higher-level reasoning stages on top. The optimism surrounding symbolic reasoning and artificial intelligence during the 1970s relegated to low-level vision the task of dividing an image into (simple) regions on the basis of brightness and colour, leading to regarding the lower levels of processing as statistical image processing problems where the major concern were to determine statistically significant changes of pixel values under the presence of noise [273]. Obviously, the generated edge or region images generally outlined not only material boundaries, but also shadows, highlights and object edges. More extensive low-level processing was considered unnecessary since higher levels were assumed to be able to understand, identify, and merge these simple regions by combining them across highlights or shadows either by matching image features with object models or by determining the physical cause of edges or regions. Physical processes in the scene have thus not been a strong point of interest in the traditional line of computer vision research.

Early work in physics-based vision dates from the 1970s, when Berthold Horn at MIT introduced optical models of reflection and imaging to the vision community, and demonstrated that these optical

laws could actually be used as a source of information to aid in the process of automatic visual interpretation [100]. Theoretically, using Horn's model, some physical characteristics of a surface, including shape, could be estimated from a single image. However, few researchers adopted a similar approach to machine vision through the 1970s and early 1980s, because of, on the one hand, the important restrictions of the model of image formation (one perfectly diffuse, perfectly reflective surface, i.e. a Lambertian surface, under one point light source, without noise and perfect cameras), and, on the other hand, because Horn's work raised more questions that it answered [94]:

- how to deal with the inherent ambiguity that remains when optical laws are applied to an image;
- how to segment the image, that is, determine the regions with uniform optical properties;
- how to obtain reflection models for real materials, or whether there is perhaps a universal reflection model;
- how much calibration of cameras is required, and how to do it;
- how much knowledge about the scene is required, and how to get it in realistic scenarios.

Throughout this period of time, there was no sizable community of researchers in the area, because of skepticism about whether there could be any satisfactory solution to these issues. Little research was done in the area and it was typically tested only on relatively simple synthetic images generated under highly idealized assumptions.

In the 1980s, an explosion of research in this area began, leading to the adoption of the term *physics-based vision* to distinguish this area from more traditional approaches. In particular, Shafer's Dichromatic Reflection Model [252] allowed researchers to begin looking at a larger class of materials, inhomogeneous dielectrics, which was later complemented by the Unichromatic Reflection Model for metals by Healey [88]. Since then, there has been an active community of researchers in physics-based vision leading to an important amount of literature mainly in the areas of shape recovery and colour image understanding [94].

3.3 State of the Art in Physics-Based Segmentation

As has been said in the previous section, relatively little effort has been devoted to applying the physics of image formation to the segmentation of an image with respect to the effort invested in traditional segmentation. In general, physics-based segmentation strategies look for changes in objects reflectance, or equivalently surface material, either by direct or by indirect means. Nevertheless, an image encode many other properties of scene objects which, from a theoretical and physics-based point of view, could be used for segmentation purposes. They will be reviewed and discussed later, in chapter 4.

Among the several physics-based segmentation strategies which have been proposed so far, some of them are based on estimating directly the reflectance of the surfaces present in the scene, others look for certain configurations of clusters in colour space, as it is predicted by the Dichromatic Reflection Model, others are based on the continuity of the image surface when no material changes are involved, and, finally, others use photometric invariants in their different forms or work over specialized colour spaces. Although these algorithms can also be categorized according to any of the taxonomies mentioned in section 3.1, the classification proposed here makes more sense for physics-based vision algorithms since: (1) the way how the segmentation problem is addressed differs significantly among these classes, (2) important differences reside at the theoretical level on some occasions, and (3) the sort of techniques employed by every class are quite different to one another. Referring to this last point and by way of example, it is worthwhile noting that methods based on looking for certain configurations of clusters in the colour space make use of, mainly, *Principal Component Analysis (PCA)*, while methods related with photometric invariants are more prone to histograming and clustering techniques.

The following sections review an important part of physics-based segmentation and edge detection algorithms on the basis of the proposed classification. Among the different segmentation algorithms proposed in the literature, this survey restricts to those such that the paper where were published explicitly mentioned a model of image formation with a minimally sound physical basis. On the other hand, every algorithm is described by means of a sort of form, which contains, in this order:

(1) the researchers involved in the different papers where the algorithm appeared,
(2) the references to those papers together with the year of publication,
(3) the model of image formation assumed in the work, using the notation of the original papers conveniently explained, and
(4) a brief description of the algorithm.

Furthermore, each algorithm is assigned a *survey entry* with the following structure: 3.3.(PB type).xx, where *xx* is a correlative number and *(PB type)* is 1, 2, 3 or 4, for, respectively, reflectance-estimation-based algorithm, cluster-analysis-based algorithm, image-surface-continuity-based algorithm, and photometric-invariant-based algorithm. Finally, the list of algorithms surveyed is ordered in chronological order of the first publication related with the work.

3.3.1 REFLECTANCE ESTIMATION

3.3.1.1 Estimation of reflectance from three non-coplanar points of Lambertian surfaces

Auhor(s): Chia-Hoang LEE [140–142] and Azriel ROSENFELD [140–142]

Reference(s): [140] 1983, [141] 1985, [142] 1989

Image formation model:

$$I = (\lambda\varrho)\cos\theta\,,$$

where I is the brightness value found in an image pixel, λ represents the illumination intensity, ϱ is the reflectance and θ is the angle between the surface normal at the scene point considered and the light source direction. A distant point light source is assumed. The work only refers to gray-level images.

Description: The authors develop a method for estimating the product $(\lambda\varrho)$ at every pixel P using only the intensity information of P and two pixels Q and R on opposite sides of P in the direction of the gradient, provided they did not come from coplanar points in the scene. $(\lambda\varrho)$ is determined as the quotient between the intensity at P, I_P, and $\cos\theta_P$. The latter is proposed to be estimated, under the non-coplanarity condition, in the following way:

$$\tan\theta_P = \frac{I_R - I_P + 0.5\Delta^2 \cdot I_P}{\Delta \cdot I_P}\,,$$
$$\Delta^2 = \frac{2I_P - I_Q - I_R}{I_P}\,,$$
$$\Delta = \Delta_Q \approx \Delta_R\,, \Delta_Q = \theta_Q - \theta_P\,, \Delta_R = \theta_P - \theta_R\,.$$

To reach this closed-form solution, the authors make use of the approximation $\cos\Delta \approx 1 - 0.5\Delta^2$ which implies $\Delta = \Delta_Q \approx \Delta_R$ are assumed small. The values estimated for $(\lambda\varrho)$ are then collected in a histogram for the whole image. The peaks of such histogram are expected to correspond to the different materials in the scene. A median pre-filtering of the image is suggested to improve the accuracy of the estimates.

3.3.1.2 Estimation of reflectance from umbilical points on Lambertian surfaces

Auhor(s): Alex P. PENTLAND [224, 226]

Reference(s): [224] 1984, [226] 1989

Image formation model:

$$I = (\rho\lambda)\boldsymbol{n}\cdot\boldsymbol{l}\,,$$

where I is the brightness value found in an image pixel, λ represents the illumination intensity, ρ is the reflectance, $\boldsymbol{n}$ is the unit surface normal at the scene point considered and $\boldsymbol{l}$ is the direction towards the light source. A distant point light source is assumed. The work only refers to gray-level images.

Description: In this work, Pentland proved that, for umbilical surface points (points with equal principal surface curvatures), first and second image derivatives allowed estimating the composite body reflectance ($\rho\lambda$) for those points and provided the corresponding formulation:

$$(\rho\lambda)^2 = \left(-\frac{I_{xy}\eta^2 R^2}{\gamma} + I_x R\right)^2 + \left(-\frac{I_{xy}\eta^2 R^2}{\chi} + I_y R\right)^2 + \left(-\frac{I_{xy}\eta^3 R^2}{\chi\gamma}\right)^2 ,$$

$$R = \frac{\chi\gamma}{2I_{xy}\eta^2}\left((\chi I_x + \gamma I_y) \pm \sqrt{(\chi I_x + \gamma I_y)^2 - \frac{4DI_{xy}\eta^2}{\chi\gamma}}\right) ,$$

$$\chi = \cos\tau\sin\sigma\,, \quad \gamma = \sin\tau\sin\sigma\,, \quad \eta = \sqrt{1-\chi^2-\gamma^2}\,,$$

$$\tan\tau = \frac{-(I_{xx}-I_{yy}) \mp \sqrt{(I_{xx}-I_{yy})^2 + 4I_{xy}^2}}{2I_{xy}}\,,$$

$$\cos\sigma = \frac{\mp(I_{xx}+I_{yy}) - \sqrt{(I_{xx}-I_{yy})^2 + 4I_{xy}^2}}{2\sqrt{I_{xx}I_{yy} - I_{xy}^2}}\,,$$

where τ and σ are, respectively, the tilt and the slant angles of the direction of the illumination, $\boldsymbol{l}$, and I_x, I_y, I_{xx}, I_{yy} and I_{xy} are the first- and second-order image derivatives.
The papers referenced in this survey entry are not essentially about image segmentation but about shape from shading. However, the equations provided could be used, under the aforementioned conditions, to estimate objects reflectance at every image pixel and to build a histogram from which determine the materials present in the scene, in the same way as Lee and Rosenfeld's approach [140].

3.3.2 CLUSTER ANALYSIS

3.3.2.1 Analysis of clusters in colour space for colour image segmentation

Auhor(s): Gudrun KLINKER [120–131, 254, 255], Steven SHAFER [122–131, 254, 255], Takeo KANADE [122–131, 254, 255], Carol NOVAK [254, 255]

Reference(s): [122–124] 1987, [120, 125–128] 1988, [129, 254] 1990, [130, 131, 255] 1992, [121] 1993

Image formation model:

$$\boldsymbol{C} = m_b\boldsymbol{C}_b + m_s\boldsymbol{C}_s\,,$$

where $\boldsymbol{C}$ is the vector of camera responses for all the colour channels, m_b is the body geometrical factor, $\boldsymbol{C}_b$ is the vector of body reflection, m_s is the surface geometrical factor and $\boldsymbol{C}_s$ is the vector of surface reflection.
Furthermore:

- the NIR model: $C_s(\lambda) = c_s(\lambda)L_d(\lambda) = c_s L_d(\lambda)$, where L_d is the SPD of the illuminant,

is assumed.

Description: Klinker et al. present a segmentation method able to cope with body and interface reflection throughout this profuse series of papers. To this end, the shape of colour clusters in RGB space predicted by the dichromatic reflection model is scrupulously taken into consideration. Besides, camera limitations such as clipping or blooming are accounted for.

The algorithm starts dividing the image in square windows and determining the dimensionality of the corresponding colour clusters in RGB space by means of PCA. In this way, every cluster is classified as a point, a line, a plane or a volume. Next, compatible adjacent windows are put together since they actually correspond to the same hypothesis of image formation; adjacent windows are considered compatible if they are not volumes but are of the same type and, if combined, the result belong to the same class. This first fusion generates a first broad segmentation of the image. During the following step, linear hypotheses coming from areas of the image receiving a linear classification are exploited by means of a re-segmentation in which pixels closest in RGB space to the same hypothesis are put together. Dark pixels, since can belong to many hypotheses, are excluded from the process. In a subsequent step of the algorithm, planar hypotheses are generated, trying to connect linear clusters amongst them so that can constitute a dichromatic plane. The heuristic of the 50% is applied, by which the specular cluster is expected above the 50% of the matte cluster, while matte clusters are expected to converge at the origin of the RGB cube. The segmentation is refined again by using the resulting new hypotheses, adding pixels not segmented to existing regions.
Several images of plastic objects are used in the experiments.

3.3.2.2 Colour image segmentation in $S_0S_1S_2$ space

Auhor(s): Ruzena BAJCSY [7–10], Sang Wook LEE [7–10], Aleš LEONARDIS [7–10]

Reference(s): [7] 1989, [8] 1990, [9] 1992, [10] 1996

Image formation model:

$$I(\lambda) = E(\lambda)\left(g_S(\mathbf{G})S_S(\lambda) + g_B(\mathbf{G})S_B(\lambda)\right) ,$$

where I is the colour image signal, E is the illumination, g_S is the geometrical factor for interface reflection, S_S is the specular reflectance, g_B is the geometrical factor for body reflection, S_B is the body reflectance, and $\mathbf{G}$ is the geometrical variable that accounts for the viewer direction and the source direction, where each is relative to the surface normal.
Furthermore:
- the NIR model: $S_S(\lambda) = S_S$

is assumed.

Description: The authors of this work model both reflectance and illumination by means of the first three Fourier bases $E_0 = S_0 = 1$, $E_1 = S_1 = \sin\lambda$ and $E_2 = S_2 = \cos\lambda$, so that, assuming the NIR model, the colour image signal can be rewritten as:

$$I = (\epsilon_0 E_0 + \epsilon_1 E_1 + \epsilon_2 E_2)\left[(g_S\sigma_{0S} + g_B\sigma_{0B})S_0 + g_B\sigma_{1B}S_1 + g_B\sigma_{2B}S_2\right] ,$$

where ϵ_i and σ_{ij} are basis coefficients. From this, $\sigma_0 = g_S\sigma_{0S}+g_B\sigma_{0B}$, $\sigma_1 = g_B\sigma_{1B}$ and $\sigma_2 = g_B\sigma_{2B}$ are defined. Besides, by whitening the illumination, the authors remove the terms $\epsilon_1 E_1$ and $\epsilon_2 E_2$. Finally, an HSI space is defined as:

$$I = \sigma_0 , \qquad H = \tan^{-1}\frac{\sigma_2}{\sigma_1} , \qquad S = \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_0} .$$

By analyzing the new $S_0S_1S_2$ space, they find that:
- in the absence of highlights, objects shading generates a straight line passing through the origin in $\mathbf{S}$ space, and, as the saturation is the slope of this line with respect to the S_0 axis, the saturation values of colour points on the linear cluster are constant;
- under white illumination, highlights shift colour points upward along the S_0 direction in $\mathbf{S}$ space, and planar clusters containing the previous straight line are created, where the planes are orthogonal to the S_1-S_2 plane; besides, highlights decrease saturation values compared to those points without highlights;

- shadows do not change the shape or the orientation of the linear cluster, but moves the colour points towards the origin, leaving, thus, unaltered the hue and the saturation values; and,
- although large inter-reflections can importantly change hue and saturation, small inter-reflections can be identified as small deviations from the linear cluster.

Therefore, local thresholding on saturation values separates body reflection from highlights and small inter-reflections, while hue segmentation allows distinguishing among the different materials of the scene.
Several images of plastic balls and glossy blocks are used in the experiments.

3.3.2.3 Physics-based region competition

Auhor(s): Song Chun ZHU [313], Alan YUILLE [313]

Reference(s): [313] 1996

Image formation model:

$$I(\lambda, \boldsymbol{r}_i) = \rho(\lambda, \boldsymbol{r}_i) F_b(\boldsymbol{v}, \boldsymbol{n}, \boldsymbol{s}) E(\lambda, \boldsymbol{n}, \boldsymbol{s}) + F_s(\boldsymbol{v}, \boldsymbol{n}, \boldsymbol{s}) E(\lambda, \boldsymbol{n}, \boldsymbol{s}) \,,$$

where I is the irradiance at the sensor cell, $\boldsymbol{r}_i$ is an image point, ρ is the body reflectance, F_b is the body geometrical factor, E is the surface irradiance, and F_s is the surface geometrical factor. As can be observed in the model, interface reflectance is assumed constant and equal to one.

Description: Although this paper presents a general approach for image segmentation, without a priori intention of embedding a physics-based model of image formation, when applying the algorithm to colour images the authors incorporate a particular case of the dichromatic reflection model. The general algorithm is a statistical and variational approach named *region competition* which minimizes a generalized Bayes/Minimum-Description-Length (MDL) criterion, which combines the length of the region boundaries with the probabilities of region pixels of belonging to them, so that a distribution probability for regions is assumed. After a random seed placement, region boundaries start moving so as to minimize the guiding criterion until all image regions touch. Next, a series of alternative boundary adjustment and region merging stages refine the segmentation until the minimum of the guiding criterion is reached, what is proved to exist by the authors in the paper. This general framework is next adapted to operate with gray-level, colour and textured images.
As for the colour case, regions are modelled as elongated clusters in RGB space, whose direction is assumed to coincide with the average direction of the RGB points involved. Next, points are projected over the plane whose normal vector is the direction of the cluster, reducing the problem to two dimensions. After the projection, colour points are assumed to correspond to a bi-dimensional Gaussian distribution, what will be used in the general algorithm as the distribution probability model. Throughout the different boundary refinement iterations, the cluster direction is updated as well as the distribution parameters. The cluster is further classified as matte or glossy depending on the elongation of the projected cluster, what is used to detect highlights in the image.
Results are provided for three more or less standard images, although they are not guaranteed to have been taken with gamma correction turned off.

3.3.2.4 Robust colour segmentation using the dichromatic reflection model

Auhor(s): Chung-Kiat ONG [194, 195], Takashi MATSUYAMA [194, 195]

Reference(s): [194] 1998, [195] 1999

Image formation model:

$$\boldsymbol{\rho}^x = m_b^x \boldsymbol{\rho}_b + m_i^x \boldsymbol{\rho}_i \,,$$

where $\boldsymbol{\rho}^x$ is the vector of sensor responses at pixel location x, m_b and m_i are, respectively, the geometrical scaling factors, and $\boldsymbol{\rho}_b$ and $\boldsymbol{\rho}_i$ are the body reflection and interface reflection vectors.

Description: Without any assumption about the illumination or the materials present in the scene, the authors of this work present an algorithm for detecting clusters in RGB space using robust statistics and then merging the resulting clusters according to the dichromatic reflection model. Further, they introduce the terms *specular interface reflection* (SIR) and *diffuse interface reflection* (DIR) to distinguish specular clusters for, respectively, rather polished surfaces and rather rough surfaces. This is because, in this last case, the specular cluster is almost indistinguishable from the body cluster, except for a small protuberance in one side of the body cluster.
The algorithm consists of the following stages: first, ellipsoidal clusters are identified by means of an iterative application of the Minimum Volume Ellipsoid estimator, removing clusters from the space as are identified; next, after removing outliers by means of the Least Median of Squares (LMedS), clusters are classified as matte, planar or unknown on the basis of the eigenvalues of the corresponding covariance matrix and the distance of colour pixels to the corresponding geometric entity: straight line or plane; finally, pairs of clusters whose combination gives rise to an entity in accordance with the dichromatic reflection model are merged. Up to six thresholds are used to model the matte and planar clusters, although all are then set as a function of the estimated camera noise, which is also used for determining the clusters dimensionality. The final pixel labelling is carried out by taking the predominant label in each 8-neighbourhood pixel.
Results for several images involving glossy paper and clothing are provided.

3.3.2.5 Supervised colour image segmentation in the HDI space

Auhor(s): Vladimir KRAVTCHENKO [132], James LITTLE [132]

Reference(s): [132] 1999

Image formation model:

$$\boldsymbol{C} = m_B \boldsymbol{C}_B + m_S \boldsymbol{C}_S \,,$$

where $\boldsymbol{C}$ is the vector of camera responses, m_B is the body geometrical factor, $\boldsymbol{C}_B$ is the body reflection vector, m_S is the surface geometrical factor and $\boldsymbol{C}_S$ is the surface reflection vector.

Description: The goal of this work is the efficient segmentation of a sequence of live images for use in real-time applications including object tracking, system navigation and material inspection. To this end, the authors analyze the cluster structure for uniformly coloured objects in a new colour model named Hue-Distance-Intensity (HDI), where *hue* is the same as in the HSV colour space, *distance* is the distance from the RGB point to the gray diagonal and *intensity* is defined as $(R+G+B)/\sqrt{3}$. After a series of experiments and taking into account camera limitations, the pixels of a uniformly coloured object are determined to lie in two planes defined by four points over the corresponding unsaturated pixels cluster in HDI space: P1 is the point with minimum intensity, P4 is the point with maximum intensity, P3 is the farthest point from the P1-P4 line and P2 is the farthest point from the P1-P3 line; from these points, planes P1-P3-P4 and P1-P2-P3 are deemed to considerably improve the recognition of object pixels. In order to cope with image noise, a certain thickness is allocated to both planes. With this structure in mind, a compressed Look-Up-Table (LUT) is designed so that, given an RGB triplet, it can be determined whether the pixel belongs to the object or not. It is therefore a supervised segmentation strategy in the sense that a LUT must be built for every object under consideration, although several can be combined provided that no overlapping colour values among objects exist.
No image results are explicitly provided in the paper.

3.3.2.6 Colour image segmentation for scenes containing vegetation and soil

Auhor(s): C. ONYANGO [196], J. MARCHANT [196]

Reference(s): [196] 2001

Image formation model:

$$L(\theta, \lambda) = m_b(\theta)L_b(\lambda) + m_s(\theta)L_s(\lambda)\,,$$

where L is the SPD of object reflection, θ represents the photometric angles, m_b is the body geometrical factor, L_b is the SPD of body reflection, m_s is the surface geometrical factor and L_s is the SPD of surface reflection.
Furthermore:
- the NIR model: $L_s(\lambda) = c_s(\lambda)L_d(\lambda) = c_sL_d(\lambda)$, where L_d is the SPD of the illuminant,

is assumed.

Description: This work presents an approach for discriminating between soil and vegetation in images of natural scenes. Although they are well contrasted in near-infrared images, it is difficult to distinguish between live and dead vegetation using only the infrared band. Because of this, the authors consider working in RGB space. However, since it is difficult to find the clusters structure predicted by Shafer and Klinker for such sort of images, although both materials satisfy the dichromatic reflection model, they adopt a different approach. Assuming the colour of the illuminant is known (they expect daylight conditions, whose model can be obtained from CIE standards), their algorithm is based on looking for the plane that, containing the illuminant and passing through the origin, intersects the least number of image pixels. The angular position of pixels with regard to this plane is then used as the discriminant.
Results are given for three real images containing both types of material and different lighting conditions which are first estimated with the aid of a plate of Barium Sulphate, a spectrally neutral material.

3.3.2.7 Efficient colour representation for image segmentation under non-white illumination

Auhor(s): Jae Byung PARK [222]

Reference(s): [222] 2003

Image formation model:

$$\boldsymbol{C}(x, y) = m_b(i, e, g)\boldsymbol{C}_b + m_s(i, e, g)\boldsymbol{C}_s\,,$$

where $\boldsymbol{C}$ is the vector of camera responses for the different colour channels, m_b is the body geometrical factor depending on photometric angles i, e and g, $\boldsymbol{C}_b$ is the body reflectance, m_s is the interface geometrical factor, and $\boldsymbol{C}_s$ is the interface reflectance.
Furthermore:
- the NIR model: $C_s(\lambda) = c_sL_d(\lambda)$, where L_d is the SPD of the directional illumination,

is assumed.

Description: This paper proposes a rotation of the RGB cube so as to make easier image segmentation using knowledge about the clusters structure in the new colour space, called $\mathrm{X}_\alpha\mathrm{Y}_\beta\mathrm{Z}_\gamma$ by the author. Such a rotation makes new axis Z_γ coincide with the direction of the illuminant which must, therefore, be estimated in advance. Vector (α, β, γ), from which the colour model gets its name, is the illuminant unit direction in RGB space. Once pixels are in $\mathrm{X}_\alpha\mathrm{Y}_\beta\mathrm{Z}_\gamma$, dichromatic planes are orthogonal to plane $\mathrm{X}_\alpha\mathrm{Y}_\beta$, so that more or less linear clusters are expected if pixels are projected over such a plane. Accordingly, a histogram of $\tan^{-1}(\mathrm{Y}_\beta/\mathrm{X}_\beta)$ should contain peaks corresponding

to the different materials present in the scene. Otsu's algorithm [216] is applied to this histogram to find optimum thresholds maximizing the *between-class variance*.
A single image of a red, a yellow and a green pepper is used to illustrate the segmentation method.

3.3.3 CONTINUITY OF THE IMAGE SURFACE

3.3.3.1 Image segmentation by directed region subdivision using the Delaunay triangulation

Auhor(s): Theo GEVERS [68, 72–74], Frans GROEN [72], Vasil KAJCOVSKI [73], Arnold SMEULDERS [74]

Reference(s): [72] 1991, [73] 1994, [74] 1997, [68] 2002

Image formation model:

$$I_C = k_C I_a + \frac{k_C I_s}{d + d_0} (\boldsymbol{N} \cdot \boldsymbol{L}) \,, \text{ for } C \in (R, G, B) \,,$$

where I_C is the intensity for colour channel C, k_C is the coefficient of reflection for colour channel C, I_a represents the amount of ambient lighting of the scene, I_s is the intensity of a directional light source, $\boldsymbol{N}$ is the unit surface normal vector at the point considered and $\boldsymbol{L}$ is the unit direction towards the light source. From the previous equation it can be deduced that ambient and directional illumination are assumed white. The work also assumes the integrated white condition, which, under the white illumination assumption, can be expressed as $\int_\Lambda \tau^R(\lambda)s(\lambda)\mathrm{d}\lambda = \int_\Lambda \tau^G(\lambda)s(\lambda)\mathrm{d}\lambda = \int_\Lambda \tau^B(\lambda)s(\lambda)\mathrm{d}\lambda$. Both gray-level and colour images are considered.

Description: Throughout the papers referenced in this survey entry, Gevers et al. present a segmentation strategy inspired in the work by Besl and Jain [14–16]. More precisely, they study different tessellation grids and image features within a split-and-merge segmentation strategy on the basis of the continuity of the image surface when a material change is not involved.
On the one hand, in [72], the authors compare two segmentation algorithms working over gray-level images and six colour spaces: RGB, normalized RGB, HSI, L*a*b*, XYZ, normalized XYZ and $I_1I_2I_3$. While one algorithm is just a version of the well-known k-means clustering scheme (with k known), the other algorithm falls within the split-and-merge class. Furthermore, the latter employs the quadtree as the base structure for splitting the image. Two homogeneity criteria are used for guiding the split-and-merge algorithm: one is based on the variance of the region considered and the other one uses the fitting error resulting from approximating the image surface by low-order bivariate polynomials. At the end of the paper, the authors conclude that segmentation results based on colour information are better than those based on gray-level information, normalized colour systems provide the best segmentation results, the split-and-merge algorithm outperforms the one based on k-means and, finally, the variance-based homogeneity criterion works better for normalized colour systems while the one based on low-order bivariate polynomials is adequate for gray-level images and colour systems not factoring out intensity information, such as RGB, XYZ, etc.
On the other hand, [73] focuses directly on the split-and-merge algorithm but using the incremental Delaunay triangulation instead of quadtrees, because, according to the authors, adapts better to the semantics of the image data. Gray-level images and colour images are considered. In the first case, low-order bivariate polynomials are used for determining the homogeneity of the region under consideration. In the case of colour images, the L*rg space is utilized, where L* is based on a scaling of intensity, and r and g are normalized red and green, respectively; the homogeneity criterion in this case is based on variance over every normalized colour channel, while L* is used when intensity falls below a certain threshold. The image points for feeding the triangulation process come from

pixels with the largest fitting errors, Canny edges, corners and the two end-points of lines found over Canny edge maps.
Finally, with the experience gained with the previous algorithms, the authors propose in [74] and [68] to use this strategy based on the image surface continuity just for gray-level images, making use of the Delaunay triangulation and a homogeneity criterion based on the fitting error to low-order bivariate polynomials. The algorithm mainly works as follows: in a first stage, the image is partitioned in triangles until all the triangles are homogeneous, using Canny edges and corners leading to the best new triangulation as new triangulation points; in a second stage, adjacent triangles whose union is still homogeneous are merged until no more fusions are possible.
The results published for this last algorithm come from a set of gray-level synthetic images of randomly placed geometric shapes over a background of quadratically varying intensity and a real image of three homogeneously coloured objects on a homogeneously painted background.

3.3.4 PHOTOMETRIC INVARIANTS AND SPECIALIZED COLOUR SPACES

3.3.4.1 Image segmentation using normalized colour

Auhor(s): Glenn HEALEY [86–91]

Reference(s): [86] 1988, [87, 88] 1989, [89] 1990, [90, 91] 1992

Image formation model:

$$R(g,\lambda) = \begin{cases} M_S(g)C_S(\lambda) & \text{Metal} \\ M_S(g)C_S(\lambda) + M_B(g)C_B(\lambda) & \text{Inhomogeneous Dielectric} \end{cases}$$

where $R(g,\lambda)$ is the light reflected for wavelength λ and photometric angles g, M_S is the interface geometrical factor, C_S is the surface reflectance, M_B is the body geometrical factor and C_B is the body reflectance.

Description: The work by Healey is based on normalized colour, using the L^2 norm to normalize:

$$\widehat{\boldsymbol{S}} = \frac{\boldsymbol{S}}{\sqrt{s_0^2 + \cdots + s_{N-1}^2}},$$

where $\boldsymbol{S} = (s_0, \ldots, s_{N-1})$ denotes a measured sensor vector. The author emphasizes he does not use the L^1 norm (i.e. $s_0 + \cdots + s_{N-1}$), which is indeed used by other researchers, because this norm makes the difference between normalized colours dependent on the precise colours which are compared not on how different they are.
Clusters corresponding to uniformly coloured objects are modelled as a single straight line or as a combination of two intersecting straight lines, one for the body cluster and the other one for the specular cluster. Besides, a multivariate normal distribution defined in normalized colour space is associated to every material found in the scene.
The algorithm starts detecting edges for one of the bands of the colour image. In a second stage, a region-splitting strategy is followed, which, using a quadtree decomposition, divides the image in rectangular zones until every zone is free of edges. Next, a list of scene materials is created by processing all rectangular zones in the following way: given a certain zone, a mean normalized colour is computed for it and the probability of belonging to any of the currently found materials is determined; if the largest probability exceeds a certain threshold T, the respective pixels are assigned the label of the corresponding material; otherwise, a new material is created and added to the materials list. This process continues until all the pixels of the image have been labelled, although dark pixels (i.e. their colour is within the hyper-sphere of radius $4\sigma_N$ centered at the origin

in sensor space, where σ_N is the standard deviation of the camera noise) are left unsegmented. In the final stage, adjacent regions are merged if: (1) they intersect but no at the origin of the colour space, and (2), along their common boundary, the intensities of one region are in all channels above the intensities of the other region; that is to say, this pair of regions is interpreted as, perhaps a part of, a dichromatic plane.
The author notices that the algorithm works better as more channels are considered, what can be obtained by placing narrow filters in front of a monochrome camera.
Images of plastic and metallic objects are used in the experiments.

3.3.4.2 Segmentation of images using the reflectance ratio photometric invariant

Auhor(s): Shree NAYAR [174–178], Ruud BOLLE [174–178]

Reference(s): [174–177] 1993, [178] 1996

Image formation model:

$$I = se\rho R(\boldsymbol{s}, \boldsymbol{v}, \boldsymbol{n}) \,,$$

where I is the brightness value found in an image pixel, s is the spectral response of the sensor, e is the spectral distribution of the incident light, ρ is the fraction of incident light that is reflected in all directions by the surface, R represents the dependence on the geometry of illumination and sensing, $\boldsymbol{n}$ is the surface normal of the considered scene point, $\boldsymbol{s}$ is the direction towards the light source and $\boldsymbol{v}$ is the direction of the viewer.
Furthermore:
- white illumination: $e(\lambda) = e$, and
- a constant response of the sensor: $s(\lambda) = s$

are assumed.

Description: Although the goal of the authors is to find a photometric invariant for object recognition purposes, in the process, a segmentation of the image is required, which is guided by the so-called *reflectance ratio*. It is invariant to shading and illumination changes and is given by:

$$p(A, B) = \frac{I_A - I_B}{I_A + I_B} \,,$$

where I_A and I_B are the brightness values for nearby image points A and B. Though the image formation model does not correspond to the output of a colour camera, p can be easily generalized to RGB images computing p for every channel.
The algorithm consists of three stages: first, reflectance ratios are determined; second, a sequential labeling segments the image into connected regions of uniform reflectance, where two neighbouring pixels A and B are considered connected if $|p(A, B)| < T$; finally, a reflectance ratio is computed for each of the segmented regions as an average of the reflectance ratios computed along the boundary of the region. Small regions caused by camera aliasing are discarded. Finally, the authors assume the scene consists of smooth objects; otherwise, several regions are created for the same object, which is the case of, for instance, a polyhedral object.
Several images involving objects of different materials are used in the experiments.

3.3.4.3 Segmentation and interpretation of images using multiple hypotheses of image formation

Auhor(s): Bruce MAXWELL [158–167], Steven SHAFER [160–167]

Reference(s): [160] 1993, [161,162], [163] 1995, [158,164,165] 1996, [166] 1997, [159] 1999, [167] 2000

Image formation model:

$$\boldsymbol{C} = m_b \boldsymbol{C}_b + m_i \boldsymbol{C}_i \,,$$

where $\boldsymbol{C}$ is the vector of camera responses for every colour channel, m_b is the body geometrical factor, $\boldsymbol{C}_b$ is the vector of body reflection, m_i is the interface geometrical factor and $\boldsymbol{C}_i$ is the vector of interface reflection.

Description: Throughout this series of papers, Maxwell and Shafer propose a general framework for image segmentation, entering in fact and in a certain sense the domain of image interpretation. No new physics-based fundamental techniques are introduced, but it is rather a framework which uses previous knowledge to produce clever segmentations in which, for instance, different patches of colour on the surface of the same object are interpreted as belonging to the same object, what is traditionally left to higher levels of processing by almost all the physics-based segmentation algorithms published so far. Because of this, the algorithm itself does not fall in any of the categories proposed in section 3.3. However, it has been put in this section grouping algorithms making use of photometric invariants because it makes use of some of them in one moment or another of the process.

The algorithm mainly comprises four stages: first, an initial segmentation is computed making use of normalized colour; second, every pair of adjacent regions is analyzed and hypotheses about whether can be merged or not are generated, where those hypotheses result from the analysis of the reflectance ratio, the coherence of the gradient direction and the smoothness of the intensity profile along scanlines across adjacent regions; third, a graph of hypotheses is built, containing interpretations about how the image could have been generated taking into account the evidences provided by the operators examined in the second stage; finally, segmentations are extracted from the graph after a clustering process in order to avoid the combinatorial explosion which would result from the graph if all the possible combinations were explored.

Several images containing glossy and matte artificial objects of a certain complexity are used throughout the papers but only for illustration purposes.

3.3.4.4 Colour edge detection in analog VLSI

Auhor(s): Frank PEREZ [227] and Christof KOCH [227]

Reference(s): [227] 1994

Image formation model:

$$C = \int_\lambda I_a(\lambda) k_a(\lambda) S_C(\lambda) \tau(\lambda)^d \mathrm{d}\lambda + \frac{\cos(\theta)}{r+k} \int_\lambda I_p(\lambda) k_d(\lambda) S_C(\lambda) \tau(\lambda)^d \mathrm{d}\lambda + \frac{k_s \cos^n(\alpha)}{r+k} \int_\lambda I_p(\lambda) S_C(\lambda) \tau(\lambda)^d \mathrm{d}\lambda \,, \text{ for } C = (R, G, B) \,,$$

where C is the intensity value for a certain colour channel, I_a is the ambient intensity, k_a is the ambient reflectance, I_p is the intensity of a point light source, r is the distance from the perspective viewpoint to the surface, k is a constant, k_d is a diffuse reflection coefficient, k_s is a specular reflection coefficient, α is the angle between the perfect specular direction and the viewing direction (Phong model), n is the roughness of the surface (Phong model), S_C is the filter transmittance for the particular colour channel C, $\tau(\lambda)$ is the transmittance per unit distance of the transmitting medium, and d is the thickness of a transparent medium. (The integrals are taken over the visible spectrum.)

Furthermore:

- the integrated white condition: $\int_\lambda I_p(\lambda) S_R(\lambda)\,\mathrm{d}\lambda = \int_\lambda I_p(\lambda) S_G(\lambda)\,\mathrm{d}\lambda = \int_\lambda I_p(\lambda) S_B(\lambda)\,\mathrm{d}\lambda$,
- proportionality between ambient and point source illumination: $I_a(\lambda) = c I_p(\lambda)$,
- a spectrally uniform transmission medium: $\tau(\lambda) = \tau_0$, and

- the neutral interface reflection (NIR) model: $k_s(\lambda) = k_s$

are assumed.

Description: With the aim of understanding the basis of colour segmentation in primate cortex for imitating those structures in analog CMOS VLSI circuits, the authors choose hue for an edge detector in accordance with electrophysiological studies that suggest that hue is computed at a high level in the nervous system. From the different definitions of the HSI space available in the vision literature, they select:

$$I = \frac{R+G+B}{3}\,, \qquad S = 1 - \frac{\min\{R,G,B\}}{I}\,, \qquad H = \tan^{-1}\frac{\sqrt{3}(G-B)}{(R-G)+(R-B)}$$

because of its straightforward implementation in hardware.
In the paper, it is proved that if the *integrated white condition*[1] holds, hue is invariant to certain types of highlights, shading and shadows, under the conditions expressed in the image formation model, due to the *additive/shift invariance*[2] property. It is also shown that the *multiplicative/scale invariance*[3] property of hue also allows invariance to transparencies, also under the conditions indicated.
The edge maps are computed looking for zero-crossings of what they call the *modulo Laplacian* operator applied to hue, which essentially takes into account the circular nature of hue. In order to cope with the instabilities of hue in regions of low saturation or low intensity, the hue is smoothed by means of a Markov Random Field-based operator. Furthermore, when designing the hue edge detector in hardware, colour filters are selected so as to satisfy the integrated white condition.
The results published involve real images including specularities, shadows, inter-reflections, etc. and different surface materials.

3.3.4.5 Colour edge detection in the HSV space

Auhor(s): Peter TSANG [284, 285], W.H. TSANG [284, 285]

Reference(s): [284] 1996, [285] 1997

Image formation model:

$$\boldsymbol{C}_L = m_d\boldsymbol{C}_d + m_s\boldsymbol{C}_s + \boldsymbol{C}_a\,,$$

where $\boldsymbol{C}_L$ is the vector of camera responses for every colour channel, m_d is the diffuse scaling factor, $\boldsymbol{C}_d$ is the vector of diffuse reflection, m_s is the specular scaling factor, and $\boldsymbol{C}_a$ is the ambient illumination vector.
Furthermore:

- the neutral interface reflection (NIR) model: $k_s(\lambda) = k_s$, and
- white illumination, both directional and ambient,

are assumed.

Description: This paper presents an edge detection algorithm using the Hue-Saturation-Value (HSV) colour model on the basis of the invariance properties of the hue component to shading, specularities and shadows, under the conditions stated in the image formation model. First, the algorithm calculates an index I as the product of the V and S components. Second, if the index I is larger than a threshold U, then the hue gradient is determined for this pixel, while if the index is below

[1] Under white illumination ($I_p(\lambda) = I_p$) the integrated white condition can be true if colour filters are chosen so as to satisfy $\int_\lambda S_R(\lambda)\mathrm{d}\lambda = \int_\lambda S_G(\lambda)\mathrm{d}\lambda = \int_\lambda S_B(\lambda)\mathrm{d}\lambda$.
[2] $H(R,G,B) = H(R+\beta, G+\beta, B+\beta)$
[3] $H(R,G,B) = H(\alpha R, \alpha G, \alpha B)$

a threshold L, the gradient is computed over the V component; otherwise, if I is between U and L, the largest of both gradients is selected.
Results for an only image are provided in the paper, but only the resulting gradient image is given. It is compared with less sophisticated approaches.

3.3.4.6 Region segmentation by photometric colour invariant

Auhor(s): Theo GEVERS [68, 75, 76], Arnold SMEULDERS [75, 76], Harro STOKMAN [76]

Reference(s): [75] 1997, [76] 1998, [68] 2002

Image formation model:

$$C = m_b(\boldsymbol{n}, \boldsymbol{s}) \int_\lambda f_c(\lambda) e(\lambda) c_b(\lambda) \mathrm{d}\lambda + m_s(\boldsymbol{n}, \boldsymbol{s}, \boldsymbol{v}) \int_\lambda f_c(\lambda) e(\lambda) c_s(\lambda) \mathrm{d}\lambda\,, \text{ for } C \in (R, G, B)\,,$$

where C is the intensity value for a certain colour channel, m_b is the body geometrical factor depending on surface normal $\boldsymbol{n}$ and direction towards the light source $\boldsymbol{s}$, f_c is the colour filter transmittance, e is the incident light, c_b is the body reflectance, m_s is the interface geometrical factor depending on $\boldsymbol{n}$, $\boldsymbol{s}$ and the direction of the viewer $\boldsymbol{v}$, and c_s is the interface reflectance. (The integrals are taken over the visible spectrum.)
Furthermore:

- the integrated white condition: $\int_\lambda f_R(\lambda) e(\lambda)\,\mathrm{d}\lambda = \int_\lambda f_G(\lambda) e(\lambda)\,\mathrm{d}\lambda = \int_\lambda f_B(\lambda) e(\lambda)\,\mathrm{d}\lambda$,
- white illumination: $e(\lambda) = e$, and
- the NIR model: $c_s(\lambda) = c_s$

are assumed.

Description: Throughout a series of papers, the authors experiment with several colour spaces and photometric invariants, as well as with different algorithms, including k-means (with k known) and split-and-merge, for different tessellation grids.
In [75], the Delaunay triangulation is used within a split-and-merge framework over the projection of RGB values over a plane passing through the (255,0,0), (0,255,0) and (0,0,255) RGB points and named kl-plane by the authors. Under white illumination and if the integrated white condition holds, pixels of uniformly coloured objects lie on a straight line over the kl-plane. The fitting error to a straight line is thus used in the homogeneity criterion.
The work of [76], on the other hand, considers different colour spaces (RGB, XYZ, $\mathrm{I_1I_2I_3}$, normalized RGB, U*V*W*, L*a*b* and HSI) and compares a k-means-based clustering scheme with a split-and-merge strategy based on quadtrees. The sum in quadrature of region colour channel variances is used in the homogeneity criterion in the case of photometric invariants, while, for the case of RGB, the knowledge about the clusters structure in colour space is taken into account.
Finally, [68] proposes a split-and-merge algorithm supported by a Delaunay triangulation tessellation grid and the photometric invariants $c_1c_2c_3$ and $l_1l_2l_3$ defined as follows:

$$\begin{aligned} c_1(R,G,B) &= \tan^{-1}\left(\tfrac{R}{G}\right) & l_1(R,G,B) &= \frac{|R-G|}{|R-G|+|B-R|+|G-B|} \\ c_2(R,G,B) &= \tan^{-1}\left(\tfrac{R}{B}\right) & l_2(R,G,B) &= \frac{|R-B|}{|R-G|+|B-R|+|G-B|} \\ c_3(R,G,B) &= \tan^{-1}\left(\tfrac{G}{B}\right) & l_3(R,G,B) &= \frac{|G-B|}{|R-G|+|B-R|+|G-B|} \end{aligned}$$

where the former achieves invariance to shading and the latter also to specularities, under the conditions stated in the image formation model.
The algorithm mainly works as follows: in a first stage, the image is partitioned in triangles until all the triangles are homogeneous, using Canny edges and corners leading to the best new triangulation as new triangulation points; in a second stage, adjacent triangles whose union is still homogeneous are merged until no more fusions are possible.
The results published for this last algorithm come from an only image of plastic objects over a background divided in four uniformly coloured squares.

3.3.4.7 Colour invariant snakes

Auhor(s): Theo GEVERS [70, 71], Sennay GHEBREAB [71] and Arnold SMEULDERS [71]

Reference(s): [71] 1998, [70] 2004

Image formation model:

$$C = m_b(\boldsymbol{n}, \boldsymbol{s}) \int_\lambda f_c(\lambda) e(\lambda) c_b(\lambda) \mathrm{d}\lambda + m_s(\boldsymbol{n}, \boldsymbol{s}, \boldsymbol{v}) \int_\lambda f_c(\lambda) e(\lambda) c_s(\lambda) \mathrm{d}\lambda\,, \text{ for } C \in (R, G, B)\,,$$

where C is the intensity value for a certain colour channel, m_b is the body geometrical factor depending on surface normal $\boldsymbol{n}$ and direction towards the light source $\boldsymbol{s}$, f_c is the colour filter transmittance, e is the incident light, c_b is the body reflectance, m_s is the interface geometrical factor depending on $\boldsymbol{n}$, $\boldsymbol{s}$ and the direction of the viewer $\boldsymbol{v}$, and c_s is the interface reflectance. (The integrals are taken over the visible spectrum.)
Furthermore:
- the integrated white condition: $\int_\lambda f_R(\lambda) e(\lambda)\,\mathrm{d}\lambda = \int_\lambda f_G(\lambda) e(\lambda)\,\mathrm{d}\lambda = \int_\lambda f_B(\lambda) e(\lambda)\,\mathrm{d}\lambda$,
- white illumination: $e(\lambda) = e$, and
- the NIR model: $c_s(\lambda) = c_s$

are assumed.

Description: This time Gevers and colleagues present an active contour supervised segmentation strategy guided by photometric invariants instead of just intensity. Normalized RGB and the typical formulation for hue are used as invariants against, respectively, shading only, and both shading and specularities (under the conditions stated in the image formation model), in [76]. Colour ratios

$$m_1(R_{\boldsymbol{x}_1}, R_{\boldsymbol{x}_2}, G_{\boldsymbol{x}_1}, G_{\boldsymbol{x}_2}) = \frac{R_{\boldsymbol{x}_1} G_{\boldsymbol{x}_2}}{R_{\boldsymbol{x}_2} G_{\boldsymbol{x}_1}}\,,$$
$$m_2(R_{\boldsymbol{x}_1}, R_{\boldsymbol{x}_2}, B_{\boldsymbol{x}_1}, B_{\boldsymbol{x}_2}) = \frac{R_{\boldsymbol{x}_1} B_{\boldsymbol{x}_2}}{R_{\boldsymbol{x}_2} B_{\boldsymbol{x}_1}}\,,$$
$$m_3(G_{\boldsymbol{x}_1}, G_{\boldsymbol{x}_2}, B_{\boldsymbol{x}_1}, B_{\boldsymbol{x}_2}) = \frac{G_{\boldsymbol{x}_1} B_{\boldsymbol{x}_2}}{G_{\boldsymbol{x}_2} B_{\boldsymbol{x}_1}}\,,$$

where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are image locations, and polar coordinates

$$\theta_1 = \tan^{-1}\left(\frac{R}{B}\right)\,, \qquad \theta_2 = \tan^{-1}\left(\frac{G}{B}\right)$$

are proposed in [70] against just shading (again, under the conditions stated in the image formation model). Intensity uncertainties are incorporated in [70] to compute adaptive thresholds. They are determined by computing the variance of homogeneously coloured surface patches in an image under controlled imaging conditions.
Images of simple plastic objects over uniform background are used in the experiments with static images, while a human figure over a slightly textured background is used in the video sequences.

3.3.4.8 Photometric invariant region detection in multi-spectral images

Auhor(s): Harro STOKMAN [80, 262–264], Theo GEVERS [80, 263, 264]

Reference(s): [263, 264] 1999, [262] 2000, [80] 2003

Image formation model:

$$h_i = m_b(\boldsymbol{n}, \boldsymbol{s}) \int_\lambda f_i(\lambda) E(\lambda) c_b(\lambda) \mathrm{d}\lambda + m_s(\boldsymbol{n}, \boldsymbol{s}, \boldsymbol{v}) \int_\lambda f_i(\lambda) E(\lambda) c_s(\lambda) \mathrm{d}\lambda \,,$$

where h_i is the camera output for filter f_i with central wavelength i, m_b is the body geometrical factor depending on surface normal $\boldsymbol{n}$ and direction towards the light source $\boldsymbol{s}$, E is the incident light, c_b is the body reflectance, m_s is the interface geometrical factor depending on $\boldsymbol{n}$, $\boldsymbol{s}$ and the direction of the viewer $\boldsymbol{v}$, and c_s is the interface reflectance.
Furthermore:

- the integrated white condition: $\int_\lambda f_1(\lambda)\mathrm{d}\lambda = \int_\lambda f_2(\lambda)\mathrm{d}\lambda = \cdots = \int_\lambda f_N(\lambda)\mathrm{d}\lambda$,
- the NIR model: $c_s(\lambda) = c_s$, and
- narrow-band camera filters

are assumed.

Description: An adaptation of the well-known k-means clustering scheme for photometric invariants and multi-spectral images is proposed in [80, 264]. Multi-spectral images are captured by means of a spectrophotometer by taking samples at different wavelengths closely spaced. The invariance is achieved through chromaticity polar angles and by means of hue polar angles. The former is invariant to shading while the latter is invariant to both shading and glossiness, under the conditions stated in the image formation model. Given a sample $\boldsymbol{c} = (c(\lambda_1), c(\lambda_2), \ldots, c(\lambda_N))$ the chromaticity polar angles are defined as

$$\theta_c(\lambda_i) = \tan^{-1}\left(\frac{c(\lambda_i)}{c(\lambda_N)}\right) \,,$$

while the expression for hue polar angles is

$$\begin{aligned}
\theta_h &= \alpha\left[c(\lambda_i) - (1-\rho_s), \phi(i,N)\right] \,, \\
\rho_s &= 1 - \min\{c(\lambda_1), c(\lambda_2), \ldots, c(\lambda_N)\} \,, \\
\phi(i,N) &= \frac{i-1}{N-1} \cdot \frac{4}{3}\pi \,, \\
\alpha(w_i, \theta_i) &= \tan^{-1}\left(\frac{\sum_{i=1}^N w_i \sin(\theta_i)}{\sum_{i=1}^N w_i \cos(\theta_i)}\right) \,,
\end{aligned}$$

where the typical formulation for hue

$$H = \tan^{-1}\left(\frac{\sqrt{3}(G-B)}{(R-G)+(R-B)}\right)$$

can be attained by assigning indexes $i = 0, 1$ and 2 to, respectively, the red, green and blue colour channels.
Finally, robust clustering is achieved by means of computing uncertainties of hyper-spectral colour samples, propagating them through the invariants and incorporating them in the k-means algorithm as sample weights when computing cluster means. The uncertainties come from a camera noise model including camera gain, shot noise and dark current; noise model parameters are estimated by means of uniformly coloured cards.
The results published involve an only image of plastic objects over a background divided in four uniformly coloured squares.

3.3.4.9 Adaptive detection and classification of colour edges

Auhor(s): Harro STOKMAN [77–79, 262, 265, 266], Theo GEVERS [69, 77–79, 265]

Reference(s): [77] 1999, [78, 262, 265, 266] 2000, [69, 79] 2003

Image formation model:

$$C = m_b(\boldsymbol{n}, \boldsymbol{s}) \int_\lambda f_c(\lambda) e(\lambda) c_b(\lambda) \mathrm{d}\lambda + m_s(\boldsymbol{n}, \boldsymbol{s}, \boldsymbol{v}) \int_\lambda f_c(\lambda) e(\lambda) c_s(\lambda) \mathrm{d}\lambda\,, \text{ for } C \in (R, G, B)\,,$$

where C is the intensity value for a certain colour channel, m_b is the body geometrical factor depending on surface normal $\boldsymbol{n}$ and direction towards the light source $\boldsymbol{s}$, f_c is the colour filter transmittance, e is the incident light, c_b is the body reflectance, m_s is the interface geometrical factor depending on $\boldsymbol{n}$, $\boldsymbol{s}$ and the direction of the viewer $\boldsymbol{v}$, and c_s is the interface reflectance. (The integrals are taken over the visible spectrum.)
Furthermore:

- the integrated white condition: $\int_\lambda f_R(\lambda) e(\lambda)\,\mathrm{d}\lambda = \int_\lambda f_G(\lambda) e(\lambda)\,\mathrm{d}\lambda = \int_\lambda f_B(\lambda) e(\lambda)\,\mathrm{d}\lambda$,
- white illumination: $e(\lambda) = e$, and
- the NIR model: $c_s(\lambda) = c_s$

are assumed.

Description: In this work, Stokman and Gevers study the detection and classification of colour edges into shadow-geometry, highlights or material transitions. To this end, the following colour spaces are utilized in the different papers:

- RGB, in all papers,
- $\mathrm{c_1c_2c_3}$ and $\mathrm{l_1l_2l_3}$, in [69, 77, 78], defined as:
$$c_1(R,G,B) = \tan^{-1}\left(\tfrac{R}{G}\right) \qquad l_1(R,G,B) = \tfrac{|R-G|}{|R-G|+|B-R|+|G-B|}$$
$$c_2(R,G,B) = \tan^{-1}\left(\tfrac{R}{B}\right) \qquad l_2(R,G,B) = \tfrac{|R-B|}{|R-G|+|B-R|+|G-B|}$$
$$c_3(R,G,B) = \tan^{-1}\left(\tfrac{G}{B}\right) \qquad l_3(R,G,B) = \tfrac{|G-B|}{|R-G|+|B-R|+|G-B|}$$
where the former is invariant to shading and the latter is invariant to shading and specularities (under the conditions stated in the image formation model),
- normalized colours r and g and opponent colours o and p, in [262]:
$$r(R,G,B) = \tfrac{R}{R+G+B} \qquad o(R,G,B) = \tfrac{R-G}{2}$$
$$g(R,G,B) = \tfrac{G}{R+G+B} \qquad p(R,G,B) = \tfrac{B-(R+G)/2}{2}$$
where, again, the former is invariant to shading and the latter is invariant to shading and specularities (under the conditions stated in the image formation model), and
- $\mathrm{c_1c_2}$ and $\mathrm{o_1o_2}$, in [79]:
$$c_1(R,G,B) = \tan^{-1}\left(\tfrac{R}{B}\right) \qquad o_1(R,G,B) = \tfrac{R-G}{2}$$
$$c_2(R,G,B) = \tan^{-1}\left(\tfrac{G}{B}\right) \qquad o_2(R,G,B) = \tfrac{B-(R+G)/2}{2}$$
where again the former is invariant to shading and the latter is invariant to shading and specularities (under the conditions stated in the image formation model).

Finally, adaptive detection is achieved by means of computing intensity uncertainties, propagating them through the invariants and incorporating them in the edge detection algorithm. The uncertainties come from a camera noise model including camera gain, shot noise and dark current. Noise model parameters are estimated by means of uniformly coloured cards.
The results published involve images of plastic objects over a uniform background.

3.3.4.10 A scale-space approach to colour invariance

Auhor(s): Jan-Mark GEUSEBROEK [61–67], Arnold SMEULDERS [62–67], Anuj DEV [62, 64, 65], Rein van den BOOMGARD [62–67], Hugo GEERTS [63–65], Frans CORNELISSEN [64, 65]

Reference(s): [64, 65] 1999, [61, 62, 66] 2000, [63, 67] 2001

Image formation model:

$$E(\lambda, \boldsymbol{x}) = e(\lambda, \boldsymbol{x}) \left(1 - \rho_f(\boldsymbol{x})^2\right) R_\infty(\lambda, \boldsymbol{x}) + e(\lambda, \boldsymbol{x}) \rho_f(\boldsymbol{x})\,,$$

where $\boldsymbol{x}$ denotes the position at the imaging plane, $e(\lambda, \boldsymbol{x})$ is the illumination spectrum, $\rho_f(\boldsymbol{x})$ is the Fresnel reflectance, $R_\infty(\lambda, \boldsymbol{x})$ is the material reflectivity and $E(\lambda, \boldsymbol{x})$ is the reflected spectrum in the viewing direction.
Furthermore:
- white illumination: $e(\lambda, \boldsymbol{x}) = e(\boldsymbol{x})$, and
- the NIR model (ρ_f does not depend on wavelength)

are assumed.

Description: On the basis of the Kubelka-Munk model of light reflection for coloured bodies [113], the authors of this work develop a set of invariants for different imaging conditions:
(1)

$$H_{\lambda^m x^n} = \frac{\partial^{m+n}}{\partial \lambda^m \partial x^n} \left\{ \tan^{-1} \left(\frac{E_\lambda}{E_{\lambda\lambda}} \right) \right\}, m, n \geq 0,$$

is invariant to shading and specularities under white but not necessarily uniform illumination;
(2)

$$C_{\lambda^m x^n} = \frac{\partial^n}{\partial x^n} \left\{ \frac{E_{\lambda^m}}{E} \right\}, m \geq 1, n \geq 0,$$

is invariant to shading under white but not necessarily uniform illumination;
(3)

$$W_{\lambda^m x^n} = \frac{E_{\lambda^m x^n}}{E}, m \geq 0, n \geq 1,$$

is invariant to shading under white but not necessarily uniform illumination for planar objects; and
(4)

$$N_{\lambda^m x^n} = \frac{\partial^{m+n-2}}{\partial \lambda^{m-1} \partial x^{n-1}} \left\{ \frac{E_{\lambda x} E - E_\lambda E_x}{E^2} \right\}, m \geq 1, n \geq 1,$$

is invariant to shading under a single not necessarily white illumination spectrum.
Derivatives in the spatio-spectral domain are achieved within the Gaussian colour model of [62], by which, in particular, expressions for E_λ and $E_{\lambda\lambda}$ are derived as convolution with, respectively, first- and second-order derivatives of a Gaussian at spectral scale σ_λ positioned at λ_0. The relationship between E, E_λ and $E_{\lambda\lambda}$ and the RGB colour space is also determined as well as a generalization to two dimensions [63].
Results of edge detection and classification are given for an only image of plastic objects over a background divided in four uniformly coloured squares.

3.3.4.11 Colour edge detection by photometric quasi-invariants

Auhor(s): Joost WEIJER [291, 292], Theo GEVERS [291, 292], Jan-Mark GEUSEBROEK [291, 292]

Reference(s): [291] 2003, [292] 2005

Image formation model:

$$\boldsymbol{F} = e(m^b \widehat{\boldsymbol{C}}^b + m^i \widehat{\boldsymbol{C}}^i),$$

where $\boldsymbol{F}$ is a vector containing the camera output for all colour channels, e is the incident light, m^b is the body geometrical factor, $\widehat{\boldsymbol{C}}^b$ is the composite body reflectance vector, m^i is the interface geometrical factor, and $\widehat{\boldsymbol{C}}^i$ is the composite interface reflectance vector.
Furthermore:
- the NIR model: $\widehat{\boldsymbol{C}}^i = \frac{1}{\sqrt{3}}(1, 1, 1)$, and
- white illumination

are assumed.

Description: An edge detection framework for colour images is proposed in these two papers. This framework main feature resides in a new class of image derivatives which the authors refer to as photometric quasi-invariants:

- $S_x^c = F_x - S_x$ with $S_x = (F_x \cdot \widehat{F})\widehat{F}$ is quasi-invariant to shadows and shading,
- $O_x^c = F_x - O_x$ with $O_x = (F_x \cdot \widehat{C}^i)\widehat{C}^i$ is quasi-invariant to specularities, and
- $H_x^c = F_x - H_x$ with $H_x = (F_x \cdot \widehat{B})\widehat{B}$ and $B = \widehat{F} \times \widehat{C}^i$ is quasi-invariant to shadows, shading and specularities,

where $\widehat{\ }$ denotes a unit vector, $\cdot$ is the dot-product and $\times$ is the cross-product.

These quasi-invariants are not invariant with respect to a photometric variable. However, they share the property with normal invariants that they are insensitive for certain edges, e.g. shadow or specular edges (under the conditions stated in the image formation model). Besides, they have better discriminative power than the full invariants. Furthermore, they do not have the inherent instabilities of full photometric invariants. Despite not achieving total invariance, they are suited to applications such as edge detection and classification, among others.

3.4 Summary and Conclusions

Table 3.1 summarizes most of the information given in section 3.3. In the table, the different columns have the following meanings:

- *algorithm* refers to the algorithm, which is referenced by the number of survey entry;
- *year(s)* is for the year or interval of years of the related publications;
- *GL* and *CO* indicate if the algorithm is suitable for, respectively, gray-level and colour images;
- *A*, *B*, *I*, *S* and *IR* refer to the tolerance of the algorithm against, respectively, ambient illumination, body reflection, interface reflection, shadows and inter-reflections;
- *WI*, *NIR* and *IWC* indicate if the algorithm assumes, respectively, white illumination, the NIR model or the integrated white condition;
- *output* refers to whether the algorithm generates a segmentation (IS) or an edge map (EM);
- *GA* is for the type of general approach: pixel-based (*pixel*), region-based (*region*) or edge-based (*edge*);
- *PB* indicates the class of the algorithm according to the classification discussed for physics-based algorithms at the beginning of section 3.3: reflectance-estimation-based (*reflectance*), cluster-analysis-based (*cluster*), continuity-of-the-image-surface-based (*continuity*) and photometric-invariant-based (*invariant*); and, finally,
- *more ...* provides relevant information for the algorithm which has not been highlighted throughout the previous items.

Several observations can be made from the general information given in the aforementioned table and from the particular pieces of work surveyed:

About the general approach and the type of image considered ...

- *Most work on physics-based segmentation has been devoted to photometric invariance.* Next category in number of approaches is cluster analysis.
- *Almost all the algorithms assume colour images.* Among the algorithms surveyed, those that work over gray-level images also tolerate colour images (survey entries 3.3.3.1 and 3.3.4.2) or can be easily adapted (survey entries 3.3.1.1 and 3.3.1.2). It is of particular relevance the fact that algorithms working only with colour images can miss boundaries among certain pairs of objects, depending on the relationship between their reflectances. This is particularly true of the use of photometric invariants, which can hide some reflectance transitions because of their specific formulation. For instance, the typical formulation for hue used in most algorithms:

$$H = \tan^{-1}\left(\frac{\sqrt{3}(G-B)}{(R-G)+(R-B)}\right),$$

Table 3.1. Summary of physics-based segmentation approaches.

algorithm	year(s)	GL	CO	A	B	I	S	IR	WI	NIR	IWC	output	GA	PB	more ...
3.3.1.1	1983-89	✓			✓							IS	pixel	reflectance	non-coplanar point triplets are required
3.3.1.2	1984-89	✓			✓							IS	pixel	reflectance	umbilical points are required
3.3.2.1	1987-93		✓		✓	✓				✓		IS	region	cluster	region growing
3.3.2.2	1989-96		✓		✓	✓	✓	✓		✓		IS	pixel	cluster	illuminant colour required
3.3.2.3	1996		✓		✓	✓				✓		IS	region	cluster	active contours
3.3.2.4	1998-99		✓		✓	✓						IS	pixel	cluster	robust statistics
3.3.2.5	1999		✓		✓	✓						IS	pixel	cluster	semi-supervised
3.3.2.6	2001		✓		✓	✓				✓		IS	pixel	cluster	illuminant colour required
3.3.2.7	2003		✓		✓	✓				✓		IS	pixel	cluster	illuminant colour required
3.3.3.1	1991-02	✓	✓	✓	✓					✓		IS	region	continuity	Delaunay triangulation tessellation grid; split-and-merge
3.3.4.1	1988-92		✓		✓	✓						IS	region	invariant	metals and dielectrics; quadtree; split-and-merge
3.3.4.2	1993-96	✓	✓		✓				✓			IS	region	invariant	constant response of sensor required
3.3.4.3	1993-00		✓		✓	✓						IS	region	invariant	image interpretation as an added value
3.3.4.4	1994		✓	✓	✓	✓				✓	✓	EM	edge	invariant	ambient illumination proportional to directional illumination required
3.3.4.5	1996-97		✓	✓	✓	✓			✓	✓		EM	edge	invariant	very simple algorithm
3.3.4.6	1997-02		✓		✓	✓			✓	✓	✓	IS	region	invariant	Delaunay triangulation tessellation grid; split-and-merge
3.3.4.7	1998-04		✓		✓	✓			✓	✓	✓	IS	region	invariant	semi-supervised; active contours; propagation of intensity uncertainties
3.3.4.8	1999-03		✓		✓	✓				✓	✓	IS	region	invariant	propagation of intensity uncertainties
3.3.4.9	1999-03		✓		✓	✓			✓	✓	✓	EM	edge	invariant	propagation of intensity uncertainties
3.3.4.10	1999-01		✓		✓	✓			✓	✓		EM	edge	invariant	scale-space invariance
3.3.4.11	2003-05		✓		✓	✓			✓	✓		EM	edge	invariant	introduction of quasi-invariants

is invariant to shading and specularities if the NIR model and the integrated white condition are assumed because hue transforms to:

$$H = \tan^{-1}\left(\frac{\sqrt{3}((L_d\rho_b)^G - (L_d\rho_b)^B)}{2(L_d\rho_b)^R - (L_d\rho_b)^G - (L_d\rho_b)^B}\right), \tag{3.1}$$

when R, G and B are replaced by the corresponding physics-based expressions. However, if two adjacent objects happen to have body reflectances $(\rho_{b1}^R, \rho_{b1}^G, \rho_{b1}^B)$ and $(\rho_{b2}^R, \rho_{b2}^G, \rho_{b2}^B)$ such that:

$$\frac{\sqrt{3}((L_d\rho_{b1})^G - (L_d\rho_{b1})^B)}{2(L_d\rho_{b1})^R - (L_d\rho_{b1})^G - (L_d\rho_{b1})^B} = \frac{\sqrt{3}((L_d\rho_{b2})^G - (L_d\rho_{b2})^B)}{2(L_d\rho_{b2})^R - (L_d\rho_{b2})^G - (L_d\rho_{b2})^B},$$

then they cannot be distinguished only by hue.

About the model of image formation ...

- *Most algorithms tolerate body and specular reflection.* Only some of them restrict to exclusively body reflection (survey entries 3.3.1.1, 3.3.1.2 and 3.3.4.2).
- *Very few algorithms consider ambient illumination.* Only 3.3.3.1, 3.3.4.4 and 3.3.4.5 include it in the image formation model. This fact is of particular importance because most photometric invariants proposed so far leave their invariance condition as soon as ambient illumination is introduced. This is the case of, for instance, the typical formulation of hue, which when R, G and B are replaced by the corresponding physics-based expressions transforms to:

$$\tan^{-1}\left(\frac{\sqrt{3}(G-B)}{(R-G)+(R-B)}\right) = \tan^{-1}\left(\frac{\sqrt{3}\left((L_a\rho_a)^G - (L_a\rho_a)^B + m_b\left((L_d\rho_b)^G - (L_d\rho_b)^B\right)\right)}{2(L_a\rho_a)^R - (L_a\rho_a)^G - (L_a\rho_a)^B + m_b\left(2(L_d\rho_b)^R - (L_d\rho_b)^G - (L_d\rho_b)^B\right)}\right), \tag{3.2}$$

which is not invariant to shading, although the NIR model and the integrated white condition hold. Most of these photometric invariants tolerate ambient illumination if a certain SPD is assumed for it, such as, for example, white illumination, which is what the papers of survey entries 3.3.3.1, 3.3.4.4 and 3.3.4.5 assume. Under those circumstances, $L_a = kL_d$ and, if the NIR model is also assumed, hue above turns again invariant to shading and specularities (equation 3.1).

The following expression involving two RGB triplets, (R, G, B) and (R', G', B'), corresponding to the same material, and their difference $(\Delta R, \Delta G, \Delta B) = (R-R', G-G', B-B')$, do achieve the desired invariance:

$$\tan^{-1}\left(\frac{\sqrt{3}(\Delta G - \Delta B)}{(\Delta R - \Delta G) + (\Delta R - \Delta B)}\right) = \tan^{-1}\left(\frac{\sqrt{3}(m_b - m_b')\left((L_d\rho_b)^G - (L_d\rho_b)^B\right)}{(m_b - m_b')\left(2(L_d\rho_b)^R - (L_d\rho_b)^G - (L_d\rho_b)^B\right)}\right) = \tan^{-1}\left(\frac{\sqrt{3}\left((L_d\rho_b)^G - (L_d\rho_b)^B\right)}{\left(2(L_d\rho_b)^R - (L_d\rho_b)^G - (L_d\rho_b)^B\right)}\right). \tag{3.3}$$

This invariant requires, however, to be able to relate pairs of RGB triplets of the same material. Observe that intensity derivatives would play the same role as $\Delta C, C \in \{R, G, B\}$, in the interior of uniform reflectance regions, but problems would arise at material boundaries.

Other algorithms not making use of photometric invariants are also affected by the incorporation of ambient illumination. This is the case of algorithms based on the analysis of clusters in colour space if they assume the body cluster is a line passing through the origin of the colour space (this is particularly true for survey entries 3.3.2.1 and 3.3.2.6). As was shown in figure 2.6, under ambient illumination, the dichromatic planes do not necessarily contain colour (0,0,0).

- *Shadows and inter-reflections are rarely considered.* It is true that there exist ad-hoc approaches to deal with them, mainly regarding shadows [47, 48, 58–60, 110, 111, 245, 246], but they are not integrated in any of the algorithms surveyed. At most, their tolerance is discussed in some of the publications included in this survey (survey entries 3.3.2.2 and 3.3.4.6).
- *It is common practice to assume the NIR model as well as to impose some constraints about scene illumination.* This is particularly true for the approaches based on photometric invariants. Although the NIR model is generally accepted within the vision community, white illumination is considered a quite idealized condition of incident light.

About the uncertainty of the irradiance measurement ...

- *Most algorithms does not take into account the way how the imaging sensing device works.* Uncertainties of the irradiance measurements provided by the camera are used only by few algorithms. However, those uncertainties are computed as the intensity variance over uniformly coloured plane cards (survey entry 3.3.4.7) or assuming camera noise models covering only part of the noise sources enumerated in section 2.4.4 (survey entries 3.3.4.8 and 3.3.4.9). Algorithms that take into account this information can, in particular, set in an adaptive way, the different thresholds that inevitably any segmentation algorithm incorporates.

About the experiments ...

- *Papers published up until now contain very little experimentation on the proposed physics-based algorithms.* Experimental results sections are mainly devoted to illustrate the algorithm proposed. Among all the algorithms surveyed, only the one presented in survey entry 3.3.3.1 was tested against noise and the corresponding results provided in a publication. Furthermore, none of the publications surveyed in this chapter presents a comparison with other algorithms, nor with physics-based, nor with non-physics-based.
- Furthermore, *test images normally correspond to quite simple scenes, mostly containing plastic objects.*

Summing up, the following conclusions can be drawn from the previous considerations, which, on the other hand, set some lines of research: (1) segmentation algorithms should ensure all the reflectance transitions can be detected; (2) a certain improvement on the models of image formation considered by the segmentation algorithms is needed in order to cope with more realistic situations; (3) the quality of the measurements provided by vision cameras should be taken into account, and this quality should be computed incorporating a noise model as more reliable as possible; and (4) more effort should be devoted to determining the performance of physics-based segmentation algorithms, both regarding their validity in front of a varied set of test images and in comparison with other algorithms.

4

General Framework for Image Segmentation and Edge Detection

Images produced by imaging sensors are the result of sensing the electromagnetic radiation coming from the scene in front of the camera. Such electromagnetic radiation carries information about scene objects as a result of their interaction with scene lighting. Being interested in identifying groups of pixels corresponding to the same object, it seems reasonable that the analysis of the measurements provided by the camera is performed in the light of the physics-based models of the formation of the image, and to process the image accordingly. To this end, this chapter focuses on the development of a general framework for image segmentation and edge detection on the basis of physics-based image formation models.

In the following, section 4.1 introduces the axis notation to be used whenever geometrical relationships between scene elements are to be used, while section 4.2 succinctly reviews the physical processes involved, already studied in chapter 2, and develops a comprehensive expression of image formation. As already mentioned in chapter 2, neither attenuating propagation media, emissivity of surfaces or transmission processes will be incorporated in the model. The whole model is analyzed then to discuss its suitability as a starting point for image segmentation and edge detection in section 4.3. Finally, section 4.4 considers the evaluation of the performance of segmentation and edge detection algorithms. While standard measures are suggested for evaluating the output of edge detectors, a new strategy is proposed for assessing the performance of segmentation algorithms.

4.1 Axis Notation

The viewer-oriented co-ordinate system depicted in figure 4.1 will be assumed from now on whenever it is necessary to make use of geometrical relationships between the different elements of the imaging scenario, namely light sources, scene objects and camera. As a consequence, surface normal vectors pointing from the scene towards the camera have negative z components, as well as vectors expressing the direction towards a light source. Besides, the origin of the co-ordinate frame lies at the center of projection and the image plane is put before the center of projection, at $z = f$, being f the focal distance.

4.2 Expanded Model of Image Formation (without noise)

As it has been discussed in chapter 2, in general, radiance emitted at any scene point comes from reflection from either non-directional and directional light sources. The case of non-directional light corresponds to ambient illumination and represents light with no preferred orientation that is the result of multiple interreflections on the walls and other scene objects, in indoor environments, or the result of multiple difractions in the atmosphere, in the outdoor case. On the other hand, the scene can be illuminated by several directional light sources, and for every one either body and interface reflection processes must be considered. Following the approximation by Hougen and Ahuja [105], the light source

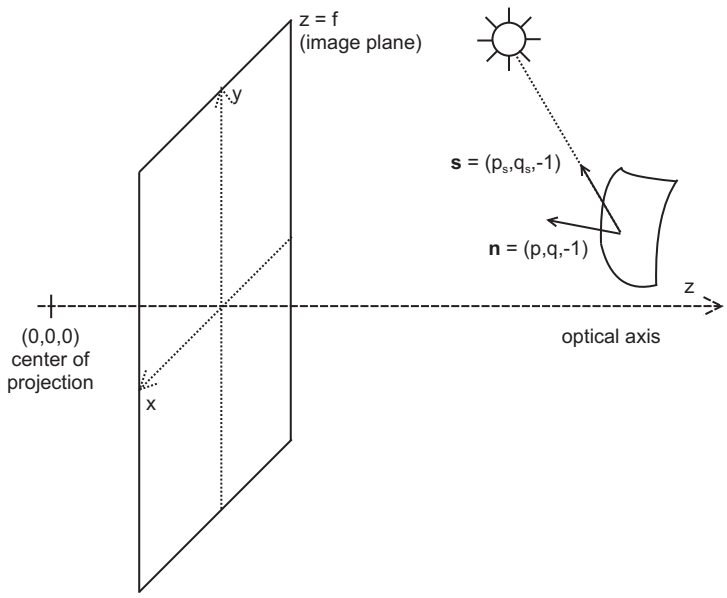


Fig. 4.1. Axis notation.

distribution may be approximated by a set of ls distinct individual directional light sources located at infinity and characterized by their own colour and radiance, and a vector defining their individual orientations. Ignoring interreflections and camera noise, and decomposing the light arriving at image cell (i, j) into the three reflection components of the DRM, ambient, body and interface, equation 4.1 results from the combination of equations 2.31, 2.32 and 2.39:

$$\begin{aligned} D^c(i,j) = & \left[q_0^c \int_\Lambda L_a(\lambda)\rho_a(i,j;\lambda)\tau^c(\lambda)s(\lambda)\mathrm{d}\lambda \right] && \text{(ambient term)} \\ & + \sum_{k=1}^{ls} m_{b,k}(i,j) \left[q_0^c \int_\Lambda L_{d,k}(\lambda)\rho_b(i,j;\lambda)\tau^c(\lambda)s(\lambda)\mathrm{d}\lambda \right] && \text{(body terms)} \\ & + \sum_{k=1}^{ls} m_{i,k}(i,j) \left[q_0^c \int_\Lambda L_{d,k}(\lambda)\rho_i(i,j;\lambda)\tau^c(\lambda)s(\lambda)\mathrm{d}\lambda \right] && \text{(interface terms)}\,, \end{aligned} \tag{4.1}$$

where:

- q_0^c is a scaling factor including Q_0^c (equation 2.32) and $(\pi/4)\ (d/f)^2$ (equation 2.38).
- $L_a(\lambda)$ is the ambient illumination radiance, while $L_{d,k}(\lambda)$ is the radiance corresponding to the directional light source k.
- $m_{b,k}(i,j)$ is the geometrical factor of the body reflection corresponding to image cell (i,j) for the directional light source k. From a theoretical point of view, for every scene point p projecting onto image cell (i,j) there is a possibly different m_b value, so that m_b depends in fact on p (by way of illustration, see figure 2.8). However, after equation 2.39, $L(p;\lambda)$ was assumed constant for all the scene points p optically reachable from image cell (i,j). Accordingly, $m_{b,k}(p)$ turns out to be constant as well, and equal to $m_{b,k}(i,j)$. Using the Lambert's Cosine Law to give form to m_b:

$$m_{b,k}(i,j) = \cos\theta_k(i,j) = \boldsymbol{n}(i,j)\cdot\boldsymbol{s}_k(i,j)\,, \tag{4.2}$$

  where $\theta_k(i,j)$ is the angle between the outward unit surface normal corresponding to image cell (i,j), $\boldsymbol{n}(i,j)$, and the unit direction towards the directional light source k, $\boldsymbol{s}_k(i,j)$. The meanings of θ_k, $\boldsymbol{n}$ and $\boldsymbol{s}_k$ are clearly related to the previous considerations about m_b.

- $m_{i,k}(i,j)$ is the geometrical factor of the interface reflection corresponding to image cell (i,j) for the directional light source k. The above sort of observations for m_b would also apply here. As for the form of $m_{i,k}$, the often-quoted Torrance-Sparrow model is taken, so that:

$$m_{i,k}(i,j) = \frac{D_k(i,j;\overline{m})\, G_k(i,j)\, F_k^{(1)}(i,j)}{\cos\theta_r(i,j)} = \frac{D_k(i,j;\overline{m})\, G_k(i,j)\, F_k^{(1)}(i,j)}{\boldsymbol{n}(i,j)\cdot\boldsymbol{v}(i,j)}\,, \tag{4.3}$$

  where: D is the probability distribution function for the orientation of micro-facet normals, dependent on the surface roughness parameter $\overline{m}$; G is the geometrical attenuation factor that accounts for the shadowing and masking of surface micro-facets by adjacent facets; and $F^{(1)}$ comes from the analysis of Healey in [88], where the Fresnel reflectance term of the Torrance-Sparrow model (equation 2.22) is determined to be quite well approximable by the product of two terms $F^{(1)}$ and $F^{(2)}$, the former depending exclusively on geometry and the latter exclusively on wavelength.
- $\rho_b(i,j;\lambda)$ is the body reflectance for image cell (i,j). The same considerations made above regarding the constancy of $L(p;\lambda)$ also apply for this and the other reflectances.
- $\rho_i(i,j;\lambda)$ is the interface reflectance at image cell (i,j). It is generally defined through the Fresnel reflectance, so that, following Healey [88], $\rho_i = F^{(2)}$.
- $\rho_a(i,j;\lambda)$ is the ambient illumination reflectance corresponding to image cell (i,j). ρ_a is assumed to be a linear combination of ρ_b and ρ_i.

A more compact form of equation 4.1 is given as follows:

$$D^c(i,j) = \overbrace{C_a^c(i,j)}^{D_a^c(i,j)} + \sum_{k=1}^{ls} \overbrace{m_{b,k}(i,j)C_{b,k}^c(i,j)}^{D_{b,k}^c(i,j)} + \sum_{k=1}^{ls} \overbrace{m_{i,k}(i,j)C_{i,k}^c(i,j)}^{D_{i,k}^c(i,j)}\,, \tag{4.4}$$

with

$$C_a^c(i,j) = (L_a\rho_a(i,j))^c = q_0^c \int_\Lambda L_a(\lambda)\rho_a(i,j;\lambda)\tau^c(\lambda)s(\lambda)\mathrm{d}\lambda\,, \tag{4.4a}$$

$$C_{b,k}^c(i,j) = (L_{d,k}\rho_b(i,j))^c = q_0^c \int_\Lambda L_{d,k}(\lambda)\rho_b(i,j;\lambda)\tau^c(\lambda)s(\lambda)\mathrm{d}\lambda\,, \text{ and} \tag{4.4b}$$

$$C_{i,k}^c(i,j) = (L_{d,k}\rho_i(i,j))^c = q_0^c \int_\Lambda L_{d,k}(\lambda)\rho_i(i,j;\lambda)\tau^c(\lambda)s(\lambda)\mathrm{d}\lambda\,. \tag{4.4c}$$

where C_a^c, C_b^c and C_i^c are, respectively, the so-called *ambient*, *body* and *interface composite reflectances*, representing the joint contribution of lighting and material reflectance to the corresponding reflection component. Observe that they have embedded, among others, the integration time T (equation 2.40) and factor $(\pi/4)(d/f)^2$ (equation 2.39), so that, for the same object, variations in exposure time and lens aperture will yield different values for those quantities.

For instance, in an RGB colour camera, the model of equation 4.4 takes the form of equation 4.5:

$$\begin{pmatrix} R(i,j) \\ G(i,j) \\ B(i,j) \end{pmatrix} = \begin{pmatrix} R_a(i,j) \\ G_a(i,j) \\ B_a(i,j) \end{pmatrix} + \sum_{k=1}^{ls} m_{b,k}(i,j) \begin{pmatrix} R_{b,k}(i,j) \\ G_{b,k}(i,j) \\ B_{b,k}(i,j) \end{pmatrix} + \sum_{k=1}^{ls} m_{i,k}(i,j) \begin{pmatrix} R_{i,k}(i,j) \\ G_{i,k}(i,j) \\ B_{i,k}(i,j) \end{pmatrix}\,, \tag{4.5}$$

where (R_a, G_a, B_a), $(R_{b,k}, G_{b,k}, B_{b,k})$ and $(R_{i,k}, G_{i,k}, B_{i,k})$ are, respectively, the ambient, body and interface colours of the corresponding surface material when illuminated by light source k.

Observe now that, in case the different light sources have all the same or similar spectral composition (this could be the case of, for instance, an office), equation 4.4 can be simplified to equation 4.6:

$$D^c(i,j) = \overbrace{C_a^c(i,j)}^{D_a^c(i,j)} + \overbrace{m_b(i,j)C_b^c(i,j)}^{D_b^c(i,j)} + \overbrace{m_i(i,j)C_i^c(i,j)}^{D_i^c(i,j)}\,, \tag{4.6}$$

where, now, $m_b = \sum_{k=1}^{ls} m_{b,k} \lesseqqgtr 1$ and $m_i = \sum_{k=1}^{ls} m_{i,k} \lesseqqgtr 1$.

As a final consideration and as a proof of validity of the previous models, it would be important to note that, in the research area of realistic computer graphics, the often used *ray tracing* rendering techniques implement subsets of equation 4.1 [83, p. 82].

Table 4.1. Scene attributes encoded in the image formation model

Class	Attribute	Symbols involved
object properties	complex refraction index	$(n - K_0 i)(\lambda)$
	surface reflectances	$\rho_a(\lambda)$, $\rho_b(\lambda)$, $\rho_i(\lambda)$
	surface roughness	$\overline{m}$
	object local shape	$\boldsymbol{n}$
light source properties	directional light source(s) radiance distribution	$\boldsymbol{s}_k$, $L_{d,k}(\lambda)$
	ambient illumination radiance	$L_a(\lambda)$
camera properties	colour filter(s) transmittance	$\tau_c(\lambda)$
	camera spectral responsivity	$s(\lambda)$

4.3 Use of a Physics-Based Model for Segmentation and Edge Detection

According to equation 4.1, it is clear that the intensity levels present in the cells of an image encode the values of different attributes of the elements of the scene. Depending whether a method can be devised to recover such pieces of information, it will be possible to rebuild the scene for the general model and elaborate an accurate description of it or not.

Table 4.1 puts together the relevant attributes of a scene whose particular values could, from a theoretical point of view, be recovered through the use of equation 4.1, just because of the fact of appearing there; obviously, some of them could be harder if not impossible to estimate, at least in isolation from the rest, by only analyzing one image. Depending on the vision task at hand, one would be interested in one or other attribute.

In the particular tasks of edge detection and image segmentation, the relevant attributes would be the ones which could be used to tell the different scene objects apart; that is to say, the scene attributes which were in turn object properties. Let us revise them one by one following the order of table 4.1:

- *Complex refraction index.* As it was indicated in chapter 2, the complex refraction index $n - K_0 i$ is related to the object optical properties through the electrical permitivity ϵ, the magnetic permeability μ and the conductivity σ. The refraction index is used to compute the Fresnel reflectance and it appears only in the interface reflection term, so that using equation 4.1 the complex refraction index could only be estimated for glossy objects. Furthermore, it has been proved that the Fresnel reflectance tends to be constant for many materials, which jeopardizes its discriminating power [144, 280].
- *Surface reflectances.* Among the different surface reflectances distinguished in the model, the body reflectance ρ_b is the one that determines what is usually considered as the *perceptual colour* of the object (i.e. when one says an object is reddish it is because $\rho_b(\lambda)$ takes a certain shape). Interface reflectance ρ_i, on the other hand, contributes to the colour of highlights, and therefore to the colour of, most times in most scenes, small areas of the surface of the object. The ambient light reflectance ρ_a is usually considered as a linear combination of ρ_b and ρ_i, so that it does not provide more discrimination power than ρ_b and ρ_i alone.
- *Surface roughness.* The roughness of the surface of an object, $\overline{m}$, is a characteristic of the distribution of micro-facet slopes. Besides, it is only applicable to glossy objects and only refers to a part of the object surface.
- *Object local shape.* The shape of an object is another property which results very useful in the real world when wishing to discriminate among objects. However, although every pixel encodes the surface normal vector $\boldsymbol{n}$ of the corresponding scene point (or area, being more precise), the global shape of an object is really encoded in the group of pixels belonging to such object, not in just one point, what implies the image has been previously segmented.

According to the previous discussion, surface reflectances, in particular ρ_b, seem to be the most appropriate object property for image segmentation and edge detection, from a physics-based point of view. This election, however, has several implications:

- Surface reflectance is not really an object property, but a property of the material of the object surface. Therefore, if the surface of an object consists of several patches with different reflectance properties, a reflectance-based segmentation algorithm will not put those patches together (think, for instance, of the case of the black stripes of a zebra). Similarly, the edge map produced by a reflectance-based edge detector will not contain exclusively the object boundaries.
- In order to be able to segment an image or detect its edges according to surface reflectance, it is essential that the image contains large enough uniform surface reflectance regions. Otherwise, image resolution must be increased in order to get meaningful regions in the segmentation output.

Provided that the segmentation/edge detection strategy has been based on surface reflectances, it results useful to consider the expression for $D^c(i,j)$ as a function $\mathcal{F}$ with the following parameters:

$$D^c(i,j) = \mathcal{F}(\text{scene geometry}, \text{lighting distribution}, \text{surface reflectances})\,. \tag{4.7}$$

In equation 4.7, the term *scene geometry* refers to either the geometry of the objects appearing in the scene and the lighting geometry. On the other side, camera properties have not been included in the definition of $\mathcal{F}$ because they do not change from pixel to pixel.

Equation 4.7 is important in the sense that states that the segmentation/edge detection algorithm must be able to find surface reflectance boundaries surpassing changes in scene geometry and lighting across the scene. This consideration systematizes physics-based segmentation and edge detection and is the basis for the segmentation/edge detection methods presented in this thesis. On the one hand, chapter 7 presents an iterative segmentation algorithm whose basis is the recovery of scene shape after having gathered information about the illumination distribution to finally determine surface reflectance and use it for joining pixels belonging to the same scene material. On the other hand, chapters 8-9 present an indirect strategy which allows locating surface material boundaries without resorting to the estimation of surface reflectance. Neither the scene shape or the illumination distribution need to be known in advance, and can be used to either provide *edge* data or closed contours, depending on the particular implementation. Because the central idea of the algorithm of chapter 7 is the recovery of (numerical) reflectance information, this approach can be called *quantitative* in opposition to the algorithms of chapters 8 and 9 which can be named *qualitative* in the sense they do not need numerical estimations but use *qualities* (i.e. properties) of uniform reflectance image areas derived from the image formation model.

4.4 Discussion about the Experimental Setup

4.4.1 Camera configuration and input data

The diagram of figure 4.2 depicts graphically the setup which is proposed within this general framework for image segmentation and edge detection. The details are discussed in the following:

- Equation 4.1, and all the expressions about the image formation model presented in chapters 2 and 4, assume the camera output is linear, or, in other words, it is proportional to scene radiance. This assumption, which is typical from physics-based vision algorithms, implies the camera correction circuit must be turned off ($\gamma = 1$ in equation 2.47) before taking any image. In case the camera does not allow such an operation, the images captured will have to be linearized as was explained in section 2.4.5.
- Depending on the physics-based image partitioning/edge detection approach, it can be necessary to make available estimates of scene parameters, such as the illumination distribution, before processing the image. This is the case of the segmentation algorithm proposed in chapter 7. To this end, chapter 5 presents a set of methods for estimating lighting information for a case of one distant directional light source and ambient illumination. Since the estimates depend on the camera configuration, some capabilities of present cameras will have to be switched off, such as the *Automatic Gain Control*. In fact, any change in the operating parameters of the camera —the colour channel gains, the camera optics aperture (i.e. the camera F-number) and the shutter speed (i.e. the exposure time)— will lead to the repetition of, in particular, the estimation of the lighting strengths.

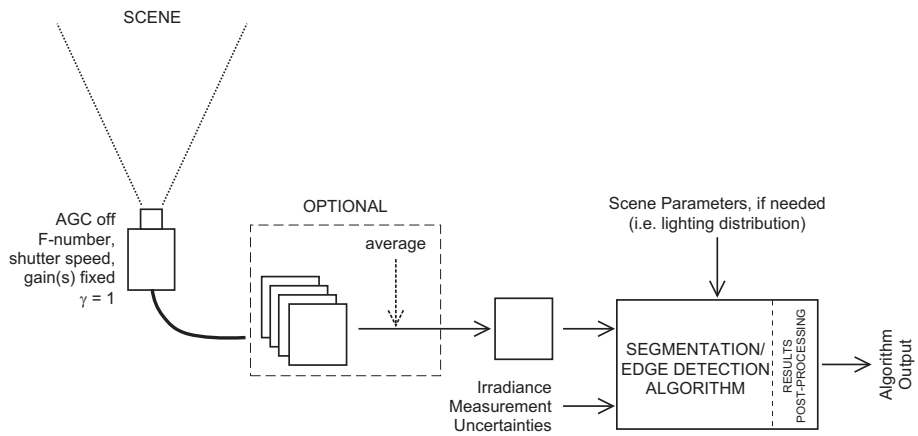


Fig. 4.2. Diagram of the experimental setup for physics-based image segmentation and edge detection.

- The different sources of noise corrupting irradiance measurements in CCD-based cameras can be classified as spatial, referring to the PRNU and the DCNU, and non-spatial, or random, for the rest of noises. Although rough, this taxonomy is important because the influence of non-spatial noises can be reduced if several images of the same scene are taken and the corresponding average image is processed instead of the individual images [98,109]. In effect, if σ is the standard deviation of random noise, the standard deviation of the mean is given by $\sigma/\sqrt{N}$, where N is the number of frames averaged [272]. In this way, it can be said the noise is reduced by the square-root of the number of frames averaged (see table 4.2).
- As will be shown in chapter 6, every digital measurement produced by a camera can be assigned an uncertainty by means of the noise model introduced in section 2.4.4, provided that the relevant parameters of equation 2.44 are estimated, or, in other words, if the camera is radiometrically calibrated. These uncertainties take into account all the noise sources mentioned in section 2.4.4 for every possible digital intensity level. That is to say, if frames were not averaged, the uncertainties would correspond to the measurements contained in the images; otherwise, those uncertainties would represent upper bounds and should be recalculated, if needed, removing the fraction of noise already accounted for.
- Finally, a post-processing stage is planned within this segmentation/edge detection framework with the aim of removing useless information in the algorithm output. In the case of image segmentation, this refers to those regions which, despite having accounted for camera noise, are still created by the algorithm because of, mainly, an anomalous behaviour of the partitioning procedure in extreme

Table 4.2. Percentage of non-spatial noise retained by frame averaging.

# frames averaged (N)	percentage of original noise retained ($100/\sqrt{N}$)
1	100.0
2	70.7
3	57.7
4	50.0
5	44.7
10	31.6
20	22.4
50	14.1
100	10.0
1000	3.2
10000	1.0

cases. Clearly, the usefulness of a region for the subsequent image processing steps should be the main criterion for its survival within this post-processing stage. Some examples of usefulness-related criteria could be linked to region size (i.e. tiny regions correspond, at best, to small details which are perhaps not relevant for the task at hand), contrast with neighbouring regions (i.e. large contrasted regions are preferred rather than small regions differing very little in colour), etc.
As for edge detection, the final post-processing stage is intended to get rid of spurious edges. They are typically characterized by a short length and for being isolated from every other edge.

4.4.2 Performance evaluation of segmentation and edge detection algorithms

General discussion

Since the innumerable edge detection and segmentation algorithms published are all based on their own working hypotheses, the development of common and reasonable criteria for evaluating and comparing their performance has always been a concern for researchers in the area. Given the theoretical complementary nature of the edge detection and the image segmentation problems, it is reasonable to think of a common framework for the evaluation of these algorithms output. That is to say, one can evaluate the quality of a segmentation by assessing the quality of the resultant region boundaries, and, therefore, perform the evaluation of both types of algorithms over exclusively edge maps. However, it is well known that edge detectors rarely produce closed contours, while this is precisely the distinguishing feature of segmentation algorithms output. Because of this, roughly speaking, two types of evaluation strategies can be found in the vision literature: assessment strategies which evaluate the quality of edge maps, and, therefore, can be applied to either edge detection and image segmentation algorithms; and strategies specifically oriented towards assessing image segmentation algorithms, mainly because they measure some qualities of the interior of the regions produced by the algorithm or because they rely on close contours.

In an often-cited survey [306], Zhang makes a deeper classification. First, he distinguishes between two general approaches: analytical and empirical assessment. On the one hand, *analytical methods* refer to the analysis of the segmentation algorithm on the basis of its properties, such as: (1) the type and amount of *a priori* knowledge that has been incorporated into the algorithm; (2) the processing strategy, whether it is parallel, sequential, iterative or mixed; or (3) the processing complexity and efficiency. *Empirical methods*, on the other hand, judge the algorithms analyzing their output. Among them, some are oriented towards measuring the "goodness" of the results (*empirical goodness methods*) while others produce some discrepancy measures between the algorithm output and the expected output, often referred to as ground truth data (*empirical discrepancy methods*). In this latter case, the reference segmentation would be automatically generated in the case of synthetic images, while, for real images, one or several manually segmented images would play the reference role.

Analytical methods avoid the implementation of the algorithm and, thus, reduce the influence of the experimental arrangement. However, not all properties of a segmentation/edge detection algorithm can be captured by analytical studies. In this sense, empirical methods are of larger scope and more general at the expense of implementing the algorithm to be able to perform the assessment. Within this class of methods, empirical goodness methods avoid generating a reference segmentation/edge map by using goodness measures. Nevertheless, those goodness measures should be devised independently of the characteristics of the image. For instance, the *intra-region uniformity* and the *inter-region contrast* are region-oriented goodness measures proposed in many studies [148, 218, 219, 244, 293] that were mostly conceived for images where scene curvature was not very noticeable and thereby object projections could be considered as more or less uniform intensity areas. An empirical discrepancy method can still be more general since the evaluation is based on contrasting the algorithm's output with an already computed reference. The number of pixels incorrectly classified as edge pixels or the number of incorrectly segmented pixels, their position and the number of regions are among the different discrepancy measures proposed in the literature [6, 81, 96, 106, 149, 150, 189, 231, 234, 239, 267, 293, 296, 297]. Recently, Zhang proposed another set of discrepancy measures based on region features such as area, eccentricity or perimeter, among others [305, 307, 308].

Clearly, a single evaluation measure is not enough to judge the result of a segmentation/edge detection algorithm. Haralick and Shapiro stated this idea in [84] in the form of three heuristic criteria for image segmentation evaluation: regions must be uniform and homogeneous, regions shape must be simple, without many holes, and adjacent regions must present significatively different values for the feature defining the partitioning process. Liu and Yang proposed then a function based on these criteria using the number of regions, their areas and the colour error for each region [152]. A more recent work by Borsotti et al. [18] improved this expression after proving empirically that the expression by Liu and Yang did not satisfy, in general, the criteria by Haralick and Shapiro. Some of the studies cited above also propose combined measures [218, 267, 296, 304].

Furthermore, many of the aforementioned evaluation methods were developed assuming image regions of uniform intensity and introduce goodness measures in the evaluation expression representing this fact. For instance, in the case of Liu and Yang, the corresponding function is given as [152]:

$$F(I) = \frac{1}{1000(N \cdot M)} \sqrt{R} \sum_{i=1}^{R} \frac{e_i^2}{\sqrt{A_i}} \,, \tag{4.8}$$

where $N \times M$ is the size of the image, R is the number of regions, A_i is the area of region i and e_i is the colour error, calculated as the sum of the differences between region pixel colours and the colour attributed to region i. Even if colour space transforms or photometric invariants are used to tolerate the change of intensity due to scene curvature, shadows and/or highlights, the inclusion of factors such as e_i in the evaluation expression implies a dependence of the evaluation results on how accurately the model of image formation assumed for the evaluation function represents real images. In the general case, according to section 4.3, e_i should refer to the difference between the reflectance(s) attributed to a given pixel and the reflectance(s) assumed for a region. Depending on the complexity of the image, the mere computation of such goodness measures can be quite difficult if not impossible (chapter 7 will include a discussion on the subject). From this point of view, it is obvious that, in general, it is preferable an evaluation function not including any reference to the colours present in the resultant regions. Clearly, this limits the applicability of goodness measures such as the intra-region uniformity and the inter-region contrast in this context.

Summing up, since using discrepancy-based measures the evaluation procedure reduces to an appropriate comparison with the reference segmentation and nothing is assumed about the characteristics of the true image regions, given their generality, these evaluation measures seem to be the most appropriate election for a general framework of image segmentation and edge detection. Bowyer et al. even propose other types of measures as secondary metrics in [20, p. 78]: *"... Metrics based on qualitative properties of detected edges may well be useful as additional secondary metrics, but they do not seem appropriate as primary performance metrics."* Besides, in the case of image segmentation evaluation, it is proposed to base part of the discrepancy on the pixel classification facet of image segmentation. Accordingly, this part of the assessment would consist in determining whether image pixels are correctly grouped or not, as well as the degree of over- and under-segmentation. The following sections indeed propose a strategy in this line. The other part of the evaluation would take into account distance information between true and predicted region contours. In this way, the evaluation procedure would also measure the accuracy of the segmentation algorithm when locating reflectance transitions.

Image segmentation evaluation and the confusion matrix

This section comments on the use of a matrix C which constitutes the basis for many discrepancy-based segmentation evaluation strategies. C gathers information about the distribution of pixels between the different regions of the image, either real or predicted. An entry of C, say (i, j), contains the number of pixels belonging to region i in the reference segmentation (*reference region* from now on) such that are labeled as pertaining to region j in the algorithm output (*output region* from now on). A particular case of this matrix, in which, among other features, C is square, is known as *confusion matrix*, *error*

Table 4.3. Confusion matrix for a two-class classification problem.

		output	
	class	**foreground**	**background**
reference	**foreground**	true positives (TP)	false negatives (FN)
	background	false positives (FP)	true negatives (TN)

matrix or *contingency matrix* according to the terminology employed in the evaluation of classifiers performance [261][1].

In a classification context, the percentage of correctly classified pixels or, conversely, the percentage of misclassified pixels are the discrepancy measures that come most readily to mind and that can be easily obtained from C. In a two-class case, e.g. a thresholding task, the pixels which are classified as belonging to the foreground figure but are actually part of the background are called *false positives* or *type I errors*, while pixels allocated to the background being part of the foreground figure are the *false negatives* or *type II errors*. According to table 4.3, the *type I error rate* is $FP/(FP+TN)$, the *type II error rate* is $FN/(FN+TP)$, while the percentage of correctly classified pixels or *accuracy* is $(TP+TN)/(TP+TN+FP+FN)$ [6].

In a multi-class case, say N classes, C is an $N \times N$ matrix and the percentage of correctly classified pixels is given by the trace of C, $\sum_{i=1}^{N} C_{ii} / \sum_{i,j=1}^{N} C_{ij}$. As for the misclassified pixels, the *multi-class type I error* for class k or percentage of false positives for class k is expressed as [297,306] (see figure 4.3):

$$M_I^{(k)} = \frac{\left(\sum_{i=1}^{N} C_{ki}\right) - C_{kk}}{\sum_{i=1}^{N} C_{ki}} \times 100\,, \tag{4.9}$$

while the *multi-class type II error* for class k or percentage of false negatives for class k is given by:

$$M_{II}^{(k)} = \frac{\left(\sum_{i=1}^{N} C_{ik}\right) - C_{kk}}{\left(\sum_{i=1}^{N}\sum_{j=1}^{N} C_{ij}\right) - \sum_{i=1}^{N} C_{ki}} \times 100\,. \tag{4.10}$$

In a recent paper [149], Lewis and Brown introduce what they claim to be a *generalized confusion matrix* for producing accurate estimates of the area covered by each class in an image, typically resulting from remote sensing. Essentially, they put matrix C as the combination of a *target matrix* T and an *output matrix* Y, where both matrices have as many rows as pixels in the image and each cell of both is the membership of the corresponding pixel to a class: T contains reference memberships, while Y contains the memberships produced by the classification algorithm. Finally, each row of both matrices sum to unity and $C = T^T Y$ is defined. With the introduction of the membership concept, they contemplate real situations where several classes may occur in the field of view of the satellite sensor represented by a single pixel. Besides, the authors introduce an *error matrix* E as the difference between the generalized confusion matrix and a *reference matrix* $R = T^T T$. Finally, they propose several measures that summarize the information contained in the error matrix, such as the average error among all the pixels, or the error for a given class.

[1] C is sometimes given as an array of probabilities, instead of pixel counts. That is to say, $C(i,j)$ is calculated as the quotient between the number of pixels in reference class i classified in output class j and the number of pixels in the image.

$$
\begin{array}{c|ccc|c|ccc|}
 & & & & k & & & \\
 & & & & \downarrow & & & \\
\hline
 & \mathrm{TN} & \cdots & \mathrm{TN} & \mathrm{FN} & \mathrm{TN} & \cdots & \mathrm{TN} \\
 & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
 & \mathrm{TN} & \cdots & \mathrm{TN} & \mathrm{FN} & \mathrm{TN} & \cdots & \mathrm{TN} \\
\hline
k \rightarrow & \mathrm{FP} & \cdots & \mathrm{FP} & \mathrm{TP} & \mathrm{FP} & \cdots & \mathrm{FP} \\
\hline
 & \mathrm{TN} & \cdots & \mathrm{TN} & \mathrm{FN} & \mathrm{TN} & \cdots & \mathrm{TN} \\
 & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
 & \mathrm{TN} & \cdots & \mathrm{TN} & \mathrm{FN} & \mathrm{TN} & \cdots & \mathrm{TN} \\
\hline
\end{array}
$$

Fig. 4.3. Confusion matrix for a multi-class classification problem. In the matrix, TP, TN, FP and FP correspond to class k.

To finish with this brief review on the use of the confusion matrix, it is important to cite a new discrepancy evaluation strategy rooted on the same principles of the confusion matrix and known as *Receiver Operating Characteristic* (ROC) curve analysis . ROC analysis comes from psychophysics and signal detection theory and has received an important amount of attention within the vision community [1, 20, 43, 51, 233, 234, 256, 259, 298]. A ROC curve plots true positive rates against false positive rates as a decision parameter of the classifier is varied (i.e. the classification threshold). It is therefore mainly suited for two-class classification problems, such as edge detection, although Rees et al. [234] addressed multi-class classification evaluation by means of ROC analysis. They mainly choose a single operating point based on a risk/cost matrix that relates the cost of misclassification between all pairs of classes and then plot this on a class by class basis. In this way, they avoid the generalization of the two-class ROC curve to an (n-1)-order hypersurface for n classes. In any case, ROC analysis is based on the same principles as the confusion matrix. In fact, no new performance measures are provided, it is rather a change of philosophy, as Bowyer et al. remarked regarding the evaluation of edge detectors [20, p. 80]: *"... What is needed is to evaluate detectors over a range of the same TP values. This is the essence of the concept of the receiver operating characteristic (ROC) curve. Sound comparisons are more appropriately made using ROC curves rather than isolated (TP,FP) performance points."*

A discrepancy-based evaluation strategy for image segmentation

Although it is true that sometimes a classification task follows a segmentation task in the pipeline of a vision application, and that, therefore, the evaluation can be applied directly to the output of the classifier, on many occasions it can be important to evaluate the segmentation output *per se*. An important difference which arises then is that, unlike the case of a typical classification task, the final number of output regions/classes does not need to agree with the number of reference regions/classes, what was already pointed out when the confusion matrix was introduced above. This fact can be the consequence of two misbehaviours of the segmentation algorithm, which can happen alone or in combination: (1) on the one hand, a reference region can appear split in several output regions, which is known as *over-segmentation*; and (2) on the other hand, several reference regions can result merged in an only output region, which is known as *under-segmentation*. A non-null degree of over- and/or under-segmentation derives in the following consequences from a practical point of view:

(1) the confusion matrix is no longer square, and, in fact, the term *overlapping area matrix* has been introduced to refer to this new structure [12];

(2) the region numbering of the reference segmentation rarely coincides with the region numbering of the algorithm's output, and
(3) due to (1) and (2), the correspondence between reference and output regions is not trivial (i.e. the k-th row and the k-th column of the confusion matrix do not refer to the same class, unlike the matrix of figure 4.3).

Because of all the aforementioned for the general case of image segmentation, a set of new performance measures are introduced next in accordance with the special features of the confusion matrix. For notation purposes, a reference region i will be denoted as R_i, while output region j will be written as $\widehat{R}_j$. Every measure is summarized in table 4.4 and explained in the following:

- The *percentage of Correctly Grouped pixels*, CG, aims at accounting for those pixels which, belonging to a reference region R_i, are put together in a single output region $\widehat{R}_j$. That is to say, the goal is to determine to which degree the segmentation algorithm does not mix pixels from different reference regions into a single output region. Being strict, only those output regions which completely overlap with a reference region should be considered in CG. However, even in the best of the segmentations, a certain level of dispersion can be found, in the sense that small fractions of regions $\widehat{R}_j$ fall outside the main region R_i. Because of this, the definition of CG has been relaxed through the parameter p, by which $\mathrm{CG}(p)$ represents the amount of pixels which belong to a region $\widehat{R}_j$ which is mostly included in a region R_i, where the meaning of *mostly* depends on p. Therefore, the amount of dispersion which is tolerated is expressed through parameter p, being $p = 100$ the most restrictive case which leads to the first definition of CG. Observe that this scattering of pixels can be easily detected looking at the columns of the confusion matrix, which will contain more than one non-null entry in case a region $\widehat{R}_j$ groups pixels from several regions R_i.
  As can be easily guessed, CG tries to determine the correspondence between reference and output regions. With the introduction of the parameter p, this is particularly true when $p > 50$, since, on those occasions, $\mathrm{SRa}(\cdot, \widehat{R}_j, p)$ takes value 1 only for one R_i, which would correspond to the region R_i which mostly overlaps with region $\widehat{R}_j$. For $p \leq 50$, $\mathrm{SRa}(\cdot, \widehat{R}_j, p)$ can take value 1 for several R_i, which makes $\mathrm{CG}(p)$ account for several fractions of $\widehat{R}_j$; since the sense of correspondence has been lost, CG can be said less meaningful for those values of p.
  As a final remark, notice that, however, even for $p > 50$, several output regions $\widehat{R}_j$ can correspond to a single reference region R_i. As was said in the first definition of CG, this measure just tries to determine to which degree the segmentation algorithm does not put together pixels belonging to different reference regions. Therefore, a segmentation with a 100 % of correctly grouped pixels can be over-segmented (see later in cases (2) and (3) of the example).
- The *percentage of Under-Segmentation*, US, represents the amount of pixels of the image which have been assigned to regions $\widehat{R}_j$ which cover several reference regions R_i. As before, parameter p allows relaxing the definition of US to tolerate slight segmentation errors, so that $\mathrm{US}(p)$ accounts for regions $\widehat{R}_j$ whose overlapping with a reference region R_i is below $p\,\%$. Observe that US accumulates all the pixels of the output region if the previous circumstance arises, contrary to CG, which accumulates fractions of output regions. This is because the concept of under-segmentation is relative to a segmentation; that is to say, it would be said that the output of a segmentation algorithm under-segments an image. CG, however, accounts for groups of pixels which are correct from the point of view of not belonging to several reference regions. In this sense, US has to take into account those output regions which under-segment the image.
- The *percentage of Over-Segmentation*, OS, accounts for pixels of output regions $\widehat{R}_j$ which split a reference region R_i. In this case, the key role is played by the rows of the confusion matrix. If a row, say i, of C contains several non-null entries is because the reference region R_i appears divided into the regions $\widehat{R}_j$ corresponding to those non-null entries. Again, the amount of dispersion tolerated is controlled through parameter p.

Other authors have proposed measures rooted on principles similar to CG, OS and US, and, therefore, aiming at evaluating the amount of misclassification in the segmentation output as well as the level of under- and over-segmentation. A representative set is summarized in tables 4.5 and 4.6 using the notation of table 4.4 and the overlapping area matrix, although the formulations appearing in the

Table 4.4. Discrepancy measures formulae.

N_r	number of reference regions
N_o	number of output regions
$n(R_i) = \sum_{k=1}^{N_o} C_{ik}$	size of region R_i
$n(\widehat{R}_j) = \sum_{k=1}^{N_r} C_{kj}$	size of region $\widehat{R}_j$
$n(I) = \sum_{k=1}^{N_r} n(R_k) = \sum_{k=1}^{N_o} n(\widehat{R}_k)$	size of image
$\mathrm{SRa}(R_i, \widehat{R}_j, p) = \begin{cases} 1 \text{ if } \frac{C_{ij}}{n(\widehat{R}_j)} \times 100 \geq p \\ 0 \text{ otherwise} \end{cases}$	p % pixels of $\widehat{R}_j$ are concentrated in a Single Reference region R_i
$\mathrm{SRb}(\widehat{R}_j, p) = \begin{cases} 1 \text{ if } \frac{\max_{k=1..N_r} \{C_{kj}\}}{n(\widehat{R}_j)} \times 100 \geq p \\ 0 \text{ otherwise} \end{cases}$	at least p % pixels of $\widehat{R}_j$ are concentrated in a Single Reference region R_i
$\mathrm{SO}(R_i, p) = \begin{cases} 1 \text{ if } \frac{\max_{k=1..N_o} \{C_{ik}\}}{n(R_i)} \times 100 \geq p \\ 0 \text{ otherwise} \end{cases}$	at least p % pixels of R_i are concentrated in a Single Output region $\widehat{R}_j$
$\mathrm{CG}(p) = \frac{\sum_{i=1}^{N_r} \sum_{j=1}^{N_o} \mathrm{SRa}(R_i, \widehat{R}_j, p) \times C_{ij}}{n(I)} \times 100$	percentage of Correctly Grouped pixels, at level p
$\mathrm{US}(p) = \frac{\sum_{j=1}^{N_o} \left(1 - \mathrm{SRb}(\widehat{R}_j, p)\right) \times n(\widehat{R}_j)}{n(I)} \times 100$	percentage of Under-Segmentation, at level p
$\mathrm{OS}(p) = \frac{\sum_{i=1}^{N_r} \left(1 - \mathrm{SO}(R_i, p)\right) \times n(R_i)}{n(I)} \times 100$	percentage of Over-Segmentation, at level p

original papers did not explicitly make use of this matrix. A brief discussion on them is provided in the following:

- On the one hand, Levine and Nazif propose in [147] the *under-merging error* UM and the *over-merging error* OM. While the first one would be a synonymous for over-segmentation (i.e. the error related to merging less than it should be is over-segmentation), the second one would correspond to under-segmentation (i.e. a segmentation output which is the result of more fusion than should be leads to under-segmentation). In both measures, the authors associate to every output region $\widehat{R}_j$ the reference region R_k with the maximum spatial coincidence, and compute the error measures

Table 4.5. Other evaluation measures based on the overlapping area matrix (I).

authors	formulae	measure description
Levine and Nazif [147]	$UM = \sum_{j=1}^{N_o} \left(n(R_k) - C_{kj}\right), \quad k = \underset{i=1..N_o}{\arg\max}\{C_{ij}\}$	under-merging error
	$OM = \sum_{j=1}^{N_o} \left(n(\widehat{R}_j) - C_{kj}\right), \quad k = \underset{i=1..N_o}{\arg\max}\{C_{ij}\}$	over-merging error
	$M = \sqrt{\left(\frac{UM}{n(I)}\right)^2 + \left(\frac{OM}{n(I)}\right)^2}$	overall performance
Huang and Dom [106]	$e_R^m = \frac{\sum_{i=1}^{N_r} \sum_{j \neq \underset{k=1..N_o}{\arg\max}\{C_{ik}\}} C_{ij}}{n(I)}$	missing rate
	$e_R^f = \frac{\sum_{j=1}^{N_o} \sum_{i \neq \underset{k=1..N_r}{\arg\max}\{C_{kj}\}} C_{ij}}{n(I)}$	false alarm rate
	$p_R = 1 - \frac{e_R^m + e_R^f}{2}$	overall performance
Hoover et al. [99]	$(R_i, \widehat{R}_j)$ is an instance of a *correct detection* if $C_{ij} \geq T \times n(\widehat{R}_j)$ and $C_{ij} \geq T \times n(R_i)$	#CC, number of correct detection instances ($0.5 < T \leq 1.0$ is a threshold percentage)
	$(R_i, \widehat{R}_{j_1}, \ldots, \widehat{R}_{j_x})$ is an instance of an *over-segmentation* if $C_{ij_t} \geq T \times n(\widehat{R}_{j_t}), \forall t$ and $\sum_{t=1}^{x} C_{ij_t} \geq T \times n(R_i)$	#OC, number of over-segmentation instances ($0.5 < T \leq 1.0$ is a threshold percentage)
	$(R_{i_1}, \ldots, R_{i_x}, \widehat{R}_j)$ is an instance of an *under-segmentation* if $\sum_{t=1}^{x} C_{i_t j} \geq T \times n(\widehat{R}_j)$ and $C_{i_t j} \geq T \times n(R_{i_t}), \forall t$	#UC, number of under-segmentation instances ($0.5 < T \leq 1.0$ is a threshold percentage)
	R_i is a *missed region* if it does not participate in any instance of correct detection, over-segmentation or under-segmentation	#MR, number of missed regions
	$\widehat{R}_j$ is a *noise region* if it does not participate in any instance of correct detection, over-segmentation or under-segmentation	#NR, number of noise regions

accumulating the resulting misclassified pixels. Therefore, they are not intended to compute the level of fragmentation of reference regions into output regions and vice versa. To combine both errors in an overall performance measure, the authors propose a sum in quadrature of UM and OM.

- Later, Huang and Dom presented two quantities very similar to the under- and over-merging errors, but not identical, which they called *missing rate* e_R^m and *false alarm rate* e_R^f [106]. e_R^m rely on determining, for every reference region, the output region for which the matching is maximum, while e_R^f needs the correspondence between every output region and its maximum reference region.

Table 4.6. Other evaluation measures based on the overlapping area matrix (II).

authors	formulae	measure description
Mezaris et al. [169]	let $\mathcal{A} = \{(R_i, \widehat{R}_j)\}$ be such that, given $\widehat{R}_j$ and $i_j = \{i \mid j = \arg\max_{k=1..N_o} \{C_{ik}\}\}$, $i = \arg\max_{k \in i_j} \{C_{kj}\}$, $\forall j = 1..N_o$ let $\mathcal{N}_R$ and $\mathcal{N}_S$ denote the set of, respectively, non-paired reference and non-paired output regions $E_i = \sum_{p \in R_i^*} f_1(d(p, R_i)) + \sum_{p \in \widehat{R}_j^*} f_2(d(p, R_i)), \forall (R_i, \widehat{R}_j) \in \mathcal{A}$ $E_i = \sum_{p \in R_i} f_1(d(p, R_i)), \forall R_i \in \mathcal{N}_R$ $F_j = \alpha \sum_{p \in \widehat{R}_j} f_1(d(p, R(p))), \forall \widehat{R}_j \in \mathcal{N}_S$ $E = \sum_{i=1}^{N_r} E_i + \sum_{\widehat{R}_j \in \mathcal{N}_S} F_j$	overall performance, where: 1. $R_i^* = \left(R_i - R_i \bigcap \widehat{R}_j\right)$, 2. $\widehat{R}_j^* = \left(\widehat{R}_j - R_i \bigcap \widehat{R}_j\right)$, 3. $R(p)$ is the reference region to which p belongs, 4. $d(p, R)$ is the distance from a pixel p and the nearest boundary pixel of region R, 5. f_1 and f_2 are: $f_1(d) = \frac{0.001}{10} d$, $f_2(d) = \begin{cases} \frac{0.001}{10} d & \text{if } d \leq 10 \\ 0.001 & \text{if } d > 10 \end{cases}$, and 6. $\alpha = 2$
Pignalberi et al. [229]	$C = \frac{\sum_{j=1}^{N_o} \left(n(R_{x_j}) - C_{x_j j}\right)}{N_o}$, $x_j = \arg\max_{i=1..N_r} \{C_{ij}\}$ $U = n(I) - \sum_{i=1}^{N_r} \sum_{j=1}^{N_o} C_{ij}$ $H_u = k \sum_{j \mid \sum_{i=1}^{N_r} m_{ij} > 1} n(\widehat{R}_j)$ $H_o = k \sum_{j \mid \sum_{i=1}^{N_r} m_{ij} = 0} n(\widehat{R}_j)$ $F = w_1 C + w_2 H_u + w_3 H_o + w_4 U$	cost of erroneously segmented pixels cost of unlabeled pixels handicaps of under- and over-segmentation, where 1. $m_{ij} = \begin{cases} 1 & \text{if } j = \arg\max_{k=1..N_o} \{C_{ik}\} \\ 0 & \text{otherwise} \end{cases}$, and 2. k is a constant to enlarge the variability range fitness function ($\sum_i w_i = 1$)

Once determined those correspondences, both measures accumulate the pixels of the non-maximum intersections between reference and output regions. As well as for Levine and Nazif, these measures are not intended to compute the level of fragmentation of reference regions into output regions and vice versa. Furthermore, as is indicated in figure 4.4, if a correspondence is established between a

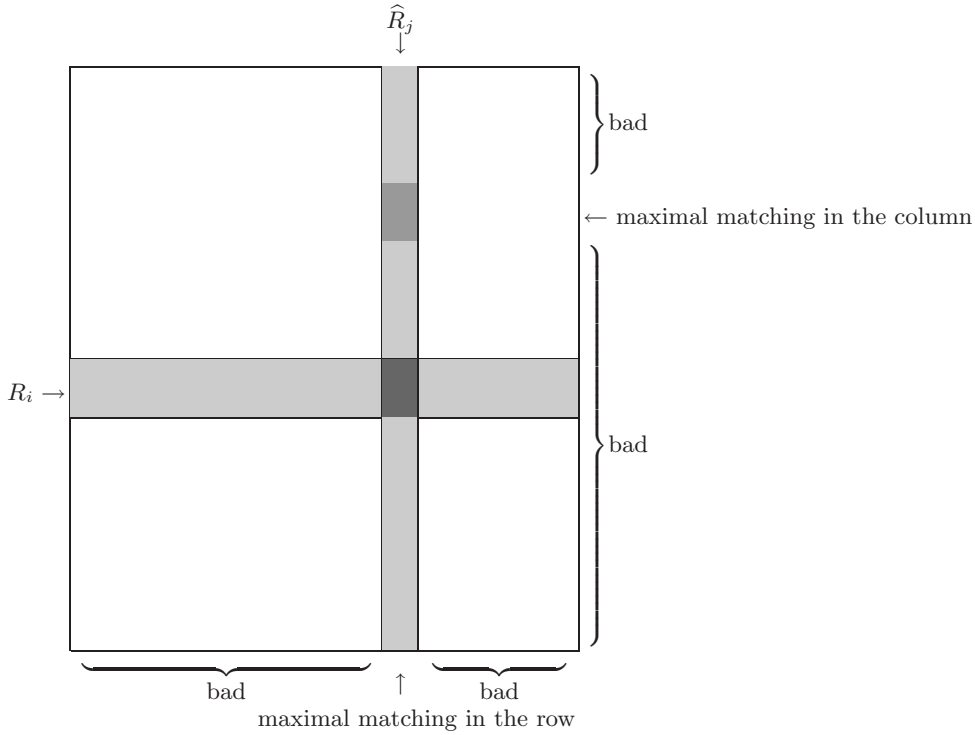


Fig. 4.4. Illustration of the missing and false alarm rates over the confusion matrix in a particular case.

reference region R_i and an output region $\widehat{R}_j$ in e_R^m, $\widehat{R}_j$ need not necessarily have the maximum intersection with R_i when computing e_R^f; as a consequence, what is considered misclassification by one of the measures can be considered a correct classification by the other one. OS and US avoid this difficulty by accumulating the whole region, respectively, R_i and $\widehat{R}_j$, when the over- and under-segmentation is detected, instead of accounting for the pixels not belonging to the maximum intersection, as in e_R^m and e_R^f. This fact can make e_R^m and e_R^f difficult to interpret. Finally, Huang and Dom introduce an overall performance measure p_R from the average between e_R^m and e_R^f, what, given the case of figure 4.4, makes this measure meaningless.

- In [99], Hoover et al. classify every pair of reference R_i and output $\widehat{R}_j$ regions as *correct detections*, *over-segmentation*, *under-segmentation*, *missed* or *noise*, and then build evaluation metrics counting instances of every case. As can be observed, the percentage threshold T used during the classification plays the same role as p in CG, OS and US. Besides, as the authors show in the paper, for $0.5 < T < 1.0$, any region can contribute to up to three classifications, one each of correct detection, over- and under-segmentation. Although the measures are easily interpretable, however, unlike CG, OS and US, the evaluation metrics depend on the number of regions in the image, which makes the metrics dependent, thus, on the image. This fact prevents using a set of test images for getting a global measurement of algorithm performance in terms of misclassification, over- and under-segmentation; that is to say, when comparing several algorithms, it will have to be done on the basis of a single image each time, instead of considering a representative set of images and computing a global performance measure for each.
- Mezaris et al. propose in [169] an overall performance measure whose cost terms take into account, apart from the spatial coincidence between reference and output regions, the distance between misclassified pixels and the corresponding reference region. The output of the evaluation process is an only measure which combines the aforementioned distance and the amount of misclassification, over- and under-segmentation in a single number. All in all, although it is clear that the measure takes lower values as the segmentation output is better and can be used to rank algorithms per-

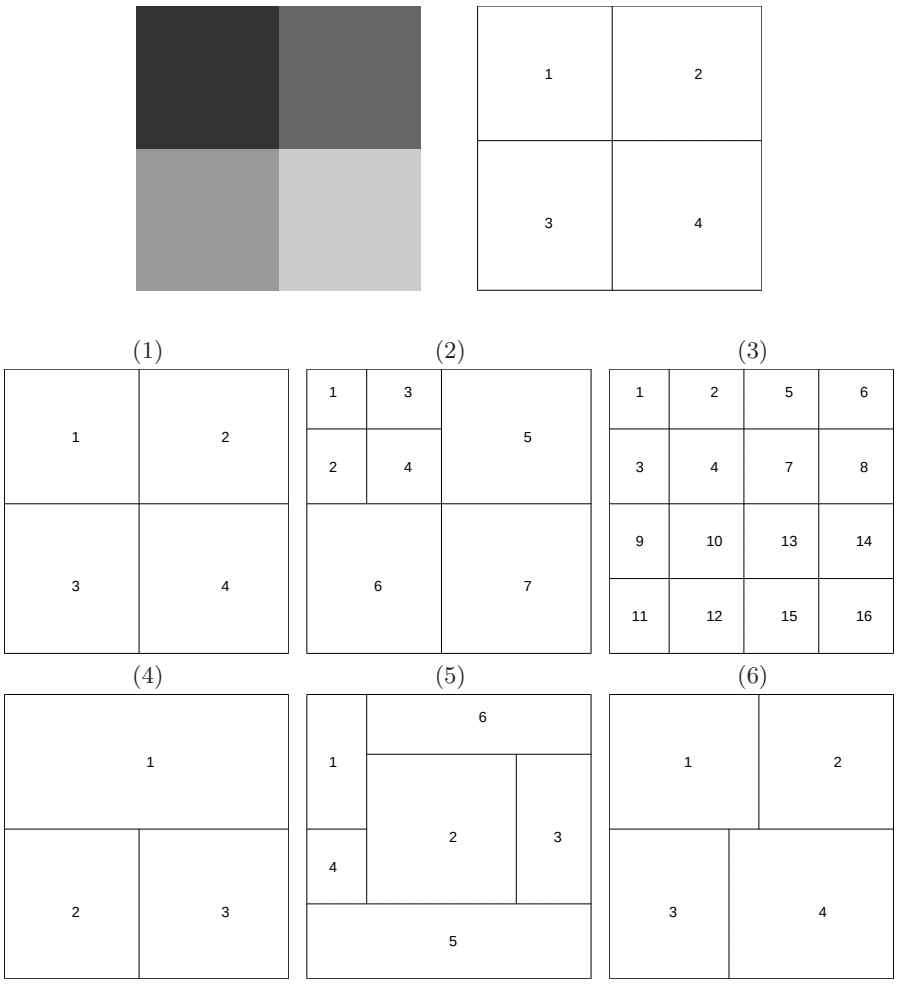


Fig. 4.5. Examples of segmentations and their discrepancy measures (I): [upper row,left] original image (20×20 pixels); [upper row,right] reference contour image; (1)-(6) examples of contour images.

formance, it results difficult to interpret because a lot of information is put together in a single number.

- Finally, Pignalberi et al. define in [229] another cost function combining the cost associated with erroneously segmented pixels together with handicap functions accounting for under- and over-segmentation; this function was intended as a fitness function for tuning the parameters of segmentation algorithms by means of evolutionary techniques. Contrary to the case of Mezaris et al., this time, terms accounting for misclassified pixels, over-segmentation and under-segmentation are clearly defined, although the overall performance measure suffers from the same drawback as in the case of Mezaris et al. Over- and under-segmentation level measurement takes into account whole regions, similarly to OS and US. Nevertheless, a percentage threshold is not included so that a single misclassified pixel can make a large region to be considered over- or under-segmented. Therefore, over- and under-segmentation are greatly penalized.

Discussion over an illustrative example

By way of illustration of CG, OS and US, figure 4.5 shows some examples of segmentation for a very simple image, while the corresponding confusion matrices and the related performance measures appear in, respectively, figure 4.6 and table 4.7. The six cases considered are discussed in the following lines:

$$C_1 = \begin{pmatrix} 100 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 100 \end{pmatrix} \quad C_2 = \begin{pmatrix} 25 & 25 & 25 & 25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 100 \end{pmatrix}$$

$$C_3 = \begin{pmatrix} 25 & 25 & 25 & 25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 25 & 25 & 25 & 25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 25 & 25 & 25 & 25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 25 & 25 & 25 & 25 \end{pmatrix}$$

$$C_4 = \begin{pmatrix} 100 & 0 & 0 \\ 100 & 0 & 0 \\ 0 & 100 & 0 \\ 0 & 0 & 100 \end{pmatrix} \quad C_5 = \begin{pmatrix} 50 & 25 & 0 & 0 & 0 & 25 \\ 0 & 25 & 25 & 0 & 0 & 50 \\ 0 & 25 & 0 & 25 & 50 & 0 \\ 0 & 25 & 25 & 0 & 50 & 0 \end{pmatrix} \quad C_6 = \begin{pmatrix} 100 & 0 & 0 & 0 \\ 10 & 90 & 0 & 0 \\ 0 & 0 & 90 & 10 \\ 0 & 0 & 0 & 100 \end{pmatrix}$$

Fig. 4.6. Examples of segmentations and their discrepancy measures (II): matrices C_1 to C_6 for, respectively, cases (1) to (6) of figure 4.5.

(1) This case corresponds to a perfect segmentation and, accordingly, the percentage of correctly grouped pixels is 100, while the percentages of over-segmentation and under-segmentation are 0.
(2) Now, reference region R_1 appears split into output regions $\widehat{R}_1 - \widehat{R}_4$. None of the regions $\widehat{R}_j$ mix pixels from several true regions R_i, what leads to a CG value of 100, but 25 % of the image is over-segmented.
(3) This is an extreme case where all the reference regions appear split in the segmentation output, what leads to a 100 % of over-segmentation. Pixels are however correctly grouped and, thus, CG = 100.
(4) In this case, two reference regions have been put together and, consequently, a certain degree of under-segmentation must be measured. Since those two regions cover 50 % of the image pixels, the degree of under-segmentation is also 50 %. The remaining 50 % of pixels are correctly grouped.
(5) A more complex situation is presented in this example, where only pixels from regions $\widehat{R}_1$ and $\widehat{R}_4$ are correctly grouped (18.75 %) and the rest belong to regions under-segmenting the image. Since the area of all the reference regions is shared by more than one region $\widehat{R}_j$, the whole image is over-segmented (OS = 100).
(6) Finally, this case shows the effect of parameter p in the discrepancy measurements. As it has been depicted before, p determines the degree of tolerance with regions $\widehat{R}_j$ that all but a reduced set of their pixels fall in a region R_i. In the example, region $\widehat{R}_1$ is almost identical to region R_1 except for a single column of pixels which belong to region R_2 which have been "stolen" to region $\widehat{R}_2$. The same happens between regions $\widehat{R}_4$ and $\widehat{R}_3$. If $p = 100$, the percentage of correctly grouped pixels attains 45 %, corresponding to regions $\widehat{R}_2$ and $\widehat{R}_3$, while for $p = 90$ this percentage goes up to 95 % because only the pixels of $\widehat{R}_1$ which fall into R_2 and the ones of $\widehat{R}_4$ which fall into R_3 are discounted. As for the percentage of over-segmentation, $p = 100$ leads to account in OS for pixels from $\widehat{R}_2$ and $\widehat{R}_3$ as well as the misclassified pixels of $\widehat{R}_1$ and $\widehat{R}_4$, what leads to consider 50 % of the image over-segmented. When $p = 90$, only the misclassified pixels are accounted for (5 % of the image). Observe also that, in this case, CG + US < 100.

In the light of the example, CG, OS and US can be said to produce accurate measures of the, respectively, correct grouping, over-segmentation and under-segmentation present in the segmentation output. Furthermore, parameter p of CG, OS and US provides the desired tolerance to low pixel dispersion among regions.

For comparison purposes, the measures for Huang and Dom [106], Hoover et al. [99] and Mezaris et al. [169] for this example are also provided in table 4.7. Several conclusions can be drawn from the resulting values:

- e_R^m and e_R^f behave exclusively as misclassification rates since they only take into account pixels out of the maximal matchings, not the whole regions involved in over- and under-segmentation. As a

Table 4.7. Examples of segmentations and their discrepancy measures (III): discrepancy measures for figure 4.5. The first group of rows is for CG, OS and US, while the rest of groups are for, respectively, Huang and Dom [106], Hoover et al. [99] and Mezaris et al. [169] measures.

case →	(1)	(2)	(3)	(4)	(5)	(6a)	(6b)
p	100	100	100	100	100	100	90
CG	100.00	100.00	100.00	50.00	18.75	45.00	95.00
OS	0.00	25.00	100.00	0.00	100.00	50.00	0.00
US	0.00	0.00	0.00	50.00	81.25	55.00	0.00
e_R^m	0.00	18.75	75.00	0.00	50.00	5.00	–
e_R^f	0.00	0.00	0.00	25.00	43.75	5.00	–
p_R	100.00	90.63	62.50	87.50	53.13	95.00	–
T	1.00	1.00	1.00	1.00	1.00	1.00	0.90
# CC	4	3	0	2	0	0	4
# OC	0	1	4	0	0	0	0
# UC	0	0	0	1	0	0	0
# MR	0	0	0	2	4	4	0
# NR	0	0	0	0	6	4	0
E	0.0000	0.0260	0.1940	0.0330	0.0955	0.0018	–

consequence, e_R^m and e_R^f measures tend to be more benevolent with the segmentation output with regard to OS and US because the latter invalidate whole regions.

- The measures by Hoover et al. provide performance information closely related to CG, OS and US, but, however, no data are available about whether the detected over- and under-segmentation is significative on the particular image considered. Furthermore, it is worth discussing the results for case (6b) of table 4.7. In this case, thanks to the value given to parameter T, the small dispersions of pixels are tolerated and four correct classifications are found. This is also tolerated by OS(90) and US(90), but, unlike the measures by Hoover et al., CG(90) is able to indicate that there are pixels which are not correctly labeled in the segmentation output.
- Finally, regarding the global performance measure of Mezaris et al., it is worth noting the following:
  (a) very similar values are produced for cases (2) and (4), corresponding to, respectively, over- and under-segmentation cases, although case (4) involves twice more misclassified pixels than case (2);
  (b) case (3), which corresponds to a massive over-segmentation, is ranked twice worse than case (5), which is visually and numerically less correct (look at the percentage of pixels correctly grouped, CG, of both cases).

  As for (a), clearly it is due to the value given to parameter α in F_j (see table 4.6), which penalizes over-segmentation (F_j) over the miss-classification (E_i calculated for set $\mathcal{A}$) and under-segmentation (E_i calculated for set $\mathcal{N}_R$) measures. Regarding (b), since in case (5) both over- and under-segmentation take place simultaneously, it is more difficult to look for the cause of the misbehaviour; nevertheless, it seems to be the same as before. As was already commented, the mixture of spatial coincidence and distance information makes interpretation more difficult.

By way of conclusion:

- It is clear that global performance measures do not provide the level of detail which can be necessary to evaluate correctly the behaviour of segmentation algorithms. Besides, since the different measures are combined in a single expression, a weight has to be chosen for every term, what is clearly not trivial.
- In the same line as the previous point, mixing spatial coincidence information with distance yields performance measures which are difficult to interpret.

Table 4.8. Some measures for evaluating the performance of edge detection algorithms.

measure	formulation	typical values
Pratt's FOM	$FOM(\alpha) = \frac{1}{\max\{n(A), n(B)\}} \sum_{x \in B} \frac{1}{1 + \alpha d(x, A)^2}$	$\alpha = 1/9$
Discrepancy Percentage	$D = \frac{n(A \Delta B)}{n(I)}$	—
Baddeley	$\Delta_w^p(A, B) = \sqrt[p]{\frac{1}{n(I)} \sum_{x \in I} \lvert w(d(x, A)) - w(d(x, B)) \rvert^p}$ $w(t) = \min\{t, c\}$	$p = 2, c = 5$

A is the set of image locations corresponding to edges in the reference edge map
B is the set of image locations corresponding to edges in the edge map under evaluation
I is the set of all image locations
$n(I)$ is the number of elements of set I
$A \Delta B = (A \setminus B) \bigcup (B \setminus A)$ (exclusive-or for sets)
$B \setminus A = \{x \in B | x \notin A\}$ (set minus operator for sets)
$d(x, Y)$ is the distance between image location x and the nearest edge in Y

- The introduction of a tolerance percentage threshold when calculating the percentage of over- and under-segmentation avoids pathological cases such as the one described before involving a single misclassified pixel.
- Nevertheless, the three previous remarks are not intended to discard the measures reviewed in this chapter as well as others not appearing here. However, it seems better to use them as secondary measures providing additional information to CG, OS and US values, as will be done later in chapter 9.

Edge detection evaluation

An important amount of strategies for assessing the quality of an edge map —either coming from an edge detection or an image segmentation task— can be found in the related literature. Among all of them, Baddeley's measure [6] was deemed the best evaluation technique in a recent survey [46], which involved, among others, the often-cited Pratt's *Figure of Merit* (FOM) [232] and the discrepancy percentage, considered in Zhang's survey as one of the best discrepancy measures [239, 306]. As can be observed, all three are based on measuring the distance between edge pixels in the reference edge map and the edge map produced by the algorithm under evaluation. Table 4.8 summarizes the three aforementioned discrepancy-based measures.

As can be observed, while Pratt's FOM scores good edge maps with larger values than bad edge maps, the percentage of discrepancy and Baddeley's measure both attain lower values with better edge maps. On the other hand, Pratt's FOM yields values between 0 and 1, as well as the discrepancy percentage, but Baddeley's scores between 0 and c (see table 4.8).

As a final remark, it is important to notice that Pratt's FOM, and other similar measures such as the *mean error distance*[2] or the *mean square error distance*[3], are mostly sensitive to type I errors (i.e. false positive edges). For example, if there are no type I errors, $B \subseteq A$, then $FOM(\cdot) = n(B)/n(A)$ regardless of the positions of the type II errors. In fact, Baddeley describes FOM as *"a kind of average localization error for the type I errors only"* [6]. Nevertheless, FOM is still used by many authors (see [35, 57, 68, 135, 282, 290] among many others).

[2] $\overline{e} = \frac{1}{n(B)} \sum_{x \in B} d(x, A)$
[3] $\overline{e^2} = \frac{1}{n(B)} \sum_{x \in B} d(x, A)^2$

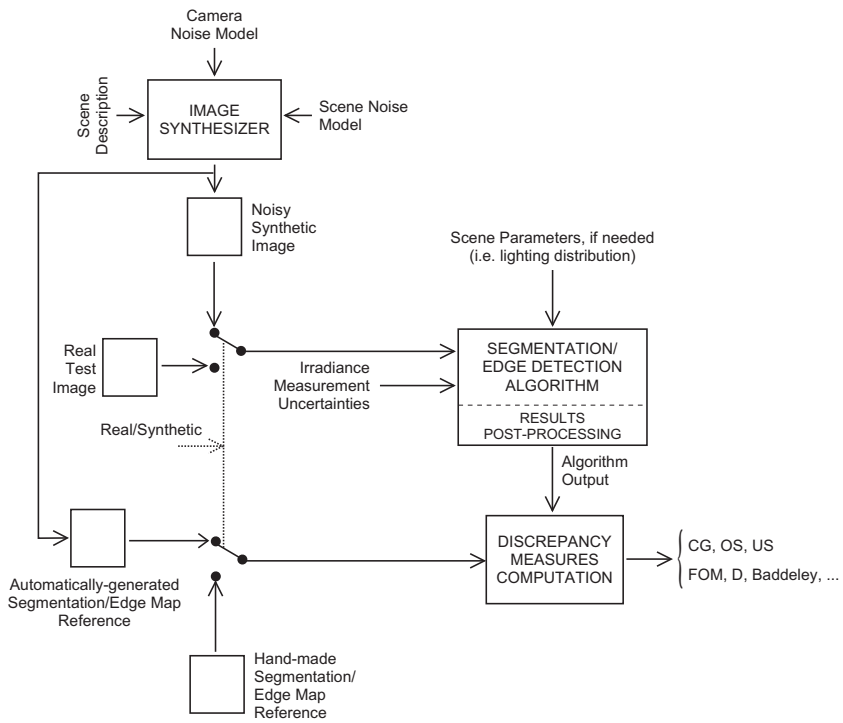


Fig. 4.7. Diagram of the performance evaluation setup for physics-based image segmentation and edge detection.

4.4.3 Performance evaluation setup

By way of conclusion of section 4.4, the diagram of figure 4.7 depicts a possible performance evaluation arrangement. In case synthetic images were to be used, the corresponding image synthesizer should be provided with the camera noise model to generate images with realistic noise. This is another application of calibrating a camera at the radiometric level, and will also be discussed in chapter 6. The reference segmentation for the computation of discrepancy measures would be automatically generated in this case. If a real test image was going to be used instead, the segmentation reference would have to be provided by other means, typically segmenting by hand the test image.

The success of a physics-based vision algorithm depends on the agreement between the images to process and the model of image formation around which the algorithm has been devised. Precisely, this is one of the most interesting features of this sort of vision algorithms, since it is possible to predict under what circumstances the algorithm should produce correct results. However, depending on its design, such an algorithm could tolerate some deviations in the image formation model. This would be an added value of the algorithm and, thus, would be interesting to know about. In figure 4.7, the term *scene noise* is introduced in this sense. From a practical point of view, the occurrence of optical phenomena such as shadows, specularities or interreflections would be examples of deviations in the image formation model if the vision algorithm only expected body reflection; lack of satisfaction of the dichromatic reflection model for some object scenes could be another example. Other deviations can have to do with the lighting (e.g. use of non-uniform lighting, or several light sources) or with the reflectance properties of scene objects (e.g. non-uniform reflectance or textured objects). In figure 4.7, the introduction of scene noise is contemplated either by means of synthetic scenes or directly using real images not covered by the image formation model assumed.

5

Estimation of Lighting Distribution Parameters

Vision algorithms relying on the physics of image formation often assume there is an *a priori* knowledge about the particular imaging infrastructure, mainly the spectral composition of scene illumination together with the geometry of the lighting configuration. Methods aiming at reconstructing objects shape from the shading of their images (mainly *Shape From Shading* methods) are the primary users of this type of data, although methods segmenting an image according to the reflectance of the objects present in the scene may also need them, as it is the case of the algorithm of chapter 7.

The study of the detection of the illumination distribution from images can be traced back to the 1980's. Since then, several light source estimators have been proposed (see [141, 224, 311], among many others, and [301] for a survey and a comparison). The primary motivation behind the work was to extend shape from shading algorithms to operate with less *a priori* information, although only cases with one parallel light source and monochrome images were considered. Research on colour vision [95] also provided some methods for, mainly, estimating the spectral composition of light sources. More recently, researchers have developed techniques for dealing with multiple point light sources, which extend the applicability of the estimations to other areas such as image-based computer graphics or photo-realistic image synthesis (see [19, 248, 289, 309, 312], among the most recent ones).

Assuming a model composed of directional lighting and ambient illumination, this chapter presents several methods using a spherical calibration object to estimate the direction related to directional lighting together with the strengths of both light sources, and even their colour if a colour camera is employed. As most physics-based vision algorithms do, directional lighting is assumed to be parallel, or, in other words, the light source is assumed distant. Concerning the projection of scene points into the image plane, two methods are proposed, one assuming orthographic projection, and the other one assuming perspective projection. Furthermore, both have been derived not only considering the interaction of light with the calibration object, but also accounting for the noise that corrupts digital pixel values in CCD cameras. To this end, the model presented in section 2.4.4 has been incorporated into the estimation methods. Besides, the use of a calibration object with known shape disambiguates the problem of estimating the parameters of the illumination and, since the estimation can be done in a controlled way, undesirable effects such as specularities or inter-reflections can be avoided. An extensive set of experimental results showing the performance of the different methods proposed are given and discussed at the end of the chapter.

The rest of the chapter is organized as follows: sections 5.1 and 5.2 describe several methods for estimating the lighting parameters from both single band and colour images assuming orthographic projection; next, section 5.3 evaluates the error related with the orthographic projection assumption, while section 5.4 extends all the previous methods to perspective projection; previous work is discussed in section 5.5; experimental results are provided in section 5.6; finally, conclusions appear in section 5.7.

5.1 Estimation from Single Band Images

Let us consider a particularization of the expanded image formation model of equations 4.1 and 4.4, consisting of a single light source and ambient lighting illuminating a uniformly coloured matte object,

so that ρ_b is constant and $\rho_i = 0$ and hence $\rho_a = \rho_b$ within the object. If noise is taken into account (equation 2.44), equation 5.1 results for the object projection pixels:

$$D^c(i,j) = K(i,j)(L_a\rho_b)^c A^c + \mu_{dc}(i,j)A^c + m_b(i,j)K(i,j)(L_d\rho_b)^c A^c + N^c(i,j)\,, \tag{5.1}$$

where $(L_a\rho_b)^c$ and $(L_d\rho_b)^c$ are, respectively, the ambient and body composite reflectances C_a^c and C_b^c of the object under study, which were introduced in equations 4.4a and 4.4b. The relevance of equation 5.1 is that it reveals a linear relationship between D^c and m_b for all the object pixels, altered by the noise coming from N^c and the spatial variation in K and μ_{dc}. Therefore, if enough m_b values can be related to the corresponding D^c values, so that some pairs (m_b, D^c) can be obtained, those pairs can be fitted by a straight line $D^c = \alpha^c + m_b\beta^c$, whose parameters, α^c and β^c, contain $(L_a\rho_b)^c$ and $(L_d\rho_b)^c$, respectively, and can, thus, be used to estimate them.

In general, getting rid of noise is a major concern when fitting a curve to a set of points. In this particular case, the noise affecting D^c can be put under control if the following steps are performed:

(1) Take and average several images of the calibration object, to produce $\mu_{D^c}(i,j)$ pixels as a measure to counteract N^c (already discussed in section 4.4.1):

$$\mu_{D^c}(i,j) = K(i,j)(L_a\rho_b)^c A^c + \mu_{dc}(i,j)A^c + m_b(i,j)K(i,j)(L_d\rho_b)^c A^c\,. \tag{5.2}$$

(2) Once N^c has been removed from equation 5.1, the variation in μ_{D^c} for pixels having the same m_b comes exclusively from the PRNU and the DCNU, and the expected intensity value results to be:

$$\mathrm{E}_{m_b}[\mu_{D^c}] = \mathrm{E}_{m_b}[K]\,(L_a\rho_b)^c\,A^c + \mathrm{E}_{m_b}[\mu_{dc}]A^c + m_b\mathrm{E}_{m_b}[K]\,(L_d\rho_b)^c\,A^c\,, \tag{5.3}$$

where $\mathrm{E}_{m_b}[K]$ and $\mathrm{E}_{m_b}[\mu_{dc}]$ are the averages of, respectively, $K(i,j)$ and $\mu_{dc}(i,j)$ for the pixels (i,j) having the m_b under consideration. If enough pixels are involved, $\mathrm{E}_{m_b}[K] \approx 1$ and $\mathrm{E}_{m_b}[\mu_{dc}] \approx \mathrm{E}_I[\mu_{dc}]$, being $\mathrm{E}_I[\mu_{dc}]$ the average of $\mu_{dc}(i,j)$ across the CCD array. This makes $\alpha^c = (L_a\rho_b)^c A^c + \mathrm{E}_I[\mu_{dc}]A^c$ and $\beta^c = (L_d\rho_b)^c A^c$ after fitting the pairs $(m_b, \mathrm{E}_{m_b}[\mu_{D^c}])$. This step copes, thus, with the spatial variation in K and μ_{dc}.

If the matte object under consideration is white ($\rho_b(\lambda) = 1$), once the straight line parameters α^c and β^c are known, $\beta^c = L_d^c A^c$ is related with the radiance of the directional lighting of the scene, while $\alpha^c - \mathrm{E}_I[\mu_{dc}]A^c = L_a^c A^c$ is related with the radiance of the ambient illumination, both for the band of the electromagnetic spectrum corresponding to colour channel c. Notice, however, that both values include, apart from the camera gain, the effect of the lens aperture and the exposure time, since they both influence the charge stored at all collection sites. Because of this, calling lighting strengths to $L_d^c A^c$ and $L_a^c A^c$ makes more sense than using the term radiance, and will be the expression used from now on to refer to them.

The knowledge about $m_b(i,j) = \cos\theta(i,j) = \boldsymbol{n}(i,j)\cdot\boldsymbol{s}(i,j)$ required by the previous procedure implies, as well, knowledge about $\boldsymbol{n}(i,j)$ and $\boldsymbol{s}(i,j)$ for the same set of pixels. Both issues are discussed in turn in the following.

5.1.1 Estimation of surface normal vectors

$\boldsymbol{n}(i,j)$ can be determined if the object shape is known beforehand, which implies the use of a calibration object. Apart from other strategies, some regular objects allow determining the parameters of their particular shape from their projection in the image. This is the case of a sphere under orthographic projection. In effect, since the projection of the sphere is a circle, the projected radius r and center (u_0, v_0) can be obtained by pointing out three points over the circle contour (see figure 5.1). From this information, the surface normal vector at every image pixel $\boldsymbol{n}(i,j)$ is given by equation 5.4: [1]

$$n(i,j) = \begin{pmatrix} \frac{u-u_0}{r} \\ \frac{v-v_0}{r} \\ -\sqrt{1-\left(\frac{u-u_0}{r}\right)^2 - \left(\frac{v-v_0}{r}\right)^2} \end{pmatrix}, \tag{5.4}$$

[1] Observe that the z component of sphere surface normal vectors is negative because a viewer-oriented coordinate system is assumed (section 4.1).

where (u, v) is the projection of the scene point (x, y, z) onto the image plane. Under orthographic projection, (u, v) happen to be equal to (x, y) and to (i, j).

5.1.2 Estimation of the lighting direction

If the light coming from the directional light source is assumed distant, $\boldsymbol{s}(i,j) = (s_x(i,j), s_y(i,j), s_z(i,j))$ turns out to be approximately constant throughout the scene. In this way:

$$\cos\theta(i,j) = \frac{u-u_0}{r}\, s_x + \frac{v-v_0}{r}\, s_y - \sqrt{1 - \left(\frac{u-u_0}{r}\right)^2 - \left(\frac{v-v_0}{r}\right)^2}\, s_z\,. \tag{5.5}$$

$\boldsymbol{s}$ can now be determined from the intensity pattern of the sphere in the image. In effect, obviating the spatial noise for the moment (i.e. $K(i,j) = 1$ and $\mu_{dc}(i,j) = \mathrm{E}_I[\mu_{dc}]$), the isophote curve of intensity L of the average image (i.e. $\mu_{D^c}(i,j) = L$), without considering the background, is given by:

$$L = \alpha^c + \left(\frac{u-u_0}{r}\, s_x + \frac{v-v_0}{r}\, s_y - \sqrt{1 - \left(\frac{u-u_0}{r}\right)^2 - \left(\frac{v-v_0}{r}\right)^2}\, s_z \right) \beta^c\,, \tag{5.6}$$

where $\alpha^c = (L_a\rho_b)^c A^c + \mathrm{E}_I[\mu_{dc}]A^c$ and $\beta^c = (L_d\rho_b)^c A^c$.

Equation 5.7 can now be obtained reordering equation 5.6:

$$\begin{aligned}\left(\frac{u-u_0}{r}\right)^2 (s_x^2 + s_z^2) + \left(\frac{v-v_0}{r}\right)^2 (s_y^2 + s_z^2) + 2\left(\frac{u-u_0}{r}\right)\left(\frac{v-v_0}{r}\right) s_x s_y \\ - 2\mathcal{C}\left(\frac{u-u_0}{r}\right) s_x - 2\mathcal{C}\left(\frac{v-v_0}{r}\right) s_y + \mathcal{C}^2 - s_z^2 = 0\,,\end{aligned} \tag{5.7}$$

where $\mathcal{C} = \frac{L-\alpha^c}{\beta^c}$. If variables p and q are defined to be $p = \frac{u-u_0}{r}$ and $q = \frac{v-v_0}{r}$, then equation 5.7 defines a rotated conic in terms of the variables p and q. [2]

The conic expressed in equation 5.7 has the following properties:

(1) If equation 5.7 is put in the form of equation 5.8:

$$(p\ q)\begin{pmatrix} a & c \\ c & b \end{pmatrix}\begin{pmatrix} p \\ q \end{pmatrix} + 2(d\ e)\begin{pmatrix} p \\ q \end{pmatrix} + f = 0\,, \tag{5.8}$$

then the matrix which defines the quadratic form associated to the conic is given by:

$$\begin{pmatrix} a & c \\ c & b \end{pmatrix} = \begin{pmatrix} s_x^2 + s_z^2 & s_x s_y \\ s_x s_y & s_y^2 + s_z^2 \end{pmatrix}\,, \tag{5.9}$$

whose eigenvalues are $\lambda_1 = s_x^2 + s_y^2 + s_z^2 = 1$ and $\lambda_2 = s_z^2$. Since both are positive (i.e. have the same sign), the conic results to be an ellipse.

(2) Since $\lambda_1 > \lambda_2$ and λ_1 and λ_2 are inversely proportional to the square of the lengths of the ellipse axes [4], λ_1 is always related to the shortest axis.

(3) The coordinates of the center of the ellipse (p_0, q_0) are given by the solution of the system of simultaneous equations 5.10 [251]:

$$\left.\begin{aligned} a\,p_0 + c\,q_0 + d = 0 \\ c\,p_0 + b\,q_0 + e = 0 \end{aligned}\right\} \tag{5.10}$$

which, for the particular values of the ellipse parameters $a, b, \ldots, e$ (equations 5.7 and 5.8), result to be $p_0 = \mathcal{C}s_x$ and $q_0 = \mathcal{C}s_y$.

[2] Note that if, instead of introducing variables p and q, equation 5.7 is reordered in terms of variables u and v, a conic over the plane $\mathcal{U} - \mathcal{V}$ is also found.

(a) (b)

Fig. 5.1. Example of lighting orientation estimation (I): (a) calibration image; (b) circumference defining the sphere projection contour, overimposed over the calibration image (introduced manually by the user).

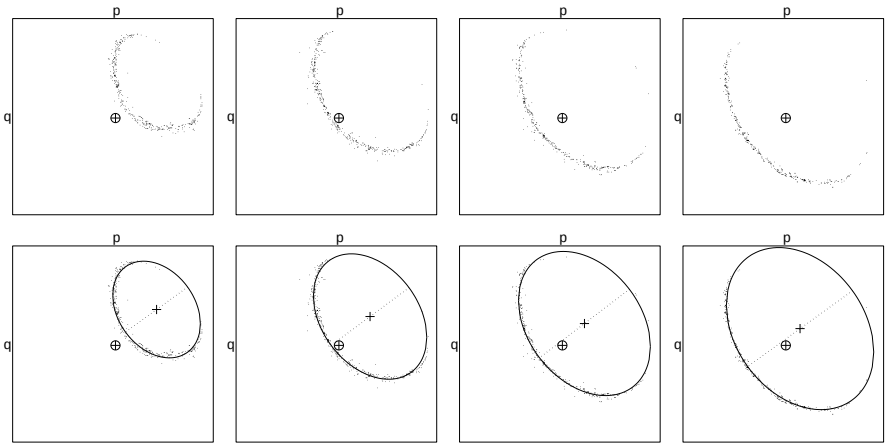


Fig. 5.2. Example of lighting orientation estimation (II): (1st row) from left to right, sets of (p, q) points for which $L = 167, 137, 107$ and 77 in the calibration image shown in figure 5.1; (2nd row) ellipses fitting the above sets of points. (In all the plots the $\oplus$ sign represents the origin of coordinates.)

(4) Besides, the slopes m_1 and m_2 of the axes of the ellipse are the roots of equation 5.11 [251]:

$$c\,m^2 + (a - b)m - c = 0\,, \tag{5.11}$$

which, for the particular values found for a, b and c, result to be $m_1 = \frac{s_y}{s_x}$, for the shortest axis (i.e. the one related to λ_1), and $m_2 = -\frac{s_x}{s_y}$, for the largest axis.

(5) Using properties (3) and (4), the expression, in parametric form, corresponding to the straight line containing the shortest axis of every ellipse results to be $(p, q) = \mathcal{C}(s_x, s_y) + t(s_x, s_y)$, where t is the parameter.

(6) Finally, the centers of the ellipses change from one isophote curve to the next since they all depend on $\mathcal{C}$, but the straight lines containing the corresponding shortest axes are co-linear and can be expressed as $(p, q) = t'(s_x, s_y)$. All the centers are thus aligned over this straight line, which in particular contains the origin of coordinates, i.e. if $t' = 0$ then $(p, q) = (0, 0)$.

Summing up the previous properties, the different isophote curves of the sphere projection define ellipses over the plane $\mathcal{P} - \mathcal{Q}$ whose centers and shortest axes all lie on the straight line passing through the origin of coordinates $(0, 0)$ and have slope $\frac{s_y}{s_x}$. [3]

[3] Regarding the plane $\mathcal{U} - \mathcal{V}$, the slope of the straight line containing the centers and the shortest axes point towards the center of the sphere projection (u_0, v_0).

By way of illustration of the aforementioned, the first row of figure 5.2 shows the sets of points of the calibration image of figure 5.1(a) for which L is, from left to right, 167, 137, 107 and 77. As can be observed, those sets clearly exhibit an elliptic shape. The fitting ellipses are given in the second row of figure 5.2. Although any conic-fitting method can be of use in this stage, it must be taken into account that the conic to fit is always an ellipse and that, depending on the light orientation, part of the ellipse lies in the hidden part of the sphere (i.e. the ellipse can appear partially occluded). Because of its performance against occlusions [49], the ellipse-fitting-specific method by Fitzgibbon et al. [50] is preferred.

The first row of figure 5.2 is also useful to illustrate the effect of noise in the isophote curves corresponding to the calibration object. Because noise makes pixels to "jump" from their original intensity level L' to a different one L'', in the noisy image, the set of pixels having, say, intensity L mix pixels from different original intensities and, therefore, isophotes. As a result, the isophotes of the noisy image consist of elliptical clouds of separated points instead of more or less continuous curves, which complicates estimating the correct ellipse in each case. However, the set of pixels with intensities in the range $[L - \Delta/2, L + \Delta/2]$, for Δ small, is likely to contain all the pixels originally belonging to the isophote of intensity L, plus some other pixels corresponding to isophotes of adjacent intensities in the noiseless image. Experimental results have shown, however, that for $\Delta = 4$ the contaminating pixels do not affect the fitting so as to significantly distort the ellipse parameters. Nevertheless, the sets of points of figure 5.2 correspond to $\Delta = 2$.

Now, on the basis of the previous properties, at least the two following methods are of application for computing the tilt τ and slant σ of s, by which $s = (\sin\sigma\cos\tau, \sin\sigma\sin\tau, \cos\sigma)$:

(1) *TEAV method (Tilt Estimation by AVeraging).* Given λ_1 and λ_2 coming from the ellipse fitting a particular isophote curve, $s_z = -\sqrt{\frac{\lambda_2}{\lambda_1}}$ [4] and $\sigma = \cos^{-1} s_z$. On the other hand, given the corresponding (p_0, q_0) or m_1, $\frac{q_0}{p_0} = m_1 = \frac{s_y}{s_x} = \tan\tau$, being determined the precise quadrant of τ by the signs of $(p_0, q_0) = (\mathcal{C}s_x, \mathcal{C}s_y)$, since $\mathcal{C} = \frac{L-\alpha^c}{\beta^c} \geq 0$.
Theoretically, the isophote curves corresponding to all the possible intensities L should all provide the same estimation of $\boldsymbol{s}$. However, due to image noise and despite the fusion of several isophotes of the noisy image to estimate $\boldsymbol{s}$, typically one has as many estimates $\widehat{s}$ as isophote curves have been examined. Therefore, a problem of aggregation of the different $\widehat{s}^L$'s to get a final vector arises. In TEAV, the tilt and slant angles are determined by averaging the different estimations $\widehat{\tau}^L$ and $\widehat{\sigma}^L$ computed for every L as indicated before. Although the different $\widehat{s}^L$'s tend to be quite close to one another, the possibility of outliers should not be discarded and means should be provided to overcome them. Solutions such as the Minimum Volume Ellipsoid (MVE) estimator [241] or a Mean-Shift based technique [27, 30] would be useful at this point.

(2) *TECA method (Tilt Estimation by Centers Alignment).* From a theoretical point of view, all the ellipse centers (p_0^L, q_0^L) lie on a straight line passing through the origin whose slope coincides with the tilt of s. On the basis of this property, this second method estimates first all the ellipse centers and then fits them by a straight line forced to pass through point $(0, 0)$. Once this line is known, the ellipse centers are projected onto the fitted line and the sphere projection isophote curves are again fitted by an ellipse, but, now, constraining their centers to be the projected ones, to ensure all of them effectively lie on a common line, while the slope of the shortest axis is also imposed to be the slope of the straight line found. To this end, this knowledge about the ellipse is introduced into the equation of a rotated non-centered ellipse. According to the formulation of table 5.1, $\tan\phi = \frac{q_0}{p_0}$, the slope of the shortest axis. Now, using $1 + \tan^2\phi = \sec^2\phi$:

[4] Observe that, because almost all curve fitting techniques perform the fitting solving an eigenvalue problem [310], the curve parameters are obtained scaled so that $\sqrt{a^2 + b^2 + ... + f^2} = 1$. Consequently, in this particular case, s_z cannot be determined directly from λ_2.

Table 5.1. Expression corresponding to a non-centered ellipse whose axes are a and b units in length, its center is (p_0, q_0) and the axis of length a is rotated ϕ degrees. ($c\phi$ and $s\phi$ stand, respectively, for $\cos\phi$ and $\sin\phi$.)

NON-ROTATED ELLIPSE:

$$\left(\tfrac{p-p_0}{a}\right)^2 + \left(\tfrac{q-q_0}{b}\right)^2 = 1$$
$$\Downarrow$$
$$(p-p_0, q-q_0)\begin{pmatrix} \frac{1}{a^2} & 0 \\ 0 & \frac{1}{b^2} \end{pmatrix}\begin{pmatrix} p-p_0 \\ q-q_0 \end{pmatrix} = 1$$

ROTATED ELLIPSE: (an angle ϕ with respect to P axis)

$$(p-p_0, q-q_0)\begin{pmatrix} c\phi & -s\phi \\ s\phi & c\phi \end{pmatrix}\begin{pmatrix} \frac{1}{a^2} & 0 \\ 0 & \frac{1}{b^2} \end{pmatrix}\begin{pmatrix} c\phi & s\phi \\ -s\phi & c\phi \end{pmatrix}\begin{pmatrix} p-p_0 \\ q-q_0 \end{pmatrix} = 1$$
$$\Downarrow$$
$$(p-p_0)^2\left(\tfrac{c^2\phi}{a^2} + \tfrac{s^2\phi}{b^2}\right) + 2\,(p-p_0)\,(q-q_0)\left(\tfrac{s\phi c\phi}{a^2} - \tfrac{s\phi c\phi}{b^2}\right) + (q-q_0)^2\left(\tfrac{s^2\phi}{a^2} + \tfrac{c^2\phi}{b^2}\right) = 1$$

$$\cos^2\phi = \frac{p_0^2}{p_0^2 + q_0^2} \tag{5.12}$$

$$\sin^2\phi = \frac{q_0^2}{p_0^2 + q_0^2} \tag{5.13}$$

$$\sin\phi\cos\phi = \frac{p_0 q_0}{p_0^2 + q_0^2} \tag{5.14}$$

After the corresponding substitutions and some reordering, equation 5.15 is obtained, from which $\frac{1}{a^2}$ and $\frac{1}{b^2}$ can be derived for every isophote curve by least squares, using the corresponding coordinates (p, q) of the points of the isophote curve. Since the squares of the axes lengths a^2 and b^2 are inversely proportional to, respectively, λ_1 and λ_2, $s_z = -\sqrt{\frac{\lambda_2}{\lambda_1}} = -\sqrt{\frac{|K/b^2|}{|K/a^2|}} = -\sqrt{\frac{a^2}{b^2}}$ (K is a constant which derives from the parameters of the ellipse [4]). However, in order to get a correct solution, values $\frac{1}{a^2}$ and $\frac{1}{b^2}$ should be constrained to be positive. To this end, non-negative least squares [138] could be used.

$$\left(\frac{1}{a^2}\right)\overbrace{\left[\frac{p_0^2}{p_0^2+q_0^2}(p-p_0)^2 + 2\frac{p_0 q_0}{p_0^2+q_0^2}(p-p_0)(q-q_0) + \frac{q_0^2}{p_0^2+q_0^2}(q-q_0)^2\right]}^{P}$$
$$+\left(\frac{1}{b^2}\right)\overbrace{\left[\frac{q_0^2}{p_0^2+q_0^2}(p-p_0)^2 - 2\frac{p_0 q_0}{p_0^2+q_0^2}(p-p_0)(q-q_0) + \frac{p_0^2}{p_0^2+q_0^2}(q-q_0)^2\right]}^{Q} = 1\,. \tag{5.15}$$

A simpler solution consists in transforming equation 5.15 into equation 5.16, using now variables P and Q as defined in equation 5.15, and determining the slope $-\frac{a^2}{b^2}$ of the resulting straight line.

$$P = -\frac{a^2}{b^2}\,Q + a^2\,. \tag{5.16}$$

To this end, points (P, Q) can be fitted to the straight line to which the sum of orthogonal distances d is minimum. According to figure 5.3, given $\boldsymbol{n}$ as the unit normal vector of the considered straight line:

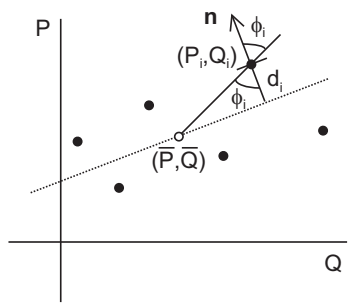


Fig. 5.3. Straight line fitting using orthogonal distances. In the figure, $\boldsymbol{n}$ is the vector normal to the straight line, while $(\overline{P}, \overline{Q})$ is the center of mass of the cloud of points.

$$\begin{aligned}
\min \sum_i d_i^2 &= \min \sum_i \left(\|(P_i, Q_i) - (\overline{P}, \overline{Q})\| \cos \phi_i \right)^2 \\
&= \min \sum_i \left(\left[(P_i, Q_i) - (\overline{P}, \overline{Q}) \right] \boldsymbol{n} \right)^2 \\
&= \min \boldsymbol{n}^T \begin{pmatrix} P_1 - \overline{P} & Q_1 - \overline{Q} \\ \vdots & \vdots \\ P_M - \overline{P} & Q_M - \overline{Q} \end{pmatrix}^T \begin{pmatrix} P_1 - \overline{P} & Q_1 - \overline{Q} \\ \vdots & \vdots \\ P_M - \overline{P} & Q_M - \overline{Q} \end{pmatrix} \boldsymbol{n} \\
&= \min \boldsymbol{n}^T S \boldsymbol{n} \,. \qquad (5.17)
\end{aligned}$$

Since the matrix S is real, symmetric and, in general, positive-definite, it is diagonalizable, all the eigenvalues are real and the corresponding eigenvectors are orthogonal [33]. Consequently, the solution to equation 5.17 is the lowest eigenvalue of S, because $\boldsymbol{n}^T S \boldsymbol{n} = \boldsymbol{n}^T \lambda \boldsymbol{n} = \lambda$, provided $\boldsymbol{n}$ is a unit vector; hence, $\boldsymbol{n} = (n_P, n_Q)$ is the eigenvector of S associated to the lowest eigenvalue. Accordingly, the vector $\boldsymbol{v} = (v_P, v_Q)$ in the same direction as the straight line is the eigenvector of S related to the largest eigenvalue. However, since the slope $\frac{v_P}{v_Q}$ of the fitting straight line must satisfy that $\frac{v_P}{v_Q} = -\frac{a^2}{b^2} < 0$, both components of vector $\boldsymbol{n}$ must have the same sign. That is to say, the problem posed through equation 5.17 must be completed in the following way:

$$\begin{gathered} \min \boldsymbol{n}^T S \boldsymbol{n} \\ \text{constrained by } n_P n_Q = k \geq 0 \end{gathered} \qquad (5.18)$$

where the product $n_P n_Q$ can be expressed in matrix form as follows:

$$n_P n_Q = \boldsymbol{n}^T \overbrace{\begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}}^{C} \boldsymbol{n} \,. \qquad (5.19)$$

In this way, the problem can be solved introducing the Lagrange multiplier λ and differentiating the Lagrangian function of equation 5.20 to reach the system of simultaneous equations 5.21.

$$\mathcal{L} = \boldsymbol{n}^T S \boldsymbol{n} + \lambda \left(\boldsymbol{n}^T C \boldsymbol{n} - k \right) \qquad (5.20)$$

$$S\boldsymbol{n} = \overbrace{(-\lambda)}^{\nu} C\boldsymbol{n} \qquad (5.21a)$$

$$\boldsymbol{n}^T C \boldsymbol{n} = k \geq 0 \qquad (5.21b)$$

Substituting $C\boldsymbol{n}$ from equation 5.21a into equation 5.21b, the latter becomes into $\frac{1}{\nu}\boldsymbol{n}^T S \boldsymbol{n} = k$. Since S is positive-definite, $\boldsymbol{n}^T S \boldsymbol{n} > 0, \forall \boldsymbol{n}$, and, therefore, only strictly positive eigenvalues ν of the generalized eigenproblem of equation 5.21a are solution of the initial problem. Because of the

Table 5.2. Example of lighting orientation estimation (III): (a) intermediate and final results for the estimation method based on the direct computation of $\boldsymbol{s}$ from the fitting ellipses (TEAV method); (b) intermediate and final results for the estimation method based on the computation of $\boldsymbol{s}$ from corrected ellipses (TECA method).

(a)

L	a	b	c	d	e	f	$\widehat{\boldsymbol{s}}^L$
167	+0.66	+0.54	+0.18	-0.37	-0.30	+0.15	(+0.56,+0.40,-0.73)
137	+0.69	+0.56	+0.21	-0.31	-0.25	-0.06	(+0.57,+0.43,-0.70)
107	+0.69	+0.58	+0.20	-0.22	-0.19	-0.24	(+0.56,+0.43,-0.71)
77	+0.68	+0.57	+0.18	-0.14	-0.13	-0.38	(+0.55,+0.40,-0.73)
							lighting direction → (+0.56,+0.41,-0.72)

(b)

L	a	b	c	d	e	f	$\widehat{p}_0^L$	$\widehat{q}_0^L$
167	+0.66	+0.54	+0.18	-0.37	-0.30	+0.15	+0.46	+0.40
137	+0.69	+0.56	+0.21	-0.31	-0.25	-0.06	+0.35	+0.32
107	+0.69	+0.58	+0.20	-0.22	-0.19	-0.24	+0.25	+0.24
77	+0.68	+0.57	+0.18	-0.14	-0.13	-0.38	+0.16	+0.19
							$\widehat{\tau}^L$ → +42.803646 °	

L	a	b	c	d	e	f	$\widehat{\sigma}^L$	$\widehat{\boldsymbol{s}}^L$
167	+0.61	+0.59	+0.18	-0.35	-0.32	+0.14	+137.676178 °	(+0.49,+0.46,-0.74)
137	+0.65	+0.61	+0.21	-0.29	-0.27	-0.06	+135.184841 °	(+0.52,+0.48,-0.71)
107	+0.65	+0.62	+0.21	-0.22	-0.20	-0.24	+135.396176 °	(+0.52,+0.48,-0.71)
77	+0.64	+0.61	+0.18	-0.14	-0.13	-0.38	+138.086710 °	(+0.49,+0.45,-0.74)
								lighting direction → (+0.50,+0.47,-0.73)

symmetry of S and C and the positive-definiteness of S, the signs of the eigenvalues of matrix C coincide with the signs of the eigenvalues of the generalized eigenproblem of equation 5.21a, according to a lemma by Fitzgibbon et al. [50]. Since the eigenvalues of C are -0.5 and 0.5, the problem posed through equation 5.18 is guaranteed to have always solution. Solving equation 5.21a through its characteristic polynomial, the corresponding eigenvalues turn out to be $\nu_1 = 2\left(S_{12} + \sqrt{S_{11}S_{22}}\right)$ and $\nu_2 = 2\left(S_{12} - \sqrt{S_{11}S_{22}}\right)$, where only $\nu_1 > 0$. Substituting ν_1 into equation 5.21a, the normal vector results to be $\boldsymbol{n} = \left(\sqrt{\frac{S_{22}}{S_{11}+S_{22}}}, \sqrt{\frac{S_{11}}{S_{11}+S_{22}}}\right)$, so that $\boldsymbol{v} = \left(-\sqrt{\frac{S_{11}}{S_{11}+S_{22}}}, \sqrt{\frac{S_{22}}{S_{11}+S_{22}}}\right)$ and $\frac{a^2}{b^2} = \sqrt{\frac{S_{11}}{S_{22}}}$.

Similarly to TEAV, every isophote curve L provides a possibly different estimation $\widehat{s}_z^L$ of s_z. Likewise, taking the average of the individual $\widehat{s}_z^L$'s, with special care of outliers, is suggested.

To illustrate both methods, the lighting orientations $\widehat{\boldsymbol{s}}^L$ estimated from everyone of the sets of points presented in the upper row of figure 5.4 are given in the corresponding row of tables 5.2(a) and 5.2(b), while the lower rows of each table gives the final orientation. The second row of figure 5.4 shows the ellipses directly fitting the sets of points given in the first row. The third row of the same image shows the ellipses corrected by aligning their centers. The straight line fitting those centers is given in figure 5.5(a), as well as the alignments of the original centers. Finally, figure 5.5(b) compares both sequences of ellipses.

At this point, a brief discussion about the goodness of both methods is noteworthy. On the one hand, note that TEAV depends more heavily on the reliability of the ellipse fitting method, since *a priori* there is no guarantee that either the ellipse axes or the ellipse centers appear aligned with one another; the final aggregation of tilt and slant angles by clustering/averaging is, thus, compulsory in order to obtain a final estimation of $\boldsymbol{s}$. On the other hand, the second approach takes specifically into account the alignment of the different ellipses. The fitting of the corresponding straight line will,

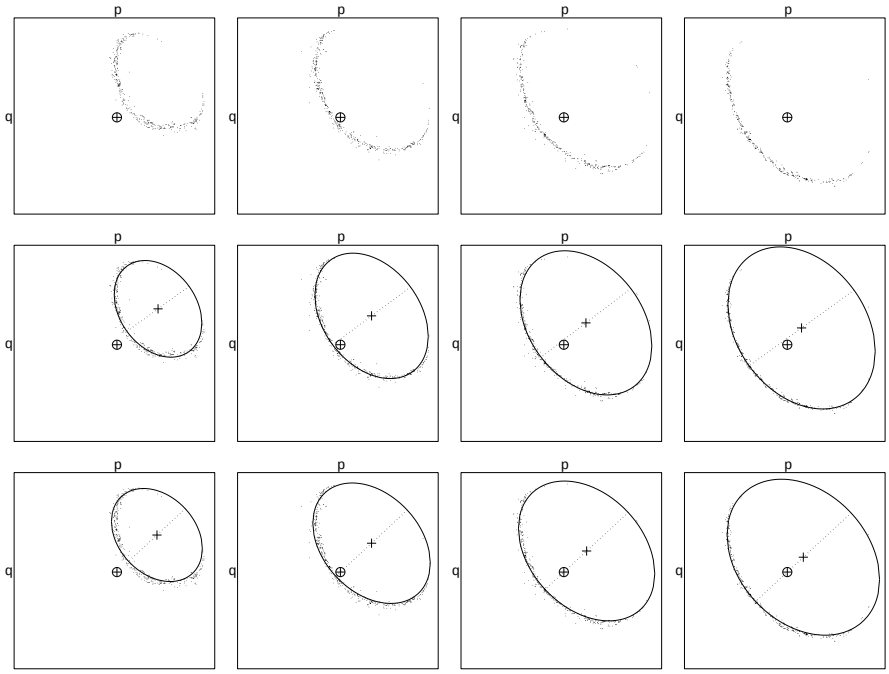


Fig. 5.4. Example of lighting orientation estimation (IV): (1st row) from left to right, sets of (p, q) points for which $L = 167, 137, 107$ and 77 in the calibration image shown in figure 5.1; (2nd row) ellipses directly fitting the above sets of points (TEAV method); (3rd row) corrected ellipses on the basis of the alignment of their centers (TECA method). (In all the plots the $\oplus$ sign represents the origin of coordinates.)

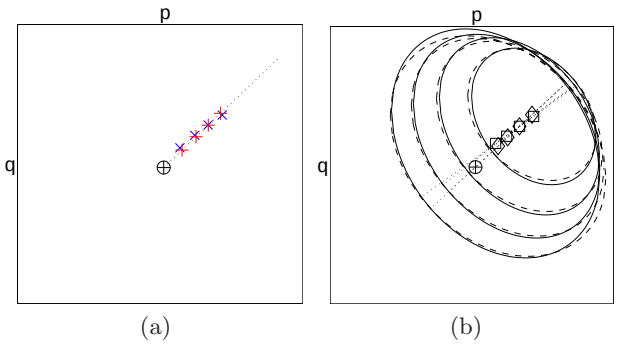


Fig. 5.5. Example of lighting orientation estimation (V): (a) centers of the ellipses shown in table 5.2 ($\times$), straight line passing through $(0, 0)$ fitting them (dotted line) and projection of ellipse centers over the fitting line ($+$); (b) sequence of direct-fitting ellipses (solid line, centers appear as $\square$) and corrected ellipses (dashed line, centers appear as $\diamond$).

however, be better or worse conditioned depending on how concentrated the ellipse centers are. None of the methods seems to be the best one beforehand. What is true is that their output need not coincide in the case of real noisy images, as can be deduced from the results given in table 5.2, where, only considering four values of L, the final orientations estimated by both methods differ in $4.50°$.

5.1.3 Final considerations

Once $\boldsymbol{s}$ is known and $\boldsymbol{n}(i, j)$ has been computed for every pixel within the circle enclosed by the circumference defined by r and (u_0, v_0), using equation 5.4, $m_b(i, j)$ can be accordingly estimated, as

well as the lighting strengths $L_a^c A^c$ and $L_d^c A^c$ by fitting the corresponding pairs $(m_b, \mathrm{E}_{m_b}[\mu_{D^c}])$ by a straight line. Since the values of r and (u_0, v_0) determine the value of $\boldsymbol{s}$ and $\boldsymbol{n}(i,j)$, and consequently of $m_b(i,j)$, the accuracy of the estimation of the lighting parameters depends on how well the sphere projection data have been calculated. In fact, if they are obtained by means of three points over the contour of the sphere projection, as suggested at the beginning of this section, it can so happen that several executions of the calibration method over the same image yield different values for $\boldsymbol{s}$, $L_a^c A^c$ and $L_d^c A^c$ because the user was not able to select all times the same circumference. This is because of, on the one hand, the spatial discretization inherent to real cameras and, on the other hand, the associated spatial aliasing. Both facts make uncertain where the real contour lies, what suggests using a sort of optimization strategy by which those parameters leading to the best estimation are chosen. Given the fact that, in the end, the sphere parameters are needed to estimate the lighting strengths through a linear model which relates m_b with the observations $\mathrm{E}_{m_b}[\mu_{D^c}]$, it seems reasonable to look for the sphere parameters leading to the lowest fitting error to the model. Accordingly, the user is asked to point out two circumferences so that one is completely included into the other and the right contour is most likely to be among both. A sample of all the circumferences among them are then evaluated by means of an elitist genetic optimization strategy [170,171], from which the circumference leading to the best fitting is kept. That is to say, from an initial guess about where the right circular contour lies, several other circumferences are generated using Gaussian mutation from the best fitting at the previous iteration, ensuring that those new circumferences are within the limits determined by the limiting circumferences. This iterative process is stopped after a certain number of generations without improvement in the fitting error.

5.2 Estimation from Colour Images

When several colour channels are available, applying the procedure described in section 5.1 to every channel separately gives rise to several estimations of vector $\boldsymbol{s}$ and an estimation of $L_a^c A^c$ and $L_d^c A^c$ for every colour channel. Since $\boldsymbol{s}$ should be the same for all the colour channels, the estimations $\widehat{\boldsymbol{s}}^{L,c}$ corresponding to the different colour channels should all contribute to the computation of the final $\widehat{\boldsymbol{s}}$, instead of aggregating over every channel separately, irrespectively of the use of TEAV or TECA methods.

On the other hand, the availability of colour data gives rise to several groups of pairs $(m_b, \mathrm{E}_{m_b}[\mu_{D^c}])$, one for every colour channel, which are fitted to determine the illumination strengths within every colour band. As a consequence, a number of fitting errors appear, as many as colour channels. In order to ensure the best matching with the theoretical linear relationships between m_b and $\mathrm{E}_{m_b}[\mu_{D^c}]$ for all c, the objective function is redefined as the maximum of the fitting errors of the different colour channels available. In this way, the same optimization procedure described in section 5.1 can also be applied to colour calibration images.

5.3 Error of Assuming Orthographic Projection

The method presented in section 5.1 assumes orthographic projection for estimating the lighting parameters. This assumption affects the computation of the surface normal vectors of the sphere and, therefore, becomes crucial as for the validity of the estimations which can be obtained. In order to throw some light about the sort of error this assumption can lead to, this section deals with the deviation which can arise between the true normal vector $\boldsymbol{n} = \left(\frac{x-x0}{R}, \frac{y-y0}{R}, \frac{z-z_0}{R}\right)$, for a sphere with parameters $(x_0, y_0, z_0; R)$, and the approximation $\widehat{\boldsymbol{n}} = \left(\frac{u-u_0}{r}, \frac{v-v_0}{r}, -\sqrt{1-\left(\frac{u-u_0}{r}\right)^2 - \left(\frac{v-v_0}{r}\right)^2}\right)$.

Assuming a pin-hole camera, the projection (u,v) of a point in the sphere surface $(x, y, z(x,y))$ onto the image plane is given by:

$$(u,v) = (x,y)\,\frac{f}{z(x,y)}\,, \tag{5.22}$$

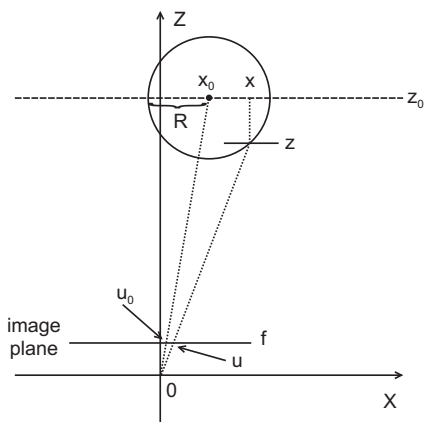


Fig. 5.6. Projection of sphere points under perspective.

where f is the focal length of the camera. Accordingly, the projection of the center of the sphere is given by $\left(u_0 = x_0 \frac{f}{z_0}, v_0 = y_0 \frac{f}{z_0}\right)$ (see figure 5.6).

Taking as a measure of the error the angle δ between $\boldsymbol{n}$ and $\widehat{\boldsymbol{n}}$:

$$\delta = \cos^{-1} \boldsymbol{n} \cdot \widehat{\boldsymbol{n}} = \cos^{-1} \left(n_x \widehat{n}_x + n_y \widehat{n}_y + n_z \widehat{n}_z\right) , \tag{5.23}$$

since both are unit vectors. In the following, equations 5.24, 5.25 and 5.26 develop expressions for the different terms in the sum of equation 5.23:

$$n_x \widehat{n}_x = \left(\frac{x - x_0}{R}\right)\left(\frac{u - u_0}{r}\right) = \left(\frac{x - x_0}{R}\right)\left(\frac{x\frac{z_0}{z} - x_0}{R}\right) \tag{5.24}$$

$$n_y \widehat{n}_y = \left(\frac{y - y_0}{R}\right)\left(\frac{v - v_0}{r}\right) = \left(\frac{y - y_0}{R}\right)\left(\frac{y\frac{z_0}{z} - y_0}{R}\right) \tag{5.25}$$

$$\begin{aligned} n_z \widehat{n}_z &= \sqrt{1 - n_x^2 - n_y^2}\sqrt{1 - \widehat{n}_x^2 - \widehat{n}_y^2} \\ &= \sqrt{1 - \left(\frac{x - x_0}{R}\right)^2 - \left(\frac{y - y_0}{R}\right)^2}\sqrt{1 - \left(\frac{x\frac{z_0}{z} - x_0}{R}\right)^2 - \left(\frac{y\frac{z_0}{z} - y_0}{R}\right)^2} \end{aligned} \tag{5.26}$$

In view of equations 5.24-5.26, it is clear that δ depends on the parameters defining the calibration sphere (x_0, y_0, z_0) and R, but not on the imaging sensor, say the focal distance f. Besides, the magnitude of the error depends on the precise point of the sphere under consideration. In particular, notice that if $x = 0$ and $y = 0$ or $z = z_0$, then $\boldsymbol{n} = \widehat{\boldsymbol{n}}$ and $\delta = 0$. In general, the error is minimum at image points within the sphere projection closest to the image center and at the border of the sphere projection.

Figure 5.7 shows some plots for the maximum and average δ across the sphere, varying R, z_0, x_0 alone and $x_0 = y_0$ simultaneously (a plot for y_0 alone is not included since it would be identical to the one for x_0 alone due to the symmetry of the sphere). In all cases, a parameter is varied and the rest keep constant to $R = 0.05$ m, $z_0 = 3$ m, $x_0 = 0$ m and $y_0 = 0$ m. As can be seen, angle δ increases as the camera approaches the sphere (z_0 decreases), the size of the sphere increases (R is enlarged) or the center of the sphere moves away from the center of the image plane (x_0 and/or y_0 shifts from the image plane center).

Although angle δ is important for measuring the deviations in the approximated surface normal vectors, what is most relevant is their effect over $\cos\theta = \boldsymbol{n} \cdot \boldsymbol{s}$. To evaluate $\Delta \cos\theta = \cos\theta - \cos\widehat{\theta} = (\boldsymbol{n} - \widehat{\boldsymbol{n}}) \cdot \boldsymbol{s}$, $\boldsymbol{s}$ must be taken into account. The plots of figure 5.8 show, for the same sphere parameters as for figure 5.7 and as a function of the tilt τ of $\boldsymbol{s}$, the percentage of sphere pixels for which the relative error $\frac{|\Delta \cos\theta|}{\cos\theta} \times 100$ is below 5%. As can be observed, the error in the estimation of $\cos\theta$ does not seem to change much as the tilt τ varies. As already discussed before, the smaller the sphere, the closer its center to the optical axis and the farther away, the better for the estimation of $\cos\theta$.

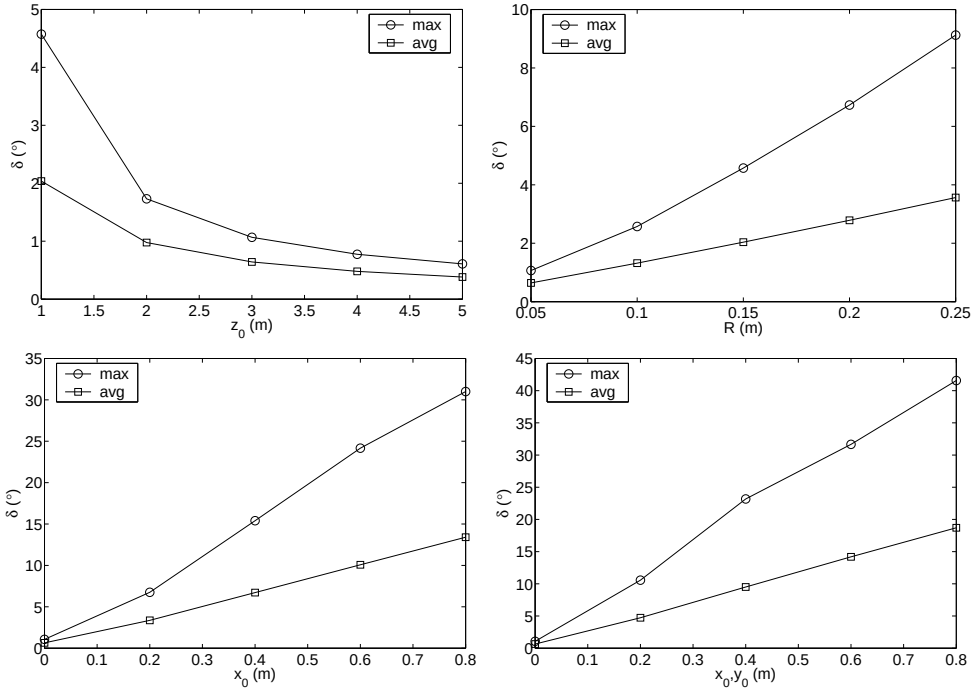


Fig. 5.7. Angle δ between true and approximated surface normal vectors, $\boldsymbol{n}$ and $\widehat{\boldsymbol{n}}$. Circles correspond to maximum angle across the whole sphere, while squares correspond to mean angle.

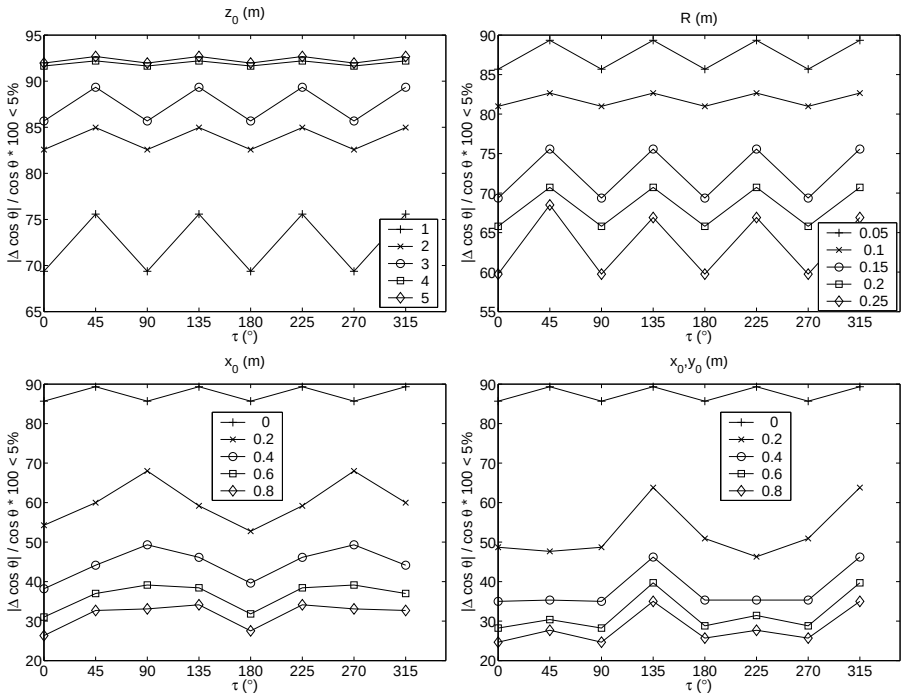


Fig. 5.8. Percentage of sphere points for which the relative error $\frac{|\Delta \cos\theta|}{\cos\theta} \times 100$ is below 5% for different values of the calibration object parameters, (x_0, y_0, z_0) and R, and the tilt τ of the illumination direction $\boldsymbol{s}$.

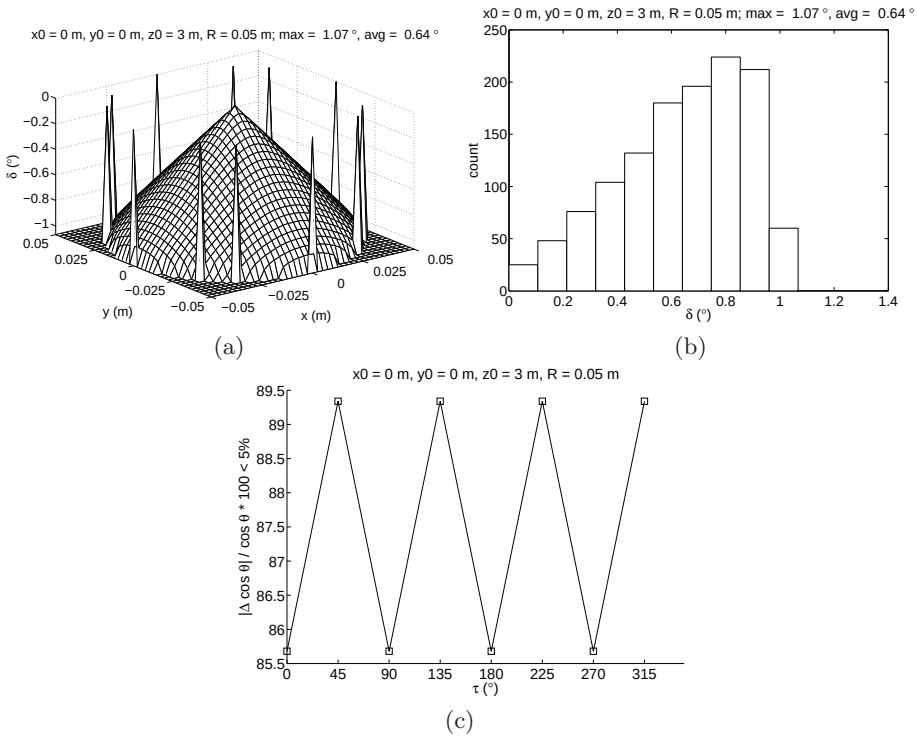


Fig. 5.9. Estimation errors for the suggested calibration setup, $R = 0.05$ m, $x_0 = 0$ m, $y_0 = 0$ m and $z_0 = 3$ m: (a) angle $-\delta$ accross the whole sphere (points outside the sphere have maximum deviation for visualization purposes); (b) histogram of δ values; (c) percentage of sphere points for which the relative error $\frac{|\Delta \cos\theta|}{\cos\theta} \times 100$ is below 5% for different values of the tilt τ of the illumination direction $\boldsymbol{s}$.

Given those results, $R \approx 0.05$ m, $z_0 \approx 3$ m, $x_0 \approx 0$ m and $y_0 \approx 0$ m seem quite reasonable calibration conditions, because they are either feasible and the related estimation error is below 5% for more than 85% of the sphere pixels. To finish this section, figure 5.9 shows more detailed plots about the estimation errors for the calibration conditions suggested. On the one hand, under those conditions, δ is always below 1.07°, while its mode is around 0.8°. On the other hand, it is worth noting that the error is higher as the border of the sphere is approached, so that it seems advisable not to consider pixels near the border of the sphere projection when using the estimation method based on orthographic projection.

5.4 Extension to Perspective Projection

As has been seen in the orthographic projection case, the estimation of the lighting strengths requires knowledge about the light source direction and the corresponding surface normal vectors at several image points of the sphere projection. When the orthographic projection assumption is not valid and central projection must be used, determining both unknowns is still possible although requires a previous stage for obtaining some parameters of the calibration sphere. The new procedure is described in the following.

5.4.1 Projection of the sphere

Under central projection, the projection of the sphere over the image plane is not always a sphere. In effect, the shape of the projection of the sphere results from the intersection of the image plane with the circular cone depicted in figure 5.10, whose vertex coincides with the center of projection. Therefore,

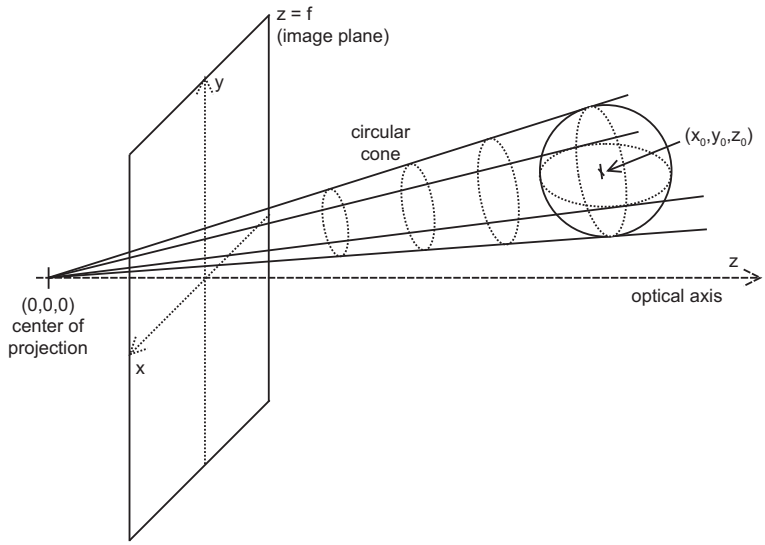


Fig. 5.10. The projection of a sphere over the image plane, under perspective projection, comes from the intersection between the circular cone of the figure and the image plane.

the projection of the sphere over the image plane is a conic since it is the intersection between a plane and a cone. Its exact parameters are derived next.

On the one hand, all the points over the sphere surface satisfy equation 5.27:

$$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = R^2 \,, \tag{5.27}$$

where, as before, $(x_0, y_0, z_0; R)$ are the parameters of the sphere.

On the other hand, the straight line connecting any point over the sphere surface (x, y, z) and the center of projection $(0, 0, 0)$ can be expressed, assuming a pin-hole camera model, as equation 5.28:

$$(x, y, z) = t(u, v, f) \,, \tag{5.28}$$

where, as in section 5.1, (u, v) is the projection of (x, y, z) onto the image plane (i.e. the intersection between the aforementioned straight line and the image plane) and f is the focal distance of the camera. Notice that (u, v) can no longer be assimilated to discrete coordinates over the image plane in contrast to orthographic projection.

Now, substituting equation 5.28 into equation 5.27 and reordering, equation 5.29 results:

$$\overbrace{\left(u^2 + v^2 + f^2\right)}^{a} t^2 \overbrace{- 2(ux_0 + vy_0 + fz_0)}^{b} t + \overbrace{x_0^2 + y_0^2 + z_0^2 - R^2}^{c} = 0 \,, \tag{5.29}$$

from which t is given as $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$. The outer points of the sphere projection correspond to the points of the sphere surface for which the straight line of equation 5.28 touches the sphere in just one point, i.e. $b^2 - 4ac = 0$. This fact leads to equation 5.30, which defines the wanted conic expressed in terms of u and v:

$$\left(R^2 - y_0^2 - z_0^2\right) u^2 + \left(R^2 - x_0^2 - z_0^2\right) v^2 + 2x_0 y_0 uv + 2f x_0 z_0 u + 2f y_0 z_0 v + f^2 \left(R^2 - x_0^2 - y_0^2\right) = 0 \,. \tag{5.30}$$

Due to their importance, a number of properties of this ellipse are discussed next:

(1) The eigenvalues of the matrix defining the quadratic form associated to the conic,

$$\begin{pmatrix} R^2 - y_0^2 - z_0^2 & x_0 y_0 \\ x_0 y_0 & R^2 - x_0^2 - z_0^2 \end{pmatrix} \,, \tag{5.31}$$

can be shown to be $\lambda_1 = R^2 - z_0^2$ and $\lambda_2 = R^2 - x_0^2 - y_0^2 - z_0^2$. Since z_0 is always larger than R —otherwise, part of the sphere would be inside the camera or behind it—, both eigenvalues are negative (i.e. have the same sign) and therefore the conic results to be in general an ellipse. Only when the center of the sphere is contained in the optical axis (i.e. x_0 and y_0 are both 0) the ellipse turns out to be a circumference. Note that this agrees with the discussion in the previous section about the error of assuming orthographic projection: the farther the center of the sphere with regard to the optical axis, the greater the error in the computation of the surface normal vectors, because the sphere projection adopts an elliptic shape and is increasingly less circular.

(2) Since $\lambda_1 < \lambda_2$ and λ_1 and λ_2 are inversely proportional to the square of the lengths of the ellipse axes [4], λ_1 is always related to the largest axis.

(3) Again using the inverse proportionality between the eigenvalues λ_1 and λ_2 and the squares of the ellipse axes lengths, $a^2 = |K/\lambda_1|$ and $b^2 = |K/\lambda_2|$ (K is once more a constant which derives from the parameters of the ellipse [4]), a relationship can be established between axes lengths a and b:

$$\frac{b^2}{a^2} = \frac{|K/\lambda_2|}{|K/\lambda_1|} = \frac{\lambda_1}{\lambda_2} = \frac{R^2 - z_0^2}{R^2 - z_0^2 - x_0^2 - y_0^2}\,, \tag{5.32}$$

which gives rise to:

$$b = a\sqrt{\frac{1 - \left(\frac{z_0}{R}\right)^2}{1 - \left(\frac{z_0}{R}\right)^2 - \left(\frac{x_0}{R}\right)^2 - \left(\frac{y_0}{R}\right)^2}}\,. \tag{5.33}$$

(4) Using equation 5.10 and the particular parameters of the ellipse, its center can be shown to be $(u_0, v_0) = f\frac{z_0^2}{z_0^2 - R^2}(x_0, y_0)$.

(5) From equation 5.11 and, again, the particular parameters of the ellipse, the slope of its largest axis (i.e. the one related with λ_1) turns out to be $\frac{y_0}{x_0}$, while the one for the shortest axis is given by $-\frac{x_0}{y_0}$.

(6) Using properties (4) and (5), the expression for the straight line containing the largest axis in parametric form is given by $(u, v) = f\frac{z_0^2}{z_0^2 - R^2}(x_0, y_0) + t(x_0, y_0)$, where t is the parameter.

(7) Finally, this straight line contains the origin of coordinates, i.e. if $t = -f\frac{z_0^2}{z_0^2 - R^2}$ then $(u, v) = (0, 0)$. In other words, the largest axis of this ellipse is always pointing towards the center of the image.

By way of illustration, figure 5.11 shows how the projection of the sphere in different positions in space would look like according to the previous properties. As predicted, when the center of the sphere lies over the optical axis, the projection is a circumference. As the sphere is displaced from this position, it leaves the circular shape, enlarging the axis pointing towards the image center. Finally, if the sphere is moved along a circular path centered on the optical axis, i.e. $x_0^2 + y_0^2 = K^2$, the eccentricity of the ellipse is always the same, though not its orientation.

5.4.2 Estimation of surface normal vectors

The surface normal vector $\boldsymbol{n}(x, y, z)$ for a sphere centered at (x_0, y_0, z_0) and having radius R is given by equation 5.34:

$$\boldsymbol{n}(x, y, z) = \begin{pmatrix} \frac{x - x_0}{R} \\ \frac{y - y_0}{R} \\ -\sqrt{1 - \left(\frac{x - x_0}{R}\right)^2 - \left(\frac{y - y_0}{R}\right)^2} \end{pmatrix}. \tag{5.34}$$

Given $(x, y) = \frac{z}{f}(u, v)$:

$$n_x = \frac{\frac{z}{f}u - x_0}{R} = \frac{u}{f}\frac{z}{R} - \frac{x_0}{R} \tag{5.35}$$

$$n_y = \frac{\frac{z}{f}v - y_0}{R} = \frac{v}{f}\frac{z}{R} - \frac{y_0}{R} \tag{5.36}$$

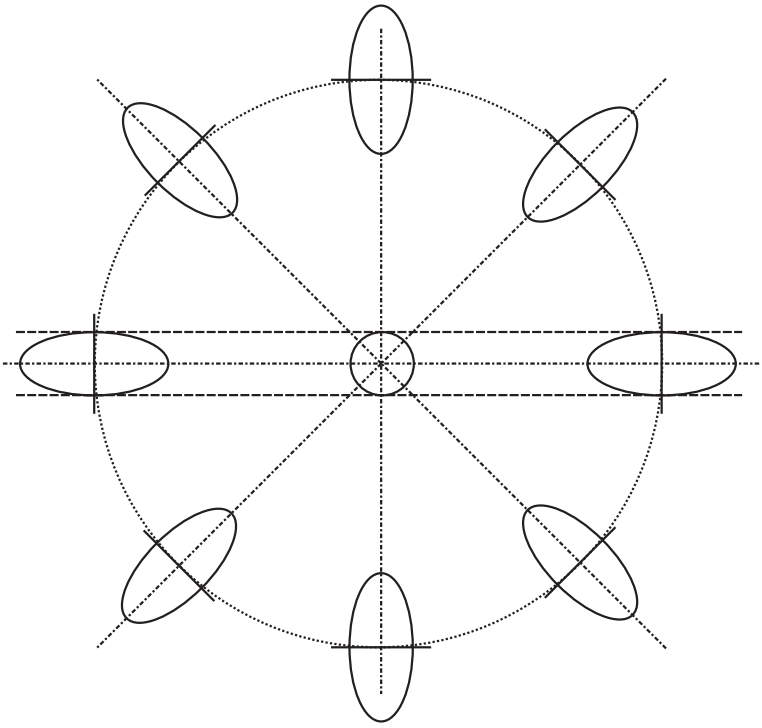


Fig. 5.11. Appearance of the projections of the sphere under different positions with regard to the camera optical axis. (The eccentricity of the ellipses has been exaggerated to improve the illustration of their properties.)

In order for these equations to be useful, f, $\frac{z}{R}$, $\frac{x_0}{R}$ and $\frac{y_0}{R}$ should be known beforehand. Although they could be derived after a geometric calibration of the camera, they can also be obtained from the ellipse fitting the contour of the sphere projection. On the one hand, $\frac{z}{R}$ is an array of values of the same size as the sphere projection which can be computed using the equation of the sphere surface, which is satisfied by all the points being projected into the image plane:

$$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = R^2 \,. \tag{5.37}$$

If, now, $x = u\frac{z}{f}$ and $y = v\frac{z}{f}$ are substituted and the whole equation is divided by R^2:

$$\left(\frac{u}{f}\frac{z}{R} - \frac{x_0}{R}\right)^2 + \left(\frac{v}{f}\frac{z}{R} - \frac{y_0}{R}\right)^2 + \left(\frac{z}{R} - \frac{z_0}{R}\right)^2 = 1 \,, \tag{5.38}$$

then, reordering equation 5.38, a second-order polynomial in $\frac{z}{R}$ is obtained as a function of u, v, f, $\frac{x_0}{R}$, $\frac{y_0}{R}$ and $\frac{z_0}{R}$:

$$\left[\left(\frac{u}{f}\right)^2 + \left(\frac{v}{f}\right)^2 + 1\right]\left(\frac{z}{R}\right)^2 - 2\left[\frac{u}{f}\frac{x_0}{R} + \frac{v}{f}\frac{y_0}{R} + \frac{z_0}{R}\right]\left(\frac{z}{R}\right) + \left[\left(\frac{x_0}{R}\right)^2 + \left(\frac{y_0}{R}\right)^2 + \left(\frac{z_0}{R}\right)^2 - 1\right] = 0 \,. \tag{5.39}$$

Then, $\frac{z}{R}$ can be computed for all the image points (u, v) covering the sphere projection, if f, $\frac{x_0}{R}$, $\frac{y_0}{R}$ and $\frac{z_0}{R}$ are known.

To determine f, $\frac{x_0}{R}$, $\frac{y_0}{R}$ and $\frac{z_0}{R}$, two methods are of application:

- On the one hand, from equation 5.30, the parameters of the ellipse are:

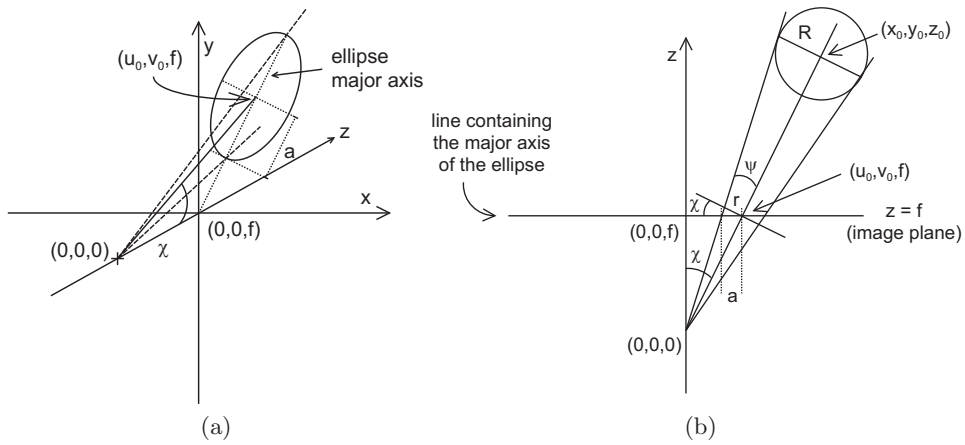


Fig. 5.12. Symbols used in the derivation of $\frac{z_0}{R}$, $\frac{x_0}{R}$ and $\frac{y_0}{R}$ as a function of f: (a) ellipse corresponding to the projection of the sphere over the image plane; (b) view of (a) after rotating the viewpoint so as to put it on top of the ellipse major axis.

$$\left.\begin{array}{l} a = k_0(R^2 - y_0^2 - z_0^2) \\ b = k_0(R^2 - x_0^2 - z_0^2) \\ c = k_0 x_0 y_0 \\ d = k_0 f x_0 z_0 \\ e = k_0 f y_0 z_0 \\ f = k_0 f^2 (R^2 - x_0^2 - y_0^2) \end{array}\right\} \tag{5.40}$$

where k_0 is just a scaling which is usually incorporated to the fitted parameters by the fitting method, most times being $\sqrt{a^2 + b^2 + ... + f^2}$. If they all are divided by R^2,

$$\left.\begin{array}{l} a = k_0' \left(1 - \left(\frac{y_0}{R}\right)^2 - \left(\frac{z_0}{R}\right)^2\right) \\ b = k_0' \left(1 - \left(\frac{x_0}{R}\right)^2 - \left(\frac{z_0}{R}\right)^2\right) \\ c = k_0' \frac{x_0}{R}\frac{y_0}{R} \\ d = k_0' f \frac{x_0}{R}\frac{z_0}{R} \\ e = k_0' f \frac{y_0}{R}\frac{z_0}{R} \\ f = k_0' f^2 \left(1 - \left(\frac{x_0}{R}\right)^2 - \left(\frac{y_0}{R}\right)^2\right) \end{array}\right\} \tag{5.41}$$

where $k_0' = k_0/R^2$, then a system of 6 non-linear equations in 5 unknowns is obtained. It can easily be shown that f, $\frac{x_0}{R}$, $\frac{y_0}{R}$ and $\frac{z_0}{R}$ can all be determined from this set of equations in all cases except when $x_0 = 0$ and $y_0 = 0$. In such a case, a and b turn out to be equal, and c, d and e are all 0, so that the system reduces to two equations and three unknowns. In this situation, $\frac{z_0}{R}$ can only be determined if the focal distance f or k_0' are known.

- On the other hand, if the focal distance f is available, either from the lens/camera manufacturer or from a previous geometric calibration of the camera, then the remaining variables $\frac{x_0}{R}$, $\frac{y_0}{R}$ and $\frac{z_0}{R}$ can be put as a function of f and some information coming from the ellipse fitting the contour of the sphere projection, which is derived next.
  First of all, given the relationship between the center of the sphere and the center of the ellipse $(x_0, y_0) = \frac{z_0}{f}(u_0, v_0)$, if both equations are divided by R, expressions for $\frac{x_0}{R}$ and $\frac{y_0}{R}$ are obtained in terms of $\frac{z_0}{R}$:

$$\frac{x_0}{R} = \frac{u_0}{f}\frac{z_0}{R}, \quad \frac{y_0}{R} = \frac{v_0}{f}\frac{z_0}{R}. \tag{5.42}$$

  Secondly, $\frac{z_0}{R}$ can be determined from some geometrical relationships between the sphere and its projection over the image plane. These relationships are established in figures 5.12(a) and (b), being the plane depicted in the latter the one containing the triangle $(0,0,0) - (0,0,f) - (u_0, v_0, f)$ of figure 5.12(a). Over this plane, angle ψ happens to be:

$$\tan\psi = \frac{R}{\sqrt{x_0^2+y_0^2+z_0^2}} = \frac{r}{\sqrt{u_0^2+v_0^2+f^2}}\,, \tag{5.43}$$

from which, substituting (x_0, y_0) by $\frac{z_0}{f}(u_0, v_0)$, the following expression for $\frac{z_0}{R}$ is obtained:

$$\frac{z_0}{R} = \frac{f}{r}\,. \tag{5.44}$$

Now, since $r = a\cos\chi$ (see figure 5.12(b)) and $\cos\chi = \frac{f}{\sqrt{u_0^2+v_0^2+f^2}}$ (see figure 5.12(a)), $\frac{z_0}{R}$ is definitely given by:

$$\frac{z_0}{R} = \frac{\sqrt{u_0^2+v_0^2+f^2}}{a} \tag{5.45}$$

From these two methods, the first one, although theoretically correct tends to be quite unstable in the sense that any error in the determination of the projection of the sphere gives rise to important errors in the computation of f, $\frac{x_0}{R}$, $\frac{y_0}{R}$ and $\frac{z_0}{R}$. However, in the second one, slight mistakes when indicating the projection contour gives rise to smooth deviations in the values taken by the remaining variables because the given f constrains the values they can take.

5.4.3 Estimation of the lighting direction

Assuming again a distant light source, and using the expressions developed for n_x and n_y (equations 5.35 and 5.36), under perspective projection, $\cos\theta$ results to be:

$$\cos\theta = \left(\frac{u}{f}\frac{z}{R} - \frac{x_0}{R}\right)s_x + \left(\frac{v}{f}\frac{z}{R} - \frac{y_0}{R}\right)s_y - \sqrt{1 - \left(\frac{u}{f}\frac{z}{R} - \frac{x_0}{R}\right)^2 - \left(\frac{v}{f}\frac{z}{R} - \frac{y_0}{R}\right)^2}\,s_z\,. \tag{5.46}$$

In the same way as for deriving $\boldsymbol{s}$ for orthographic projection (equations 5.6 and 5.7), the isophote curve of intensity L of the μ_{D^c} image is given by:

$$L = \alpha^c + \left[\left(\frac{u}{f}\frac{z}{R} - \frac{x_0}{R}\right)s_x + \left(\frac{v}{f}\frac{z}{R} - \frac{y_0}{R}\right)s_y - \sqrt{1 - \left(\frac{u}{f}\frac{z}{R} - \frac{x_0}{R}\right)^2 - \left(\frac{v}{f}\frac{z}{R} - \frac{y_0}{R}\right)^2}\,s_z\right]\beta^c\,, \tag{5.47}$$

where, again, $\alpha^c = (L_a\rho_b)^c A^c + \mathrm{E}_I[\mu_{dc}]A^c$ and $\beta^c = (L_d\rho_b)^c A^c$.

Equation 5.48 can now be obtained reordering equation 5.47:

$$\begin{aligned}\left(\frac{u}{f}\frac{z}{R} - \frac{x_0}{R}\right)^2(s_x^2+s_z^2) + \left(\frac{v}{f}\frac{z}{R} - \frac{y_0}{R}\right)^2(s_y^2+s_z^2) + 2\left(\frac{u}{f}\frac{z}{R} - \frac{x_0}{R}\right)\left(\frac{v}{f}\frac{z}{R} - \frac{y_0}{R}\right)s_x s_y \\ - 2\mathcal{C}\left(\frac{u}{f}\frac{z}{R} - \frac{x_0}{R}\right)s_x - 2\mathcal{C}\left(\frac{v}{f}\frac{z}{R} - \frac{y_0}{R}\right)s_y + \mathcal{C}^2 - s_z^2 = 0\,,\end{aligned} \tag{5.48}$$

where $\mathcal{C} = \frac{L-\alpha^c}{\beta^c}$. If variables p and q are defined, in this case, as $p = \frac{u}{f}\frac{z}{R} - \frac{x_0}{R}$ and $q = \frac{v}{f}\frac{z}{R} - \frac{y_0}{R}$, then equation 5.48 defines again a rotated conic in terms of the variables p and q. Notice, that, unlike the orthographic projection case, variables p and q do not define in this case a linear transformation common to all the points over the sphere projection. However, after introducing variables p and q, equation 5.48 results to define exactly the same ellipse as in the orthographic projection case (equation 5.7), so that all the properties discussed for that ellipse hold also here. Consequently, the lighting direction s can be estimated, under central projection, using the same methods explained for orthographic projection, after applying the transformation from (u, v) to (p, q).

To finish, remember that variables (u, v) do not represent discrete coordinates and that they must be expressed in the same units as $\frac{x_0}{R}$, $\frac{y_0}{R}$, $\frac{z}{R}$ and f, unlike the orthographic projection case, where all the magnitudes —u, v, u_0, v_0 and r— naturally shared the same units and this issue did not care. If $\frac{x_0}{R}$, $\frac{y_0}{R}$, $\frac{z}{R}$ and f are expressed in, say, meters, then (u, v) must also be expressed in meters. The wanted transformation can usually be obtained from the particular CCD size (1/4", 1/3", 1/2", 2/3", etc.) and the number of cells for every dimension of the CCD array, both of which are normally available from the camera manufacturer.

Pseudocode 5.1 Lighting parameters estimation pseudocode for the perspective projection case.

(1) Determine the parameters of the elliptic projection of the calibration sphere over the image (ensure u and v are expressed in the proper units)
(2) Estimate $\frac{x_0}{R}$, $\frac{y_0}{R}$, $\frac{z_0}{R}$ and f from the elliptic shape of the projection
(3) Compute the array $\frac{z}{R}$ for all the points within the sphere projection using the values found for $\frac{x_0}{R}$, $\frac{y_0}{R}$, $\frac{z_0}{R}$ and f (ensure u and v are expressed in the proper units)
(4) Transform (u, v) co-ordinates to (p, q) co-ordinates (ensure u and v are expressed in the proper units)
(5) Compute the lighting direction $\boldsymbol{s}$ considering the ellipses corresponding to several isophote curves of the sphere projection image defined over the plane $\mathcal{P} - \mathcal{Q}$
(6) Obtain estimations for the ambient and directional lighting strengths fitting a straight line to points $(m_b, \mathrm{E}_{m_b}[\mu_{D^c}])$

5.4.4 Summary of the extended estimation method

Pseudocode 5.1 sums up the estimation method for the central projection case. As in the orthographic projection case, all the estimations are done on the basis of the indicated contour of the projection of the sphere. In order to overcome the uncertainty on the true location of the projection, again an optimization strategy is employed. First, two circles among which the true contour must lie are given, together with a first guess about the ellipse corresponding to the real contour. Next, an elitist genetic strategy is applied [170,171]. That is to say, several ellipses are generated by Gaussian mutation of the parameters of the first guess. The one leading to the best fitting of points $(m_b, \mathrm{E}_{m_b}[\mu_{D^c}])$ is kept for the next generation, where again other ellipses are produced by mutating it. After some generations without improvement, the best ellipse is considered to be the best estimation of the real projection contour. At all times, every ellipse is ensured to satisfy all the properties discussed before and be enclosed within the limits of the two circumferences given at the beginning. Notice that, for being sure that any ellipse satisfies all those properties is sufficient, given f, to specify a center (u_0, v_0) and one of the axes, say the longest one, having length a; from them, the orientation of the ellipse ϕ is determined as $\tan^{-1} \frac{v_0}{u_0}$, while b is computed using equation 5.33.

5.5 Previous Work

5.5.1 Illuminant orientation determination

The study of the detection of the illuminant direction from images can be traced back to the 1980's. Since then, several light source estimators have been proposed. The primary motivation behind the work was to extend shape from shading algorithms to operate with less *a priori* information. In this sense, [103] is an important source of information.

Among the classical approaches, some of them estimate the orientation of the illuminant on the basis of reasonable assumptions about the statistical distribution of surface orientations, such as the methods by Pentland [224], Lee and Rosenfeld [141] and Zheng and Chellappa [311]. The first approach assumes that change in surface normal is distributed isotropically within a region, so that along any image direction the mean change in the z component of the surface normal is zero and, consequently, the mean image gradient along that direction depends only on the x and y components of the light orientation vector in a proportional way. After some algebraic manipulation, the author shows how, by means of a maximum likelihood analysis over several image directions, the tilt and the slant of the direction of the illuminant can be estimated. The approach by Lee and Rosenfeld considers only the derivatives along the x and y directions and approximates the local surface geometry by spherical patches; hence, their method is also based on assuming an isotropic distribution of surface orientations. Without a maximum likelihood analysis, the authors show that the expected values of image gradients along the x and y directions are enough to compute the tilt of the illumination. Regarding the slant, it

is estimated considering a sampling distribution of the slant of surface normals which takes into account foreshortening and avoids self-shadowing, so that, using the expected values of intensity and intensity squared, a non-linear equation on the slant is developed. Finally, Zheng and Chellappa present the most sophisticated method together with a modification of Lee and Rosenfeld's approach, incorporating self-shadows into their calculations. In their own approach, they develop an expression for the slant based on seventh-order polynomials approximating the intensity and intensity squared averages derived on the basis of a statistical model of the distribution of surface normals taking into account self-shadows and foreshortening [311]. As for the tilt, two methods are proposed: one of them assumes local spherical patches and makes use of the maximum likelihood analysis of Pentland for deriving the tilt of the illuminant on a pixel-by-pixel basis; the other approach uses shading information along image contours.

Another family of methods incorporate the estimation of the lighting parameters into optimization schemes for the recovery of shape. In [24], Brooks and Horn present a method which alternately determines the surface shape and the light source direction. The scheme is derived using a variational approach and is extended to the case in which, in addition to a point light source, there is a distributed "sky" source. In fact, the iterative scheme is then extended to arbitrary reflectance maps. The algorithms are simple to implement but do not enforce the integrability of the recovered surfaces (i.e the requirement for a physically plausible shape). Using also a global optimization scheme, Leclerc and Bobick propose in [139] a method computing first depth information equating the derivative of the objective function to zero and solving for it using a conjugate gradient technique together with a hierarchical structure to solve simultaneous equations. Once depth is known, the derivatives of the objective function with respect to the three components of the light source direction is equated to zero and the three simultaneous linear equations are solved. As initial data, their method uses depth estimated from stereo. Later, a paper by Hougen and Ahuja describes a method based on Leclerc and Bobick's approach able to cope with several light sources located at infinity, although only their strengths are estimated since intensity-independent reflectance maps are known in advance [105]. More recently, Samaras and Metaxas proposed a scheme coupling, within an iterative process, their shape estimation method with a method for light orientation estimation; this scheme improves shape estimation on the basis of improvements on the light direction estimation and vice versa [247]. Due to the characteristics of the method, the authors are able to incorporate more complex models of image formation such as the Oren and Nayar extension to the Lambertian model [197]. The lighting estimations compare favourably with the approach by Zheng and Chellapa, among others.

5.5.2 Estimation of the illuminant chromaticity

One of the classical problems in colour vision is the determination of the colour of the illumination. Throughout the years, researchers have provided several methods for solving this problem, mainly for achieving colour constancy or, in other words, the recovery of robust descriptors of scene objects from colour images which are not affected by changes in the illumination. [95] constitutes an important reference on the subject.

In [280], Tominaga and Wandell use the Dichromatic Reflection Model to predict that, under the assumption of wavelength-independent specular reflectances, planes containing scene object pixels in colour space intersect along a line that describes the spectral distribution of the light source. In [156], Maloney and Wandell develop an algorithm able to recover the spectral power distribution (SPD) of the illuminant together with the spectral reflectance at each location in the scene, on the assumption of a single SPD illuminating the scene and that both the SPD of the illuminant and the surfaces spectral reflectance are representable by finite dimensional linear models. Later, Forsyth extends and improves the algorithm of Maloney and Wandell realizing that the colour of the illuminant is constrained by the colours observed in the image because surfaces can reflect no more light than is cast on them [52]. Ho et al. present another algorithm for recovering the scene surfaces spectral reflectance and the illuminant SPD assuming finite dimensional linear models and significant structural differences on both. This allows the recovery of three parameters describing the illuminant and three more describing the surface reflectance from a sampled reflected colour signal [97]. Lee proposes a method based on the Dichromatic Reflection Model for estimating the chromaticity of the illuminant assuming that the spectral composition of the interface reflectance is the same as the spectral composition of the

illumination [143]. Using such a model, the locus of points in chromaticity space for a surface is predicted to lie on a line connecting the chromaticity of the illuminant and the chromaticity of the scene objects body reflection. While the original method needs the presence of several objects in the scene, an improved version removes this constraint [144].

5.5.3 General comment on the use of a calibration object

All the above methods make the estimations over the whole image, or over selected areas of the image under consideration, but do not use a calibration object. Although they are more attractive because they do not require the calibration step and therefore can be applied to uncalibrated images, their success on the correct estimation of the lighting orientation greatly depends on the strict matching between the assumed image formation model and the image, and on the stability of the numerical algorithm used for working out the different equations involved. With a calibration object, all the lighting conditions can be controlled, avoiding specularities, inter-reflections and other undesirable effects, which allows obtaining accurate estimations of the lighting parameters.

Some recent papers propose methods for estimating part of the lighting parameters by means of a calibration-object-based approach. In the first one, Powell et al. present a strategy for estimating the locations of several light sources without assuming parallel lighting [230]. The procedure involves the use of a novel calibration object that consists of three specular spheres at known relative positions. The 3D position of the specularities visible on their surfaces are used to triangulate for the light source positions. In another publication, Zhang and Yang propose a method for the detection of the direction of multiple illuminants based on a Lambertian sphere, looking for critical points in the scene, i.e. points whose surface normal is orthogonal to the direction of at least one light source [309]. Light sources are assumed distant and their strengths and directions are found by means of least-squares minimization over a set of over-determined equations derived using the critical points found. Wang and Samaras extended the work of Zhang and Yang allowing the use of any scene object of known geometry to estimate lighting information, avoiding thus the utilization of a calibrating sphere [289]. Zhou and Kambhamettu proposed another scheme for locating distant multiple light sources and estimating their strengths from a pair of stereo images of a sphere in [312]. The material of the sphere surface is required to be a dielectric so that both body and interface reflection takes place. Using the dependence of the specular reflection on the viewpoint, the image of the sphere is separated into two images: one with the specularities and the other one with the diffuse reflection. The first image is used to determine the direction of the light sources while the second one provides their strengths. Finally, Bouganis and Brookes proposed the V2R algorithm for, once more, distant multiple light source detection [19]. The V2R algorithm uses a Lambertian sphere and is based again on the detection of the critical points of the scene. The image of the sphere is first segmented into regions that are each illuminated by a different sum of light sources, called virtual lights by the authors, using the critical points found. Next, light source vectors containing either the direction of the virtual lights and their strengths are determined by linear regression using all the points in each region. Individual light source vectors are finally computed subtracting the virtual light source vectors of adjacent regions.

Despite the limitations of the lighting estimation methods proposed in this chapter as for the number of light sources, several advantages make them attractive in single-source cases against multi-source methods: (1) in general, its complexity is lower than the multi-source methods; (2) a previous geometric calibration of the camera is not required and the estimation of the surface normal vectors of the calibration sphere is embedded within the method itself; (3) all the image pixels belonging to the sphere projection are used in the estimation of the lighting parameters; (4) no threshold or parameter needs to be set up. Besides, the mathematical formulation of the method has been developed taking into account a model of the operation of CCD cameras, which has allowed identifying the noise sources corrupting digital pixel values in order to counteract their effects throughout the estimation process; furthermore, such a formulation has revealed that any estimation of the strength of ambient illumination always includes dark current, so that it must be estimated apart to get the real strength of ambient illumination, if needed.

Finally, in essence, the methods of Zhang and Yang [309] and Bouganis and Brookes [19], for the case of one directional light source, and no ambient illumination in the first work, are similar to the

ones proposed in this chapter in the sense that their critical points correspond to the sphere points whose isophote curve in the image has the least intensity over the sphere.

5.6 Experimental Results

A number of experiments have been performed in order to evaluate the robustness of the estimation procedures described in this chapter and the accuracy of their results. The real images used in some experiments were captured using a JAI CV-M70 progressive scan colour CCD camera with linear response and 8 bits per colour channel and pixel, halogen illumination ranging from 400 W to 650 W and a COMET Matrox frame grabber. According to the manufacturer specifications, the CCD of this camera consists of 640×480 effective square pixels 9.9 μm wide. On the other hand, the focal distance of the camera is computed as the sum of the optics own focal distance, 16 mm, and the flange back distance associated to the type of camera lens mount, 17.526 mm (C-mount); consequently, $f = 33.526$ mm [107, 275]. Those parameters have also been used to generate the synthetic images required by some experiments.

As a general rule throughout this section, uncertainties $\delta\widehat{x}$ corresponding to estimates $\widehat{x}$ are calculated through standard formulas when $\widehat{x}$ comes from a regression analysis, while, for averaged values, the Standard Deviation Of Mean, $\sigma_{\overline{x}} = \frac{\sigma_x}{\sqrt{N}}$, is used [272]. As it is well known for normally distributed measurements and $\delta\widehat{x}$ calculated in this way, $\widehat{x} \pm \delta\widehat{x}$ provides a 68% confidence interval for x, while $\widehat{x} \pm 2\delta\widehat{x}$ corresponds to a 95% confidence interval, etc [151, 272]. Besides, for further reference, the estimation method assuming orthographic projection will be referred to as OPAM, *Orthographic Projection Assuming Method*, while CPAM, *Central Projection Assuming Method*, will be used for the method assuming central projection. On the other hand, TEAV and TECA will stand for *Tilt Estimation by AVeraging* and *Tilt Estimation by Centers Alignment*, as they were introduced in section 5.1.

In all the experiments performed with real images, the calibration arrangement consisted in a sphere made of white cork lying over a black background to avoid shadows (see figure 5.13(1st row,left)). The surface of the sphere was painted using a matte white paint to avoid specularities and increase the spatial uniformity of surface reflectance. Note, however, that, although the sphere has always been required white, the fact that it is more or less white is not so important as whether its reflectance is more or less constant throughout the visible spectrum. This is because, as pointed before, the F-number is embedded into the estimations of $L_a^c A^c$ and $L_d^c A^c$, so that, for the same lighting conditions, different estimations are obtained if the lens aperture is changed. Since the reflectance of the sphere ρ_b affects both $L_a^c A^c$ and $L_d^c A^c$ equally, as the F-number does, the actual value of ρ_b, if it is independent of wavelength, acts as if the aperture was less than the one corresponding to the actual F-number of the camera. Therefore, guaranteeing $\rho_b = 1$ is important only if the effect of the optics aperture is going to be removed from the estimations.

5.6.1 Example of lighting parameters estimation

In this section, TEAV and TECA methods will be illustrated step by step for the general case of central projection, using the calibration image shown in figure 5.13(1st row,right). Table 5.3 collects some illustrative data about this particular experiment.

First, an initial guess about the location of the sphere projection contour is generated by separating the sphere projection from the background using as threshold an intensity of 18 in all colour channels, and fitting the resulting contour by an ellipse satisfying the properties enumerated in section 5.4. The dashed line of figure 5.13(1st row,right) indicates the resultant ellipse. Next, the genetic optimization strategy outlined in section 5.4 is executed for TEAV and TECA methods, using 20 ellipses per generation and constraining them to be within two concentric circles whose radius are 2 image units below and above the largest axis of the initial ellipse. Once 5 generations happen to introduce no improvement, convergence is assumed to have been attained, which, in these particular cases, occurred after 25 and 14 generations for, respectively, TEAV and TECA methods. The optimum ellipses found are indicated as solid lines in figures 5.13(2nd row,a) and (3rd row,a). For those ellipses, the values

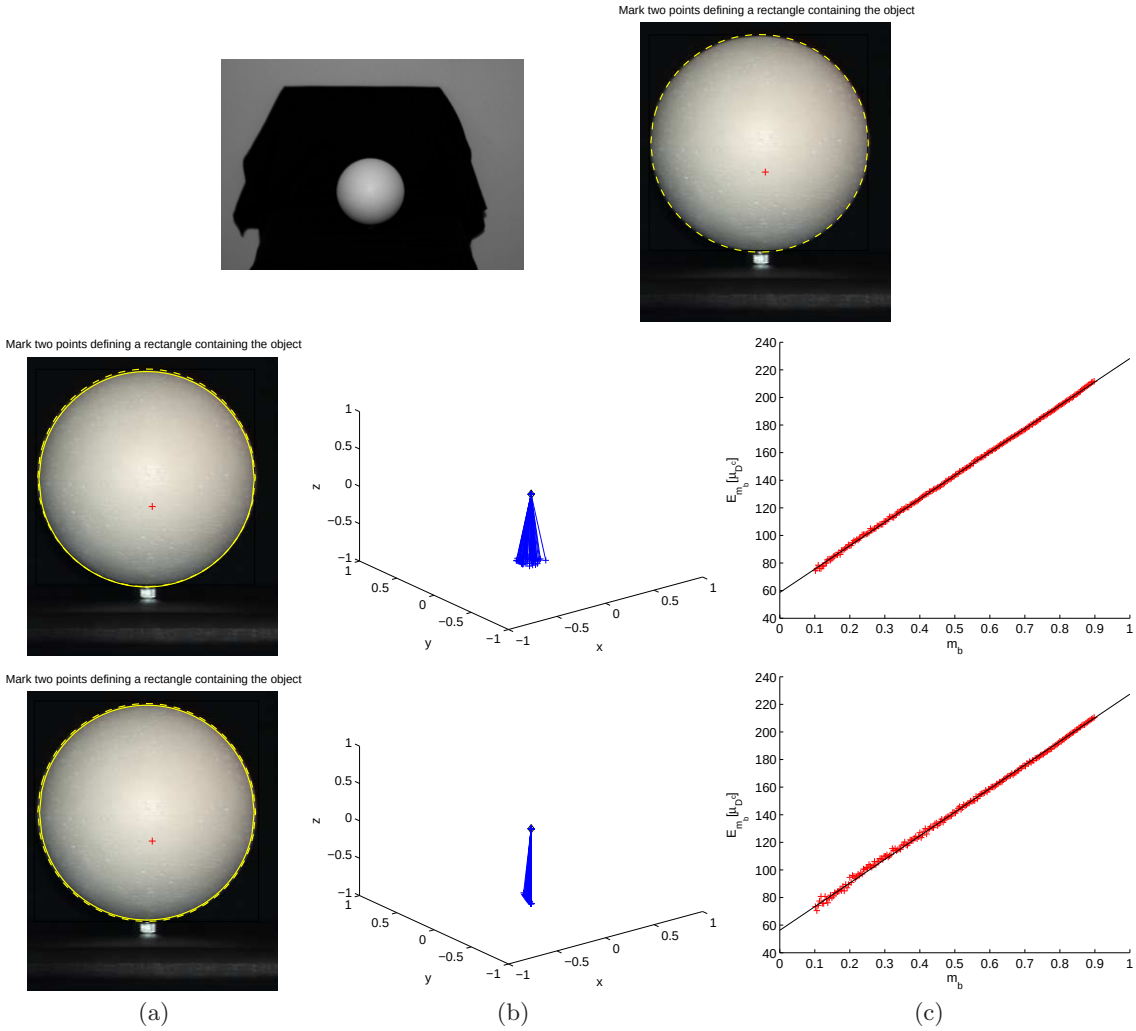


Fig. 5.13. Example of estimation for CPAM: [1st row,left] calibration setup; [1st row,right] calibration image with initial guess superimposed as a dashed line; [2nd row] TEAV method results; [3rd row] TECA method results. In second and third rows, (a) calibration image with initial guess superimposed as a dashed line and final ellipse superimposed as a solid line; (b) light source directions found; (c) straight line corresponding to points $(m_b, \mathrm{E}_{m_b}[\mu_{D^c}])$ for the red channel.

estimated for $\frac{x_0}{R}$, $\frac{y_0}{R}$ and $\frac{z_0}{R}$ are given in table 5.3. The values between parenthesis correspond to x_0, y_0 and z_0 after having measured $2R$ as 11.3 cm.

Table 5.3 also collects the initial and final fitting errors for every colour band, as well as the corresponding linear correlation coefficients. On the one hand, the fitting error values given come from scaling m_b from the interval $[0, 1]$ to the interval $[0, 255]$, while $\mathrm{E}_{m_b}[\mu_{D^c}]$ values kept, from origin, within $[0, 255]$. On the other hand, it can be clearly seen that the optimization improves significantly the fitting of points $(m_b, \mathrm{E}_{m_b}[\mu_{D^c}])$ to the model. By way of illustration, the resultant sets of pairs $(m_b, \mathrm{E}_{m_b}[\mu_{D^c}])$, for the red colour channel, are given in figures 5.13(2nd row,c), TEAV case, and (3rd row,c), TECA case, together with the fitting line. As can be seen in these last plots, m_b goes from 0.1 to 0.9, instead of covering the whole interval $[0, 1]$; it is also constrained to values for which the slant of the surface normal $\cos^{-1} n_z$ is above 95°. On the one hand, $m_b \geq 0.1$ in order to diminish the effect of the inaccuracies of the Lambertian model when angle θ is large, as suggested by Wolff et al. [295]. Besides, $\cos^{-1} n_z > 95°$ reduces the influence of the real shape of the calibration sphere, which will hardly be perfectly spherical. This influence is more harmful at the border of the projection, since can

Table 5.3. Illustrative data for the lighting estimation example. In the table, $\widehat{\boldsymbol{\alpha}}$ stands for the vector of estimates of $L_a^c A^c + \mathrm{E}_I[\mu_{dc}]A^c$ and $\widehat{\boldsymbol{\beta}}$ for the vector of estimates of $L_d^c A^c$, $c \in \{R, G, B\}$, while δx is the uncertainty associated to value x.

	TEAV	TECA
initial fitting errors	(2.7991,3.5306,3.5446)	(6.8137,8.1313,7.9425)
initial linear corr. coef.	(0.9981,0.9969,0.9962)	(0.9940,0.9931,0.9926)
# iterations	25	14
final fitting errors	(0.6570,0.5376,0.5879)	(1.3957,1.2446,1.3713)
final linear corr. coef.	(0.9999,0.9999,0.9998)	(0.9994,0.9995,0.9992)
$\frac{x_0}{R}$ (x_0) [cm]	-0.05 (-0.29)	-0.05 (-0.29)
$\frac{y_0}{R}$ (y_0) [cm]	+0.25 (+1.43)	+0.26 (+1.49)
$\frac{z_0}{R}$ (z_0) [cm]	+30.28 (+171.08)	+30.38 (+171.63)
$\widehat{\boldsymbol{s}}$	(0.07,0.17,-0.98)	(0.02,0.04,-1.00)
$\delta\widehat{\boldsymbol{s}}$	(0.00,0.00,0.00)	(0.00,0.01,0.00)
$\widehat{\boldsymbol{\alpha}}$	(58.68,62.85,63.58)	(56.25,58.73,57.98)
$\delta\widehat{\boldsymbol{\alpha}}$	(0.09,0.10,0.09)	(0.22,0.16,0.12)
$\widehat{\boldsymbol{\beta}}$	(169.52,160.72,143.33)	(171.13,164.72,149.71)
$\delta\widehat{\boldsymbol{\beta}}$	(0.14,0.15,0.15)	(0.29,0.20,0.16)
$\overline{\left\lvert \boldsymbol{\mu_D} - (\widehat{\boldsymbol{\alpha}} + \widehat{m_b}\widehat{\boldsymbol{\beta}}) \right\rvert}$	(3.59,4.74,5.33)	(8.78,6.09,4.52)
$\max_{75\%}\{\left\lvert \boldsymbol{\mu_D} - (\widehat{\boldsymbol{\alpha}} + \widehat{m_b}\widehat{\boldsymbol{\beta}}) \right\rvert\}$	(4,6,7)	(12,8,5)

give rise to the introduction of points of the black background in the analysis (i.e. $D^c \approx 0$). On the other hand, $m_b \leq 0.9$ keeps the estimations from being affected by clipping (i.e. when the dynamic range of the camera is overflowed due to an excessive light level).

For the computation of the lighting direction, Δ was set to 4 intensity levels, what gave rise to finding the unit directions presented in figure 5.13(2nd row,b) and (3rd row,b). In both cases, TEAV and TECA, the final estimation of $\boldsymbol{s}$ is obtained through clustering by means of the computation of the MVE enclosing 75% of the tilt and slant values separately, TEAV, or just the slant values, TECA. In this last case, the tilt of $\boldsymbol{s}$ was previously estimated from the alignment of ellipse centers.

Finally, lighting strengths were determined by fitting points $(m_b, \mathrm{E}_{m_b}[\mu_{D^c}])$ for all the colour channels, giving rise to the estimates given in table 5.3. As a final remark, notice the closeness of the linear correlation coefficients to +1, which indicates a strong alignment of points $(m_b, \mathrm{E}_{m_b}[\mu_{D^c}])$ for the different colour channels, which, on the other hand, can be visually checked for the red channel in plots 5.13(2nd row,c) and (3rd row,c). That is to say, the results agree with the image formation model, as they were supposed to be, which reinforces the validity of the estimation methods. Furthermore, the intensity for pixels enclosed by the sphere projection were re-synthesized using the estimates of $\alpha^c = L_a^c A^c + \mathrm{E}_I[\mu_{dc}]A^c$, $\beta^c = L_d^c A^c$ and m_b, and difference $\left\lvert \mu_{D^c} - (\widehat{\alpha}^c + \widehat{m_b}\widehat{\beta}^c) \right\rvert$ was evaluated for all those pixels. The result was that, for 75% of pixels, the difference was always below (4,6,7) for TEAV and (12,8,5) for TECA, while the average difference was, for the whole sphere, (3.59,4.74,5.33) and (8.78,6.09,4.52), respectively. To fully understand the meaning of these figures take into account that, on the one hand, the real sphere is not geometrically perfect, but consists of small pits and bumps all scattered across its surface, apart from other slight imperfections in its global shape. On the other hand, the camera spatial noise effects could not be discounted from μ_{D^c} when computing the difference between actual and synthesized intensity.

Table 5.4. Lighting parameters estimates using synthetic images generated under orthographic projection. $\widehat{\alpha}^R$ stands for the estimate of $L_a^R A^R$ and $\widehat{\beta}^R$ for the estimate of $L_d^R A^R$. The angle between the true value of $\boldsymbol{s}$ and the estimation $\widehat{\boldsymbol{s}}$ is given between square brackets.

OPAM		
$\boldsymbol{s}$	TEAV	TECA
(+0.71,+0.00,-0.71)	$\widehat{\boldsymbol{s}}$ = (+0.71,+0.00,-0.71) [0.04°] $\widehat{\alpha}^R = 20.26$, $\widehat{\beta}^R = 234.71$	$\widehat{\boldsymbol{s}}$ = (+0.71,+0.00,-0.71) [0.00°] $\widehat{\alpha}^R = 20.38$, $\widehat{\beta}^R = 234.59$
(+0.50,+0.50,-0.71)	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.07°] $\widehat{\alpha}^R = 20.68$, $\widehat{\beta}^R = 234.36$	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.04°] $\widehat{\alpha}^R = 20.56$, $\widehat{\beta}^R = 234.47$
(+0.00,+0.71,-0.71)	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.71,-0.71) [0.04°] $\widehat{\alpha}^R = 20.26$, $\widehat{\beta}^R = 234.71$	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.71,-0.71) [0.00°] $\widehat{\alpha}^R = 20.38$, $\widehat{\beta}^R = 234.59$
(-0.50,+0.50,-0.71)	$\widehat{\boldsymbol{s}}$ = (-0.50,+0.50,-0.71) [0.07°] $\widehat{\alpha}^R = 20.68$, $\widehat{\beta}^R = 234.36$	$\widehat{\boldsymbol{s}}$ = (-0.50,+0.50,-0.71) [0.04°] $\widehat{\alpha}^R = 20.56$, $\widehat{\beta}^R = 234.47$
(-0.71,+0.00,-0.71)	$\widehat{\boldsymbol{s}}$ = (-0.71,+0.00,-0.71) [0.04°] $\widehat{\alpha}^R = 20.26$, $\widehat{\beta}^R = 234.71$	$\widehat{\boldsymbol{s}}$ = (-0.71,+0.00,-0.71) [0.00°] $\widehat{\alpha}^R = 20.38$, $\widehat{\beta}^R = 234.59$
(-0.50,-0.50,-0.71)	$\widehat{\boldsymbol{s}}$ = (-0.50,-0.50,-0.71) [0.07°] $\widehat{\alpha}^R = 20.68$, $\widehat{\beta}^R = 234.36$	$\widehat{\boldsymbol{s}}$ = (-0.50,-0.50,-0.71) [0.04°] $\widehat{\alpha}^R = 20.56$, $\widehat{\beta}^R = 234.47$
(-0.00,-0.71,-0.71)	$\widehat{\boldsymbol{s}}$ = (+0.00,-0.71,-0.71) [0.04°] $\widehat{\alpha}^R = 20.26$, $\widehat{\beta}^R = 234.71$	$\widehat{\boldsymbol{s}}$ = (+0.00,-0.71,-0.71) [0.00°] $\widehat{\alpha}^R = 20.38$, $\widehat{\beta}^R = 234.59$
(+0.50,-0.50,-0.71)	$\widehat{\boldsymbol{s}}$ = (+0.50,-0.50,-0.71) [0.07°] $\widehat{\alpha}^R = 20.68$, $\widehat{\beta}^R = 234.36$	$\widehat{\boldsymbol{s}}$ = (+0.50,-0.50,-0.71) [0.04°] $\widehat{\alpha}^R = 20.56$, $\widehat{\beta}^R = 234.47$

5.6.2 Orthographic projection versus perspective projection

This section compares estimations produced by OPAM and CPAM. At this point, it is important to remember that OPAM does not need more than the sphere projection parameters, so that, under the adequate conditions, can be easier to use than CPAM. To determine its scope of usage, several experiments with synthetic images of a white sphere over black background were run and the resultant estimations for both approaches are given in tables 5.4, 5.5, 5.6, 5.7 and 5.8. In all of them, the estimation algorithms were provided with the exact projection parameters. Besides, noise was not incorporated into the images, so that, in particular, $\mathrm{E}_I[\mu_{dc}] = 0$ and, therefore, $L_a^c A^c$ could be directly measured from points $(m_b, \mathrm{E}_{m_b}[\mu_{D^c}])$. Finally, $L_a^c A^c = 0.08 \times 255 = 20.4$ and $L_d^c A^c = 0.92 \times 255 = 234.6$, for every colour channel c.

First of all, OPAM was faced with a set of images generated under orthographic projection where the tilt of the lighting was varied from 0 to 360° in 45° steps. The results are given in table 5.4. As can be seen, the estimations are correct in all cases leaving aside some inaccuracies due to the digitalization of the intensity, together with the spatial discretization. Since the error associated to the latter increases as the size of the projection of the sphere decreases, higher inaccuracies must be expected on those cases.

Secondly, the same experiment as before, but with images generated under perspective projection, was performed, setting $R = 0.05$ m, $x_0 = 0$, $y_0 = 0$ and $z_0 = 3$ m, as indicated in section 5.3. Now, results for both OPAM and CPAM are given in table 5.5. Those results clearly show that, under those conditions, the inaccuracies in the lighting direction $\boldsymbol{s}$ estimated by OPAM are of the same order as the inaccuracies of CPAM, and almost negligible. However, the estimates of $L_a^c A^c$ and $L_d^c A^c$ start suffering from the orthographic assumption in OPAM, with deviations from the true values of around 4 units in each case. Despite those errors, the sets of points $(m_b, \mathrm{E}_{m_b}[\mu_{D^c}])$ were still aligned with mean fitting errors around 0.1 and linear correlation coefficients of 0.999998, which suggests the idea that, under the conditions stated above, OPAM generates points $(m_b, \mathrm{E}_{m_b}[\mu_{D^c}])$ still lying in a straight line, although different to the correct one because of the error in the estimation of m_b.

As a third experiment, z_0 was varied from 1 to 8 meters in 1-meter steps, keeping the lighting orientation constant at $(0, 0, -1)$ and $(x_0, y_0) = 0$ in all eight cases. As it is shown in table 5.6, $\boldsymbol{s}$ is again properly estimated and mostly unaffected by changes in z_0, while the estimates of $L_a^c A^c$ and

Table 5.5. Lighting parameters estimates using synthetic images generated under central projection, varying the tilt of the illumination. $\widehat{\alpha}^R$ stands for the estimate of $L_a^R A^R$ and $\widehat{\beta}^R$ for the estimate of $L_d^R A^R$. The angle between the true value of $\boldsymbol{s}$ and the estimation $\widehat{\boldsymbol{s}}$ is given between square brackets.

OPAM		
$\boldsymbol{s}$	TEAV	TECA
(+0.71,+0.00,-0.71)	$\widehat{\boldsymbol{s}}$ = (+0.71,+0.00,-0.70) [0.32°] $\widehat{\alpha}^R = 24.47$, $\widehat{\beta}^R = 230.48$	$\widehat{\boldsymbol{s}}$ = (+0.71,+0.00,-0.70) [0.31°] $\widehat{\alpha}^R = 24.44$, $\widehat{\beta}^R = 230.51$
(+0.50,+0.50,-0.71)	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.70) [0.41°] $\widehat{\alpha}^R = 24.76$, $\widehat{\beta}^R = 230.26$	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.70) [0.38°] $\widehat{\alpha}^R = 24.65$, $\widehat{\beta}^R = 230.38$
(+0.00,+0.71,-0.71)	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.71,-0.70) [0.32°] $\widehat{\alpha}^R = 24.47$, $\widehat{\beta}^R = 230.48$	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.71,-0.70) [0.31°] $\widehat{\alpha}^R = 24.44$, $\widehat{\beta}^R = 230.51$
(-0.50,+0.50,-0.71)	$\widehat{\boldsymbol{s}}$ = (-0.50,+0.50,-0.70) [0.41°] $\widehat{\alpha}^R = 24.76$, $\widehat{\beta}^R = 230.26$	$\widehat{\boldsymbol{s}}$ = (-0.50,+0.50,-0.70) [0.38°] $\widehat{\alpha}^R = 24.65$, $\widehat{\beta}^R = 230.38$
(-0.71,+0.00,-0.71)	$\widehat{\boldsymbol{s}}$ = (-0.71,+0.00,-0.70) [0.32°] $\widehat{\alpha}^R = 24.47$, $\widehat{\beta}^R = 230.48$	$\widehat{\boldsymbol{s}}$ = (-0.71,+0.00,-0.70) [0.31°] $\widehat{\alpha}^R = 24.44$, $\widehat{\beta}^R = 230.51$
(-0.50,-0.50,-0.71)	$\widehat{\boldsymbol{s}}$ = (-0.50,-0.50,-0.70) [0.41°] $\widehat{\alpha}^R = 24.76$, $\widehat{\beta}^R = 230.26$	$\widehat{\boldsymbol{s}}$ = (-0.50,-0.50,-0.70) [0.38°] $\widehat{\alpha}^R = 24.65$, $\widehat{\beta}^R = 230.38$
(-0.00,-0.71,-0.71)	$\widehat{\boldsymbol{s}}$ = (+0.00,-0.71,-0.70) [0.32°] $\widehat{\alpha}^R = 24.47$, $\widehat{\beta}^R = 230.48$	$\widehat{\boldsymbol{s}}$ = (+0.00,-0.71,-0.70) [0.31°] $\widehat{\alpha}^R = 24.44$, $\widehat{\beta}^R = 230.51$
(+0.50,-0.50,-0.71)	$\widehat{\boldsymbol{s}}$ = (+0.50,-0.50,-0.70) [0.41°] $\widehat{\alpha}^R = 24.76$, $\widehat{\beta}^R = 230.26$	$\widehat{\boldsymbol{s}}$ = (+0.50,-0.50,-0.70) [0.38°] $\widehat{\alpha}^R = 24.65$, $\widehat{\beta}^R = 230.38$

CPAM		
$\boldsymbol{s}$	TEAV	TECA
(+0.71,+0.00,-0.71)	$\widehat{\boldsymbol{s}}$ = (+0.71,+0.00,-0.71) [0.00°] $\widehat{\alpha}^R = 20.41$, $\widehat{\beta}^R = 234.54$	$\widehat{\boldsymbol{s}}$ = (+0.71,+0.00,-0.71) [0.03°] $\widehat{\alpha}^R = 20.28$, $\widehat{\beta}^R = 234.69$
(+0.50,+0.50,-0.71)	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.04°] $\widehat{\alpha}^R = 20.25$, $\widehat{\beta}^R = 234.73$	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.06°] $\widehat{\alpha}^R = 20.18$, $\widehat{\beta}^R = 234.80$
(+0.00,+0.71,-0.71)	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.71,-0.71) [0.00°] $\widehat{\alpha}^R = 20.41$, $\widehat{\beta}^R = 234.54$	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.71,-0.71) [0.03°] $\widehat{\alpha}^R = 20.28$, $\widehat{\beta}^R = 234.69$
(-0.50,+0.50,-0.71)	$\widehat{\boldsymbol{s}}$ = (-0.50,+0.50,-0.71) [0.04°] $\widehat{\alpha}^R = 20.25$, $\widehat{\beta}^R = 234.73$	$\widehat{\boldsymbol{s}}$ = (-0.50,+0.50,-0.71) [0.06°] $\widehat{\alpha}^R = 20.18$, $\widehat{\beta}^R = 234.80$
(-0.71,+0.00,-0.71)	$\widehat{\boldsymbol{s}}$ = (-0.71,+0.00,-0.71) [0.00°] $\widehat{\alpha}^R = 20.41$, $\widehat{\beta}^R = 234.54$	$\widehat{\boldsymbol{s}}$ = (-0.71,+0.00,-0.71) [0.03°] $\widehat{\alpha}^R = 20.28$, $\widehat{\beta}^R = 234.69$
(-0.50,-0.50,-0.71)	$\widehat{\boldsymbol{s}}$ = (-0.50,-0.50,-0.71) [0.04°] $\widehat{\alpha}^R = 20.25$, $\widehat{\beta}^R = 234.73$	$\widehat{\boldsymbol{s}}$ = (-0.50,-0.50,-0.71) [0.06°] $\widehat{\alpha}^R = 20.18$, $\widehat{\beta}^R = 234.80$
(-0.00,-0.71,-0.71)	$\widehat{\boldsymbol{s}}$ = (+0.00,-0.71,-0.71) [0.00°] $\widehat{\alpha}^R = 20.41$, $\widehat{\beta}^R = 234.54$	$\widehat{\boldsymbol{s}}$ = (+0.00,-0.71,-0.71) [0.03°] $\widehat{\alpha}^R = 20.28$, $\widehat{\beta}^R = 234.69$
(+0.50,-0.50,-0.71)	$\widehat{\boldsymbol{s}}$ = (+0.50,-0.50,-0.71) [0.04°] $\widehat{\alpha}^R = 20.25$, $\widehat{\beta}^R = 234.73$	$\widehat{\boldsymbol{s}}$ = (+0.50,-0.50,-0.71) [0.06°] $\widehat{\alpha}^R = 20.18$, $\widehat{\beta}^R = 234.80$

$L_d^c A^c$ improve as z_0 increases for OPAM. As a final fact, the fitting error in all the cases studied lied between 0.2 and 0.4 for OPAM, while the linear correlation coefficients were above 0.99995. Again this reinforces the idea introduced in the last paragraph that points $(m_b, \mathrm{E}_{m_b}[\mu_{D^c}])$ still lie in a line in OPAM, although it is not the correct one.

Finally, in the fourth experiment, the center of the sphere was moved off the optical axis, keeping $\boldsymbol{s} = (1/2, 1/2, 1/\sqrt{2})$ all the time. The center of the sphere was displaced along a circular path centered at the optical axis, i.e. $x_0^2 + y_0^2 = \kappa^2$, with $z_0 = 3$ m. Tables 5.7 and 5.8 show the results obtained for, respectively, $\kappa = 0.01$ m and $\kappa = 0.15$ m under those circumstances. From the two tables, it can be noticed that the deviations in $L_a^c A^c$ and $L_d^c A^c$ for OPAM seem to keep constant and similar to the ones observed in experiment 3. The estimations of $\boldsymbol{s}$, however, tend to get worse as κ is increased (observe the angle between true and estimated $\boldsymbol{s}$ becomes 2-3° when $\kappa = 0.15$ m).

Table 5.6. Lighting parameters estimates using synthetic images generated under central projection, varying z_0. $\widehat{\alpha}^R$ stands for the estimate of $L_a^R A^R$ and $\widehat{\beta}^R$ for the estimate of $L_d^R A^R$. The angle between the true value of $\boldsymbol{s}$ and the estimation $\widehat{\boldsymbol{s}}$ is given between square brackets.

OPAM		
z_0 [m]	TEAV	TECA
1	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 36.71$, $\widehat{\beta}^R = 219.52$	$\widehat{\boldsymbol{s}}$ = (-0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 36.71$, $\widehat{\beta}^R = 219.52$
2	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 28.26$, $\widehat{\beta}^R = 227.48$	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 28.26$, $\widehat{\beta}^R = 227.48$
3	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 25.79$, $\widehat{\beta}^R = 229.64$	$\widehat{\boldsymbol{s}}$ = (+0.00,-0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 25.79$, $\widehat{\beta}^R = 229.64$
4	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 24.57$, $\widehat{\beta}^R = 230.70$	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 24.57$, $\widehat{\beta}^R = 230.70$
5	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 23.64$, $\widehat{\beta}^R = 231.54$	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 23.64$, $\widehat{\beta}^R = 231.54$
6	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 22.99$, $\widehat{\beta}^R = 232.27$	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 22.99$, $\widehat{\beta}^R = 232.27$
7	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 23.08$, $\widehat{\beta}^R = 232.01$	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 23.08$, $\widehat{\beta}^R = 232.01$
8	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 22.14$, $\widehat{\beta}^R = 233.00$	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 22.14$, $\widehat{\beta}^R = 233.00$

CPAM		
z_0 [m]	TEAV	TECA
1	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 20.40$, $\widehat{\beta}^R = 234.61$	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 20.40$, $\widehat{\beta}^R = 234.61$
2	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 20.43$, $\widehat{\beta}^R = 234.53$	$\widehat{\boldsymbol{s}}$ = (+0.00,-0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 20.43$, $\widehat{\beta}^R = 234.53$
3	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 20.44$, $\widehat{\beta}^R = 234.57$	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 20.44$, $\widehat{\beta}^R = 234.57$
4	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 20.34$, $\widehat{\beta}^R = 234.69$	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 20.34$, $\widehat{\beta}^R = 234.69$
5	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 20.40$, $\widehat{\beta}^R = 234.64$	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 20.40$, $\widehat{\beta}^R = 234.64$
6	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 20.07$, $\widehat{\beta}^R = 234.98$	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 20.07$, $\widehat{\beta}^R = 234.98$
7	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 20.32$, $\widehat{\beta}^R = 234.61$	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 20.32$, $\widehat{\beta}^R = 234.61$
8	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 19.84$, $\widehat{\beta}^R = 235.48$	$\widehat{\boldsymbol{s}}$ = (+0.00,+0.00,-1.00) [0.00°] $\widehat{\alpha}^R = 19.84$, $\widehat{\beta}^R = 235.48$

About the representativeness of the two values chosen for κ, notice that $\kappa = 0.01$ m corresponds to a small displacement $d = f \times \frac{\kappa}{z_0} = 0.033526 \times \frac{0.01}{3} = 1.12 \times 10^{-4}$ m or, in image units, $d = 1.12 \times 10^{-4}/9.9 \times 10^{-6} = 11.29$ pixels, using the size assumed for the cells, while a larger displacement of $d = 1.68e-3$ m or $d = 169.32$ pixels corresponds to $\kappa = 0.15$ m. In this last case, the images which are synthesized cover an area of 470×470 pixels, so that slightly increasing κ from 0.15 would mean the image would require a larger CCD than the one available from the camera used in the experiments with real images.

Summing up, CPAM behaves correctly in all the cases analyzed here, while OPAM, in general, yields more accurate estimations as z_0 is increased and the center of the sphere lies closer to the optical axis, as predicted in section 5.3. The inaccuracies in OPAM affect more to the estimation of $L_a^c A^c$ and $L_d^c A^c$, although they tend to be quite constant, as if, while $L_a^c A^c$ and $L_d^c A^c$ are not changed, points

Table 5.7. Lighting parameters estimates using synthetic images generated under central projection, varying x_0 and y_0 so that $x_0^2 + y_0^2 = 0.01^2$. $\widehat{\alpha}^R$ stands for the estimate of $L_a^R A^R$ and $\widehat{\beta}^R$ for the estimate of $L_d^R A^R$. The angle between the true value of $\boldsymbol{s}$ and the estimation $\widehat{\boldsymbol{s}}$ is given between square brackets.

OPAM		
(x_0, y_0) [m]	TEAV	TECA
(0.010,0.000)	$\widehat{\boldsymbol{s}} = (+0.51,+0.50,-0.70)$ [0.54°] $\widehat{\alpha}^R = 24.58$, $\widehat{\beta}^R = 230.40$	$\widehat{\boldsymbol{s}} = (+0.51,+0.50,-0.70)$ [0.56°] $\widehat{\alpha}^R = 24.65$, $\widehat{\beta}^R = 230.32$
(0.007,0.007)	$\widehat{\boldsymbol{s}} = (+0.51,+0.51,-0.70)$ [0.58°] $\widehat{\alpha}^R = 24.49$, $\widehat{\beta}^R = 230.47$	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.70)$ [0.54°] $\widehat{\alpha}^R = 24.31$, $\widehat{\beta}^R = 230.64$
(0.000,0.010)	$\widehat{\boldsymbol{s}} = (+0.50,+0.51,-0.70)$ [0.54°] $\widehat{\alpha}^R = 24.58$, $\widehat{\beta}^R = 230.40$	$\widehat{\boldsymbol{s}} = (+0.50,+0.51,-0.70)$ [0.56°] $\widehat{\alpha}^R = 24.65$, $\widehat{\beta}^R = 230.32$
(-0.007,0.007)	$\widehat{\boldsymbol{s}} = (+0.50,+0.51,-0.70)$ [0.42°] $\widehat{\alpha}^R = 24.53$, $\widehat{\beta}^R = 230.48$	$\widehat{\boldsymbol{s}} = (+0.50,+0.51,-0.70)$ [0.42°] $\widehat{\alpha}^R = 24.51$, $\widehat{\beta}^R = 230.48$
(-0.010,0.000)	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.70)$ [0.29°] $\widehat{\alpha}^R = 24.71$, $\widehat{\beta}^R = 230.26$	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.70)$ [0.29°] $\widehat{\alpha}^R = 24.63$, $\widehat{\beta}^R = 230.34$
(-0.007,-0.007)	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.71)$ [0.08°] $\widehat{\alpha}^R = 24.07$, $\widehat{\beta}^R = 230.86$	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.71)$ [0.06°] $\widehat{\alpha}^R = 24.00$, $\widehat{\beta}^R = 230.90$
(-0.000,-0.010)	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.70)$ [0.29°] $\widehat{\alpha}^R = 24.71$, $\widehat{\beta}^R = 230.26$	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.70)$ [0.29°] $\widehat{\alpha}^R = 24.63$, $\widehat{\beta}^R = 230.34$
(0.007,-0.007)	$\widehat{\boldsymbol{s}} = (+0.51,+0.50,-0.70)$ [0.42°] $\widehat{\alpha}^R = 24.53$, $\widehat{\beta}^R = 230.48$	$\widehat{\boldsymbol{s}} = (+0.51,+0.50,-0.70)$ [0.42°] $\widehat{\alpha}^R = 24.51$, $\widehat{\beta}^R = 230.48$

CPAM		
(x_0, y_0) [m]	TEAV	TECA
(0.010,0.000)	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.71)$ [0.04°] $\widehat{\alpha}^R = 20.54$, $\widehat{\beta}^R = 234.46$	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.71)$ [0.04°] $\widehat{\alpha}^R = 20.27$, $\widehat{\beta}^R = 234.71$
(0.007,0.007)	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.71)$ [0.07°] $\widehat{\alpha}^R = 20.14$, $\widehat{\beta}^R = 234.84$	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.71)$ [0.12°] $\widehat{\alpha}^R = 19.97$, $\widehat{\beta}^R = 234.97$
(0.000,0.010)	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.71)$ [0.04°] $\widehat{\alpha}^R = 20.54$, $\widehat{\beta}^R = 234.46$	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.71)$ [0.04°] $\widehat{\alpha}^R = 20.27$, $\widehat{\beta}^R = 234.71$
(-0.007,0.007)	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.71)$ [0.02°] $\widehat{\alpha}^R = 20.29$, $\widehat{\beta}^R = 234.72$	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.71)$ [0.09°] $\widehat{\alpha}^R = 20.06$, $\widehat{\beta}^R = 234.91$
(-0.010,0.000)	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.71)$ [0.02°] $\widehat{\alpha}^R = 20.49$, $\widehat{\beta}^R = 234.51$	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.71)$ [0.01°] $\widehat{\alpha}^R = 20.40$, $\widehat{\beta}^R = 234.59$
(-0.007,-0.007)	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.71)$ [0.09°] $\widehat{\alpha}^R = 20.03$, $\widehat{\beta}^R = 234.96$	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.71)$ [0.11°] $\widehat{\alpha}^R = 19.95$, $\widehat{\beta}^R = 235.06$
(-0.000,-0.010)	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.71)$ [0.02°] $\widehat{\alpha}^R = 20.49$, $\widehat{\beta}^R = 234.51$	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.71)$ [0.01°] $\widehat{\alpha}^R = 20.40$, $\widehat{\beta}^R = 234.59$
(0.007,-0.007)	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.71)$ [0.02°] $\widehat{\alpha}^R = 20.29$, $\widehat{\beta}^R = 234.72$	$\widehat{\boldsymbol{s}} = (+0.50,+0.50,-0.71)$ [0.09°] $\widehat{\alpha}^R = 20.06$, $\widehat{\beta}^R = 234.91$

$(m_b, \mathrm{E}_{m_b}[\mu_{D^c}])$ lied in a line different to the correct one but always the same. The fitting error increases as the sphere moves off the optical axis. As for methods TEAV and TECA, results for synthetic images do not show a higher supremacy of none of them over the other.

To finish, figure 5.14 summarizes the results presented throughout the section. In the graphs, the length of every bar represents the maximum error attained during the experiment, while the average value appears as another bar superimposed over the previous one.

5.6.3 Robustness measures

Four experiments to prove the robustness of the estimation procedures are presented in the following. In the first experiment, the same calibration image served as input to CPAM and OPAM several times, using each time a different starting contour, in order to determine the sensitivity of the estimation

Table 5.8. Lighting parameters estimates using synthetic images generated under central projection, varying x_0 and y_0 so that $x_0^2 + y_0^2 = 0.15^2$. $\widehat{\alpha}^R$ stands for the estimate of $L_a^R A^R$ and $\widehat{\beta}^R$ for the estimate of $L_d^R A^R$, both for the red channel. The angle between the true value of $\boldsymbol{s}$ and the estimation $\widehat{\boldsymbol{s}}$ is given between square brackets.

OPAM		
(x_0, y_0) [m]	TEAV	TECA
(0.150,0.000)	$\widehat{\boldsymbol{s}}$ = (+0.54,+0.50,-0.68) [2.75°] $\widehat{\alpha}^R = 24.53$, $\widehat{\beta}^R = 230.43$	$\widehat{\boldsymbol{s}}$ = (+0.54,+0.50,-0.68) [2.80°] $\widehat{\alpha}^R = 24.42$, $\widehat{\beta}^R = 230.56$
(0.106,0.106)	$\widehat{\boldsymbol{s}}$ = (+0.53,+0.53,-0.67) [3.23°] $\widehat{\alpha}^R = 24.74$, $\widehat{\beta}^R = 230.17$	$\widehat{\boldsymbol{s}}$ = (+0.53,+0.53,-0.67) [3.23°] $\widehat{\alpha}^R = 24.72$, $\widehat{\beta}^R = 230.19$
(0.000,0.150)	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.54,-0.68) [2.75°] $\widehat{\alpha}^R = 24.53$, $\widehat{\beta}^R = 230.43$	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.54,-0.68) [2.80°] $\widehat{\alpha}^R = 24.42$, $\widehat{\beta}^R = 230.56$
(-0.106,0.106)	$\widehat{\boldsymbol{s}}$ = (+0.48,+0.53,-0.70) [2.07°] $\widehat{\alpha}^R = 24.71$, $\widehat{\beta}^R = 230.29$	$\widehat{\boldsymbol{s}}$ = (+0.48,+0.53,-0.70) [1.99°] $\widehat{\alpha}^R = 24.68$, $\widehat{\beta}^R = 230.34$
(-0.150,0.000)	$\widehat{\boldsymbol{s}}$ = (+0.47,+0.50,-0.73) [2.30°] $\widehat{\alpha}^R = 24.26$, $\widehat{\beta}^R = 230.69$	$\widehat{\boldsymbol{s}}$ = (+0.47,+0.50,-0.73) [2.24°] $\widehat{\alpha}^R = 24.27$, $\widehat{\beta}^R = 230.68$
(-0.106,-0.106)	$\widehat{\boldsymbol{s}}$ = (+0.48,+0.48,-0.74) [2.52°] $\widehat{\alpha}^R = 24.26$, $\widehat{\beta}^R = 230.71$	$\widehat{\boldsymbol{s}}$ = (+0.48,+0.48,-0.74) [2.54°] $\widehat{\alpha}^R = 24.17$, $\widehat{\beta}^R = 230.81$
(-0.000,-0.150)	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.47,-0.73) [2.30°] $\widehat{\alpha}^R = 24.26$, $\widehat{\beta}^R = 230.69$	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.47,-0.73) [2.24°] $\widehat{\alpha}^R = 24.27$, $\widehat{\beta}^R = 230.68$
(0.106,-0.106)	$\widehat{\boldsymbol{s}}$ = (+0.53,+0.48,-0.70) [2.07°] $\widehat{\alpha}^R = 24.71$, $\widehat{\beta}^R = 230.29$	$\widehat{\boldsymbol{s}}$ = (+0.53,+0.48,-0.70) [1.99°] $\widehat{\alpha}^R = 24.68$, $\widehat{\beta}^R = 230.34$

CPAM		
(x_0, y_0) [m]	TEAV	TECA
(0.150,0.000)	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.10°] $\widehat{\alpha}^R = 20.40$, $\widehat{\beta}^R = 234.57$	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.10°] $\widehat{\alpha}^R = 20.37$, $\widehat{\beta}^R = 234.59$
(0.106,0.106)	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.05°] $\widehat{\alpha}^R = 20.91$, $\widehat{\beta}^R = 234.15$	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.04°] $\widehat{\alpha}^R = 20.56$, $\widehat{\beta}^R = 234.45$
(0.000,0.150)	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.10°] $\widehat{\alpha}^R = 20.40$, $\widehat{\beta}^R = 234.57$	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.10°] $\widehat{\alpha}^R = 20.37$, $\widehat{\beta}^R = 234.59$
(-0.106,0.106)	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.05°] $\widehat{\alpha}^R = 20.57$, $\widehat{\beta}^R = 234.45$	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.09°] $\widehat{\alpha}^R = 20.42$, $\widehat{\beta}^R = 234.59$
(-0.150,0.000)	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.06°] $\widehat{\alpha}^R = 20.53$, $\widehat{\beta}^R = 234.51$	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.08°] $\widehat{\alpha}^R = 20.49$, $\widehat{\beta}^R = 234.54$
(-0.106,-0.106)	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.03°] $\widehat{\alpha}^R = 20.89$, $\widehat{\beta}^R = 234.15$	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.00°] $\widehat{\alpha}^R = 20.80$, $\widehat{\beta}^R = 234.23$
(-0.000,-0.150)	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.06°] $\widehat{\alpha}^R = 20.53$, $\widehat{\beta}^R = 234.51$	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.08°] $\widehat{\alpha}^R = 20.49$, $\widehat{\beta}^R = 234.54$
(0.106,-0.106)	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.05°] $\widehat{\alpha}^R = 20.57$, $\widehat{\beta}^R = 234.45$	$\widehat{\boldsymbol{s}}$ = (+0.50,+0.50,-0.71) [0.09°] $\widehat{\alpha}^R = 20.42$, $\widehat{\beta}^R = 234.59$

methods against the initial circle/ellipse. The second experiment consisted in facing OPAM and CPAM against synthetic images with varying levels of zero-mean Gaussian additive noise. In the third experiment, a set of images taken under the same lighting direction but with different lens apertures, i.e. the same $\boldsymbol{s}$ but different apparent values for L_a and L_d, were analyzed. Finally, the fourth experiment consisted in changing $\boldsymbol{s}$ but keeping constant L_a and L_d. In all the experiments but the second one, real images were used.

Sensitivity regarding the initial contour

In order to determine the sensitivity of the different estimation methods against the contour given as input, the calibration image of figure 5.15(left) was processed up to four times using a different

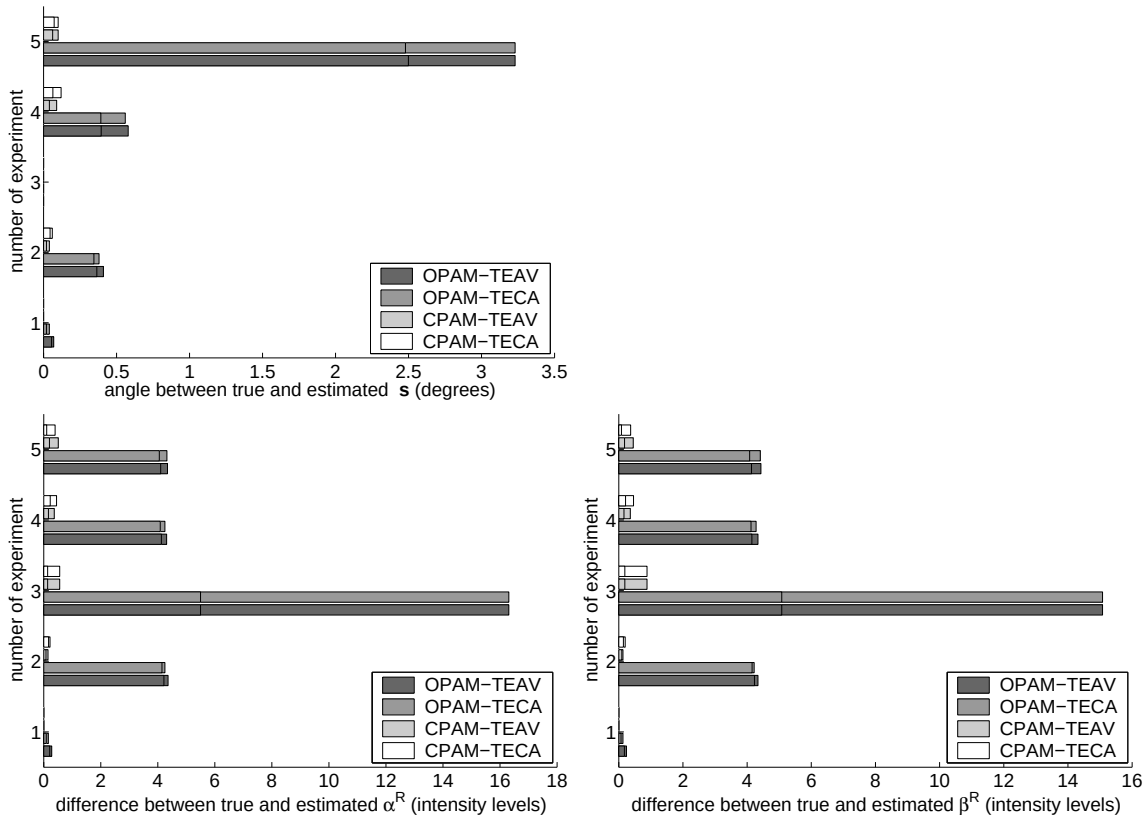


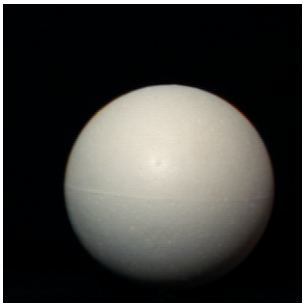
Fig. 5.14. Summary of results for *orthographic versus perspective* projection experiments: [top,left] angle between true and estimated $\boldsymbol{s}$; [bottom,left] difference between true and estimated α^R; [bottom,right] difference between true and estimated β^R. (Average-value bars are shown superimposed over maximum-value bars.)

starting contour in each trial. Since the estimations are obtained through an evolutionary optimization approach, justified by the complexity of the estimation task, the corresponding values must always be expected sub-optimum solutions, and, hence, different values will likely result in each trial. Nevertheless, the estimates produced by the algorithms are quite close to one another, as it is shown in tables 5.15(right,a-d). Coherently with other experiments, the estimations of $\boldsymbol{s}$ tend to be more concentrated than the estimates of the illumination strengths.

Robustness against Gaussian noise

To evaluate how robust the estimation methods were against noise, a set of noisy images, for which L_a and L_d was kept constant but the light direction tilt varied in 45° steps, was considered. For every level of noise, 5 images were generated and analyzed, and the mean estimate was kept. By way of illustration, figure 5.16 shows the sort of degradation introduced by every level noise in one of the synthetic images used in the experiment. It is important to notice that this experiment did not intend to observe the behaviour of the estimated methods against images with a noise distribution derived from the radiometric performance of the camera, what will be discussed in section 6.2.1. The purpose was rather to put the estimation methods on increasingly more challenging situations.

Results for OPAM are given in figure 5.17, where: the uppermost row of plots shows the angle in degrees between true and estimated $\boldsymbol{s}$, true and estimated vectors $\boldsymbol{L}_a$ and true and estimated vectors $\boldsymbol{L}_d$; and the lowermost row corresponds to the absolute difference between the norms of the true and the estimated lighting strengths: i.e. $\left|\left\|\boldsymbol{L}_a\right\| - \left\|\widehat{\boldsymbol{L}}_a\right\|\right|$ and $\left|\left\|\boldsymbol{L}_d\right\| - \left\|\widehat{\boldsymbol{L}}_d\right\|\right|$, respectively. Figure 5.18 shows the same type of results as before but regarding CPAM.



(a)

trial	$\widehat{\boldsymbol{s}}$	$\widehat{\boldsymbol{\alpha}}$	$\widehat{\boldsymbol{\beta}}$
1	(-0.18,0.55,-0.82)	(41.63, 42.81, 42.43)	(188.70,182.63,165.40)
2	(-0.18,0.55,-0.82)	(42.70, 43.64, 43.28)	(187.55,181.78,164.49)
3	(-0.18,0.55,-0.82)	(45.16, 46.01, 45.52)	(184.91,179.17,162.03)
4	(-0.18,0.55,-0.82)	(42.92, 43.86, 43.46)	(187.22,181.36,164.20)

(b)

trial	$\widehat{\boldsymbol{s}}$	$\widehat{\boldsymbol{\alpha}}$	$\widehat{\boldsymbol{\beta}}$
1	(-0.16,0.54,-0.82)	(41.86, 43.24, 42.92)	(188.34,181.83,164.68)
2	(-0.17,0.55,-0.82)	(44.14, 45.27, 44.81)	(186.04,179.91,162.78)
3	(-0.16,0.55,-0.82)	(42.02, 43.18, 42.88)	(188.24,182.11,164.84)
4	(-0.17,0.55,-0.82)	(42.59, 43.62, 43.30)	(187.64,181.72,164.42)

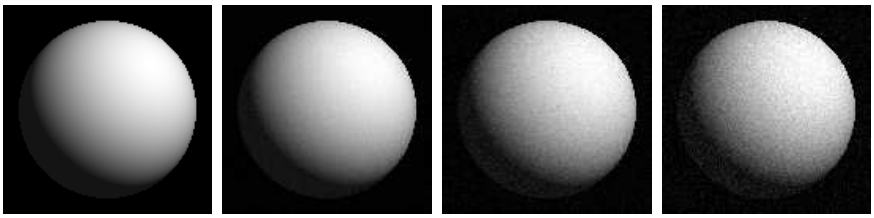
(c)

trial	$\widehat{\boldsymbol{s}}$	$\widehat{\boldsymbol{\alpha}}$	$\widehat{\boldsymbol{\beta}}$
1	(-0.18,0.55,-0.81)	(37.78, 38.73, 38.70)	(192.02,186.34,168.77)
2	(-0.18,0.55,-0.81)	(37.74, 38.58, 38.62)	(192.20,186.57,168.98)
3	(-0.18,0.55,-0.81)	(36.53, 37.42, 37.44)	(193.38,187.80,170.21)
4	(-0.18,0.55,-0.81)	(36.60, 37.53, 37.54)	(193.30,187.66,170.07)

(d)

trial	$\widehat{\boldsymbol{s}}$	$\widehat{\boldsymbol{\alpha}}$	$\widehat{\boldsymbol{\beta}}$
1	(-0.17,0.55,-0.81)	(38.68, 39.89, 39.84)	(191.05,185.01,167.42)
2	(-0.17,0.55,-0.82)	(34.63, 35.63, 35.74)	(195.43,189.58,172.05)
3	(-0.17,0.55,-0.82)	(35.01, 36.15, 36.24)	(194.98,188.97,171.42)
4	(-0.17,0.55,-0.82)	(35.88, 36.88, 36.93)	(194.08,188.25,170.73)

Fig. 5.15. Results for robustness experiment 1: [left] calibration image used; [right] (a) OPAM-TEAV, (b) OPAM-TECA, (c) CPAM-TEAV, (d) CPAM-TECA. In the tables, $\widehat{\boldsymbol{\alpha}}$ stands for the vector of estimates of $L_a^c A^c + \mathrm{E}_I[\mu_{dc}]A^c$ and $\widehat{\boldsymbol{\beta}}$ for the vector of estimates of $L_d^c A^c$, $c \in \{R, G, B\}$.



(a) (b) (c) (d)

Fig. 5.16. Examples of noisy images used in robustness experiment 2: (a) original image; (b) $\sigma = 5$; (c) $\sigma = 10$; (d) $\sigma = 15$.

As can be seen, the estimates of $\boldsymbol{s}$ are quite accurate even for high levels of noise. The same applies to the angle between true and estimated L_a and L_d, which means the colour of ambient and directional illumination is accurately estimated even under noisy conditions. The magnitude of the illumination strengths are, however, more affected by noise. Comparing OPAM and CPAM, it can be seen that both present the same level of robustness, taking into account that OPAM tends to introduce an almost constant error in the estimation of L_a and L_d, as shown in previous experiments. As for the behaviour of TEAV and TECA, the plots clearly show that the former is, in general, more robust, although only the deviations in the estimated illumination strengths are of importance in the latter.

Robustness of s estimations against changes in L_a and L_d

In this third experiment, 20 images were taken, varying the F-number of the camera from one image to the next, but without altering the geometrical relationship between camera and calibration object, so

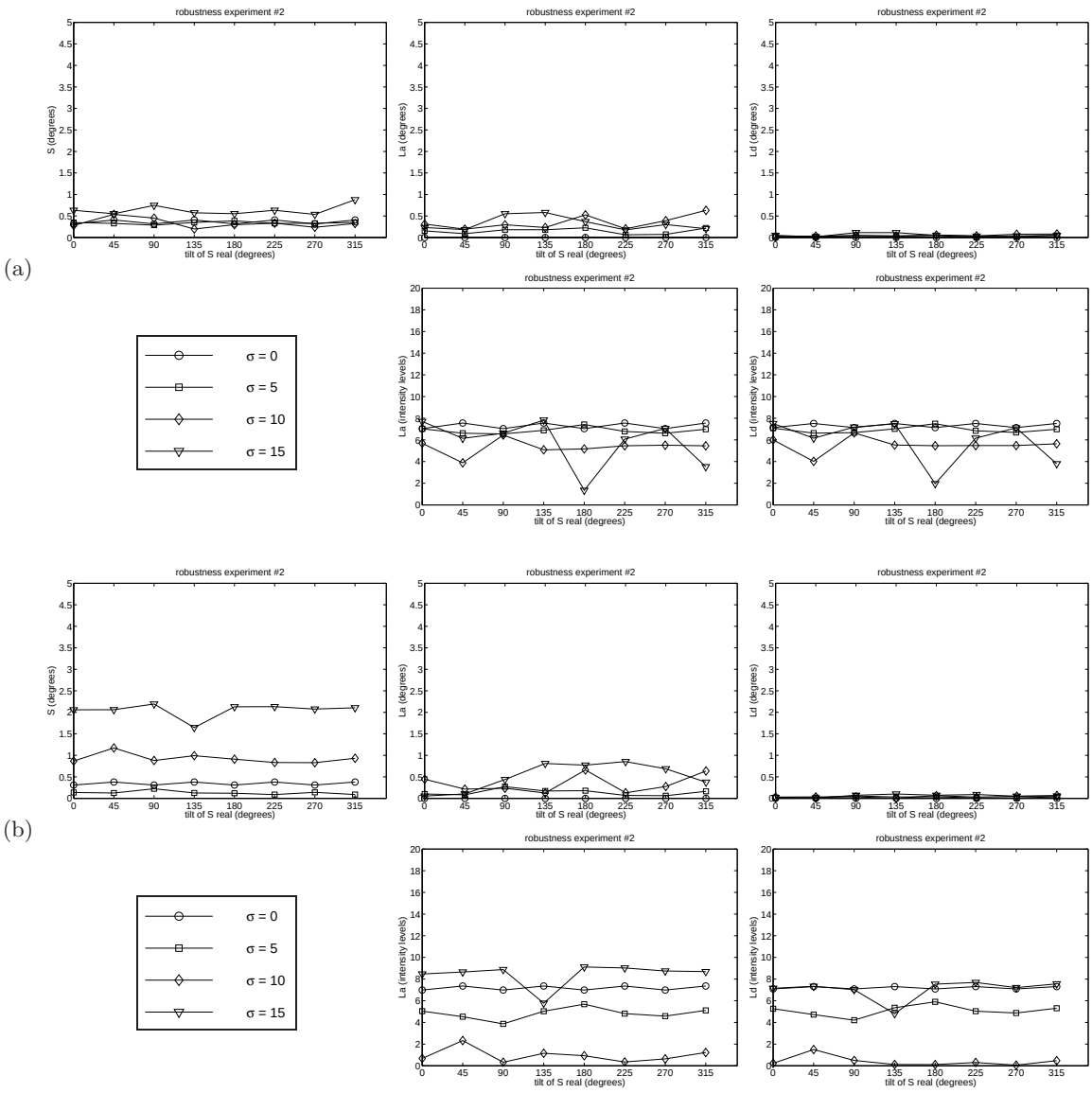


Fig. 5.17. Results for robustness experiment 2 and OPAM: (a) TEAV, (b) TECA. (σ is the standard deviation, in intensity levels, of the zero-mean Gaussian noise added to images.)

that $\boldsymbol{s}$ kept constant. Therefore, the application of the estimation methods to these 20 images should yield very similar estimates for $\boldsymbol{s}$ in the different cases, although $L_a^c A^c + \mathrm{E}_I[\mu_{dc}]A^c$ and $L_d^c A^c$ would vary accordingly. Figure 5.19 shows histograms of the difference angles between the estimations of $\boldsymbol{s}$ for every possible pair of images of the set. The mean and maximum difference angles for the different cases are given underneath every histogram. As can be appreciated, OPAM does not seem to be as robust as CPAM in this case, although both provide a quite adequate performance. On the other hand, TECA seems to be more robust than TEAV in both CPAM and OPAM, although for CPAM the difference is more remarkable.

Robustness of L_a and L_d estimations against changes in s

Finally, in the fourth experiment, the lighting orientation was varied without modifying the amount of light entering the camera to test the robustness of $L_a^c A^c + \mathrm{E}_I[\mu_{dc}]$ and $L_d^c A^c$ estimations. First, three

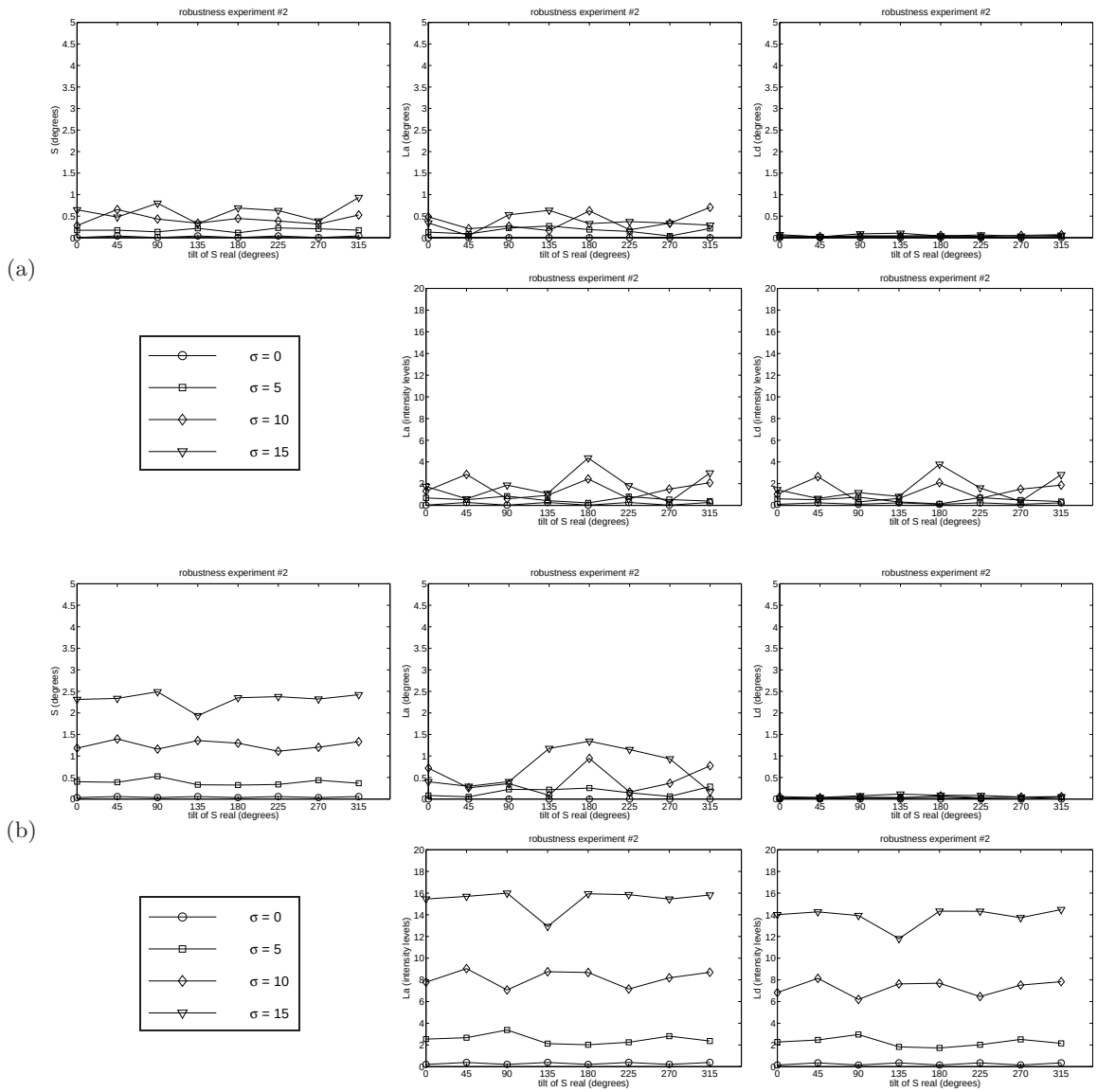


Fig. 5.18. Results for robustness experiment 2 and CPAM: (a) TEAV, (b) TECA. (σ is the standard deviation, in intensity levels, of the zero-mean Gaussian noise added to images.)

real calibration images were rotated four times 90 degrees, as well as the corresponding parameters of the sphere elliptical contour in each case. Results appear in figures 5.20 and 5.21, showing there is no appreciable difference in the estimates of $L_a^c A^c + \mathrm{E}_I[\mu_{dc}]$ and $L_d^c A^c$.

In a second step of this experiment, the camera was rotated and moved eight times, without modifying the position and strength of the light source. The corresponding estimates are given, respectively for OPAM and CPAM, in tables 5.9 and 5.10. Besides, some statistics are provided at the bottom in the same tables, where $|\cdot|$ is the vectors norm, $\mathrm{avg}\{\cdot\}$ and $\sigma.$ stand for, respectively, mean and standard deviation, $255\sqrt{3}$ is the maximum achievable norm of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and $\widehat{\boldsymbol{v},\boldsymbol{w}}$ is the angle between vectors $\boldsymbol{v}$ and $\boldsymbol{w}$. As can be observed, the behaviour is quite similar in all four cases considered, with almost the same statistical figures. As for the observed deviations in the estimates of $\boldsymbol{s}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ from one configuration to the next, it must be taken into account that the rotation of the camera between configurations implied recomputing the parameters of the contour fitting the projection of the sphere,

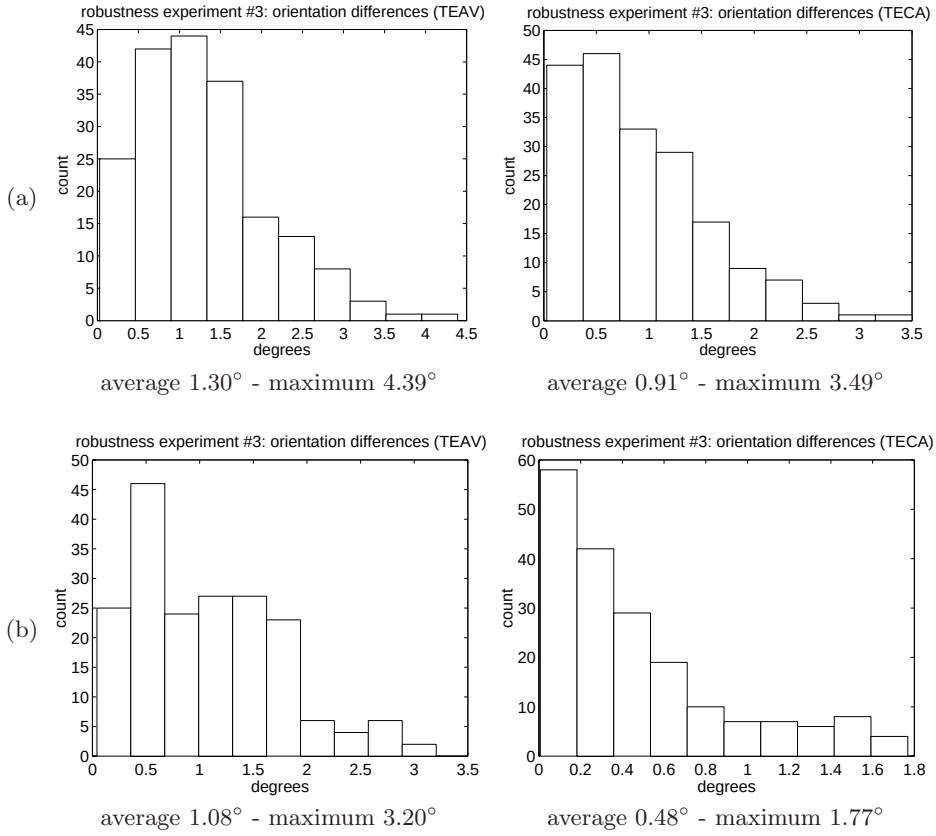


Fig. 5.19. Results for robustness experiment 3: (a) OPAM, (b) CPAM.

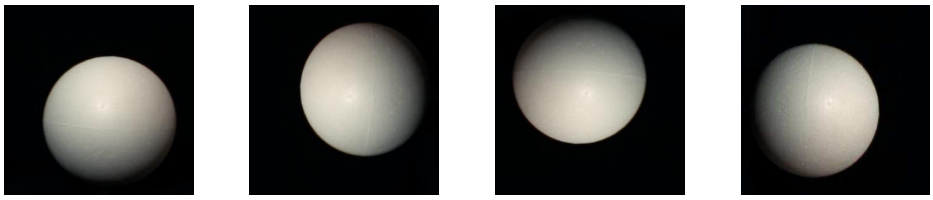
so that the deviations also observed in robustness experiment 1 (see figure 5.15) must also be expected here.

5.6.4 Comparison with other lighting direction estimation methods

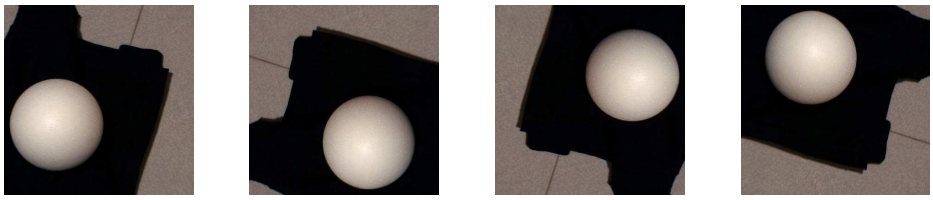
This section compares OPAM and CPAM with two well-known methods often cited in the literature about shape from shading: Lee and Rosenfeld's (L&R) [141] and Zheng and Chellappa's (Z&C) [311]. Both methods estimate the light source direction from the statistics of the image, assuming a certain distribution for surface normals, as already explained in section 5.5, but without the need of estimating shape information; in other words, both are *stand-alone* estimators. Besides, they share the common assumption of locally spherical scene surfaces. In this sense, comparing them with OPAM and CPAM, which also are stand-alone, makes sense.

The comparison is performed over two sets of noiseless synthetic images generated under different slant and tilt angles of the light source: one set consists of images generated under orthographic projection, while perspective projection is assumed in the other one; furthermore, in the second set, the sphere is located slightly off the axis so that the sphere projection is not exactly a circle [5]. Since neither L&R or Z&C methods were designed on the basis of an image formation model including ambient lighting, results are also provided discounting first the ambient reflection term corresponding to the sphere. Furthermore, as for the L&R method, both the original formulation and the modified version proposed by Zheng and Chellapa in [311] have been included in the comparison. The implementation details are also according to [311]. OPAM and CPAM are given the exact projection parameters, while

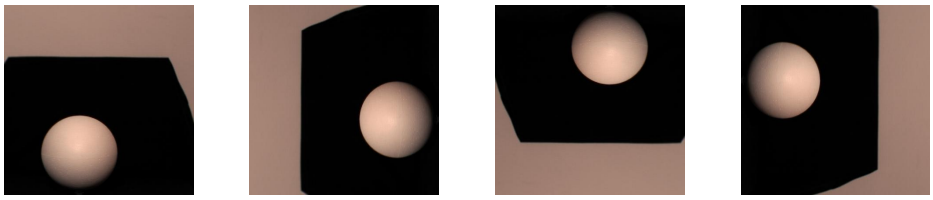
[5] The center of the sphere was shifted $0.025 \times \frac{0.033526}{3}/9.9 \times 10^{-6} \approx 28$ pixels from the center of the image (see section 5.6.2 with regard to this calculation).



$\widehat{\boldsymbol{s}}$	(-0.18,0.55,-0.82)	(-0.55,-0.18,-0.82)	(0.18,-0.55,-0.82)	(0.55,0.18,-0.82)
$\widehat{\boldsymbol{\alpha}}$	(43.22, 44.21, 43.81)	(43.22, 44.21, 43.81)	(43.22, 44.21, 43.81)	(43.22, 44.21, 43.81)
$\widehat{\boldsymbol{\beta}}$	(187.00,181.15,163.90)	(187.00,181.15,163.90)	(187.00,181.15,163.90)	(187.00,181.15,163.90)
$\widehat{\boldsymbol{s}}$	(-0.17,0.55,-0.82)	(-0.55,-0.17,-0.82)	(0.17,-0.55,-0.82)	(0.55,0.17,-0.82)
$\widehat{\boldsymbol{\alpha}}$	(42.83, 43.84, 43.50)	(42.83, 43.84, 43.50)	(42.83, 43.84, 43.50)	(42.83, 43.84, 43.50)
$\widehat{\boldsymbol{\beta}}$	(187.40,181.52,164.22)	(187.40,181.52,164.22)	(187.40,181.52,164.22)	(187.40,181.52,164.22)



$\widehat{\boldsymbol{s}}$	(-0.15,0.28,-0.95)	(-0.28,-0.15,-0.95)	(0.15,-0.28,-0.95)	(0.28,0.15,-0.95)
$\widehat{\boldsymbol{\alpha}}$	(75.97, 62.20, 60.69)	(75.97, 62.20, 60.69)	(75.97, 62.20, 60.69)	(75.97, 62.20, 60.69)
$\widehat{\boldsymbol{\beta}}$	(155.76,157.08,140.95)	(155.76,157.08,140.95)	(155.76,157.08,140.95)	(155.76,157.08,140.95)
$\widehat{\boldsymbol{s}}$	(-0.21,0.20,-0.96)	(-0.20,-0.21,-0.96)	(0.21,-0.20,-0.96)	(0.20,0.21,-0.96)
$\widehat{\boldsymbol{\alpha}}$	(71.99, 56.98, 55.80)	(71.99, 56.98, 55.80)	(71.99, 56.98, 55.80)	(71.99, 56.98, 55.80)
$\widehat{\boldsymbol{\beta}}$	(160.27,163.63,147.25)	(160.27,163.63,147.25)	(160.27,163.63,147.25)	(160.27,163.63,147.25)

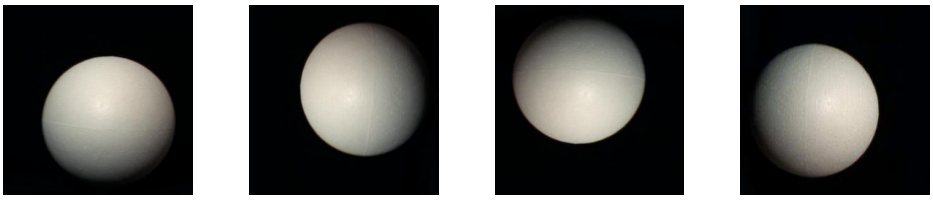


$\widehat{\boldsymbol{s}}$	(-0.05,0.49,-0.87)	(-0.49,-0.05,-0.87)	(0.05,-0.49,-0.87)	(0.49,0.05,-0.87)
$\widehat{\boldsymbol{\alpha}}$	(50.52, 38.41, 35.70)	(50.52, 38.41, 35.70)	(50.52, 38.41, 35.70)	(50.52, 38.41, 35.70)
$\widehat{\boldsymbol{\beta}}$	(191.46,158.62,135.05)	(191.46,158.62,135.05)	(191.46,158.62,135.05)	(191.46,158.62,135.05)
$\widehat{\boldsymbol{s}}$	(-0.05,0.49,-0.87)	(-0.49,-0.05,-0.87)	(0.05,-0.49,-0.87)	(0.49,0.05,-0.87)
$\widehat{\boldsymbol{\alpha}}$	(49.82, 38.19, 35.66)	(49.82, 38.19, 35.66)	(49.82, 38.19, 35.66)	(49.82, 38.19, 35.66)
$\widehat{\boldsymbol{\beta}}$	(192.34,158.85,135.08)	(192.34,158.85,135.08)	(192.34,158.85,135.08)	(192.34,158.85,135.08)

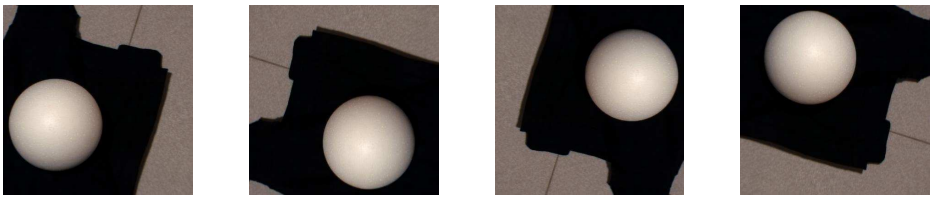
Fig. 5.20. Results for robustness experiment 4-1 and OPAM. In the tables, $\widehat{\boldsymbol{\alpha}}$ stands for the vector of estimates of $L_a^c A^c + \mathrm{E}_I[\mu_{dc}]A^c$ and $\widehat{\boldsymbol{\beta}}$ for the vector of estimates of $L_d^c A^c$, $c \in \{R, G, B\}$. For every set of images, the first table corresponds to TEAV, while the second one is for TECA.

L&R and Z&C methods are computed only over the pixels enclosed by the true sphere projection contour. Finally, $L_a^c A^c = 0.08 \times 255 = 20.4$ and $L_d^c A^c = 0.92 \times 255 = 234.6$, for every colour channel c.

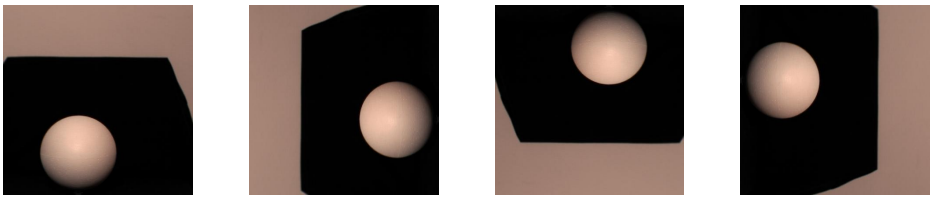
To begin with, table 5.11 gives the results corresponding to spheres generated under orthographic projection. OPAM results are thus collected. In the table, lighting directions are indicated through the slant and tilt angles notation, instead of using vector notation, unlike other occasions. This is to make easier the interpretation of results for the different slant and tilt angles examined. Besides, columns labeled as Δ correspond to the angle, in degrees, between the estimated and the true lighting orientation. As can be observed, TEAV and TECA methods produce the best estimations. Both versions of the L&R method behave quite correctly but only after having discounted the ambient reflection term of the sphere; otherwise, the errors are quite important. As for the Z&C method, surprisingly, it



$\widehat{\boldsymbol{s}}$	(-0.18,0.56,-0.81)	(-0.56,-0.18,-0.81)	(0.18,-0.56,-0.81)	(0.56,0.18,-0.81)
$\widehat{\boldsymbol{\alpha}}$	(37.07, 38.14, 38.49)	(37.07, 38.14, 38.49)	(37.07, 38.14, 38.49)	(37.07, 38.14, 38.49)
$\widehat{\boldsymbol{\beta}}$	(192.95,187.29,169.25)	(192.95,187.29,169.25)	(192.95,187.29,169.25)	(192.95,187.29,169.25)
$\widehat{\boldsymbol{s}}$	(-0.17,0.56,-0.81)	(-0.56,-0.17,-0.81)	(0.17,-0.56,-0.81)	(0.56,0.17,-0.81)
$\widehat{\boldsymbol{\alpha}}$	(36.62, 37.71, 38.11)	(36.62, 37.71, 38.11)	(36.62, 37.71, 38.11)	(36.62, 37.71, 38.11)
$\widehat{\boldsymbol{\beta}}$	(193.25,187.62,169.57)	(193.25,187.62,169.57)	(193.25,187.62,169.57)	(193.25,187.62,169.57)



$\widehat{\boldsymbol{s}}$	(-0.13,0.29,-0.95)	(-0.29,-0.13,-0.95)	(0.13,-0.29,-0.95)	(0.29,0.13,-0.95)
$\widehat{\boldsymbol{\alpha}}$	(81.25, 67.26, 67.21)	(81.25, 67.26, 67.21)	(81.25, 67.26, 67.21)	(81.25, 67.26, 67.21)
$\widehat{\boldsymbol{\beta}}$	(150.78,151.62,133.27)	(150.78,151.62,133.27)	(150.78,151.62,133.27)	(150.78,151.62,133.27)
$\widehat{\boldsymbol{s}}$	(-0.18,0.26,-0.95)	(-0.26,-0.18,-0.95)	(0.18,-0.26,-0.95)	(0.26,0.18,-0.95)
$\widehat{\boldsymbol{\alpha}}$	(82.14, 66.61, 65.28)	(82.14, 66.61, 65.28)	(82.14, 66.61, 65.28)	(82.14, 66.61, 65.28)
$\widehat{\boldsymbol{\beta}}$	(148.32,152.30,135.85)	(148.32,152.30,135.85)	(148.32,152.30,135.85)	(148.32,152.30,135.85)



$\widehat{\boldsymbol{s}}$	(-0.04,0.51,-0.86)	(-0.51,-0.04,-0.86)	(0.04,-0.51,-0.86)	(0.51,0.04,-0.86)
$\widehat{\boldsymbol{\alpha}}$	(46.07, 36.22, 35.02)	(46.07, 36.22, 35.02)	(46.07, 36.22, 35.02)	(46.07, 36.22, 35.02)
$\widehat{\boldsymbol{\beta}}$	(196.05,159.90,134.70)	(196.05,159.90,134.70)	(196.05,159.90,134.70)	(196.05,159.90,134.70)
$\widehat{\boldsymbol{s}}$	(-0.04,0.51,-0.86)	(-0.51,-0.04,-0.86)	(0.04,-0.51,-0.86)	(0.51,0.04,-0.86)
$\widehat{\boldsymbol{\alpha}}$	(45.78, 35.98, 34.84)	(45.78, 35.98, 34.84)	(45.78, 35.98, 34.84)	(45.78, 35.98, 34.84)
$\widehat{\boldsymbol{\beta}}$	(196.29,160.12,134.84)	(196.29,160.12,134.84)	(196.29,160.12,134.84)	(196.29,160.12,134.84)

Fig. 5.21. Results for robustness experiment 4-1 and CPAM. In the tables, $\widehat{\boldsymbol{\alpha}}$ stands for the vector of estimates of $L_a^c A^c + \mathrm{E}_I[\mu_{dc}]A^c$ and $\widehat{\boldsymbol{\beta}}$ for the vector of estimates of $L_d^c A^c$, $c \in \{R, G, B\}$. For every set of images, the first table corresponds to TEAV, while the second one is for TECA.

behaves better when the ambient reflection term is not removed from the sphere projection, although, in general, gives rise to higher errors than TEAV and TECA methods.

The results for images under perspective projection are given in table 5.12. As can be seen, L&R and Z&C algorithms behave in the same way as for the set of orthographically-generated images, with very small variations among the estimations for both set of images. Under CPAM in this case, TEAV and TECA methods also produce estimations with inaccuracies below one tenth of degree, as before.

(a)

# image	$\widehat{s}$	$\widehat{\boldsymbol{\alpha}}$	$\widehat{\boldsymbol{\beta}}$
1	(-0.73,-0.24,-0.64)	(27.48,22.13,19.93)	(184.97,181.74,172.97)
2	(-0.75,-0.12,-0.65)	(28.33,23.39,21.41)	(181.02,176.86,167.83)
3	(-0.75,0.01,-0.66)	(30.57,25.84,23.75)	(178.14,173.72,164.43)
4	(-0.75,0.10,-0.66)	(30.98,26.68,24.91)	(177.98,172.94,164.63)
5	(-0.72,0.22,-0.66)	(33.03,28.70,27.00)	(179.77,175.41,166.89)
6	(-0.73,-0.26,-0.62)	(30.11,26.08,22.96)	(170.63,165.35,158.23)
7	(-0.74,-0.29,-0.61)	(34.98,30.65,27.27)	(169.63,164.10,157.97)
8	(-0.73,-0.35,-0.58)	(32.23,27.78,24.42)	(176.45,171.04,164.95)

$\frac{\sigma_{|\boldsymbol{\alpha}|}}{\text{avg}\{|\boldsymbol{\alpha}|\}} \times 100 = 9.31\%$ $\frac{\sigma_{|\boldsymbol{\alpha}|}}{255\sqrt{3}} \times 100 = 1.00\%$ $\text{avg}\{\widehat{\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j}\} = 1.33^\circ$ $\max\{\widehat{\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j}\} = 2.88^\circ$

$\frac{\sigma_{|\boldsymbol{\beta}|}}{\text{avg}\{|\boldsymbol{\beta}|\}} \times 100 = 3.08\%$ $\frac{\sigma_{|\boldsymbol{\beta}|}}{255\sqrt{3}} \times 100 = 2.07\%$ $\text{avg}\{\widehat{\boldsymbol{\beta}_i, \boldsymbol{\beta}_j}\} = 0.24^\circ$ $\max\{\widehat{\boldsymbol{\beta}_i, \boldsymbol{\beta}_j}\} = 0.44^\circ$

(b)

# image	$\widehat{s}$	$\widehat{\boldsymbol{\alpha}}$	$\widehat{\boldsymbol{\beta}}$
1	(-0.74,-0.20,-0.64)	(26.93,21.79,19.79)	(185.59,181.84,172.94)
2	(-0.75,-0.08,-0.65)	(27.37,22.82,20.84)	(181.77,176.93,168.03)
3	(-0.75,0.03,-0.66)	(30.11,25.56,23.35)	(178.48,173.75,164.67)
4	(-0.74,0.14,-0.66)	(30.30,26.38,24.59)	(178.61,172.91,164.74)
5	(-0.70,0.26,-0.66)	(32.32,28.56,26.79)	(180.36,174.80,166.56)
6	(-0.74,-0.23,-0.63)	(29.30,25.30,22.33)	(171.82,166.62,159.21)
7	(-0.75,-0.25,-0.62)	(34.36,30.09,26.78)	(170.53,164.93,158.65)
8	(-0.74,-0.32,-0.58)	(32.09,27.67,24.22)	(176.57,171.15,165.16)

$\frac{\sigma_{|\boldsymbol{\alpha}|}}{\text{avg}\{|\boldsymbol{\alpha}|\}} \times 100 = 9.63\%$ $\frac{\sigma_{|\boldsymbol{\alpha}|}}{255\sqrt{3}} \times 100 = 1.01\%$ $\text{avg}\{\widehat{\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j}\} = 1.34^\circ$ $\max\{\widehat{\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j}\} = 3.00^\circ$

$\frac{\sigma_{|\boldsymbol{\beta}|}}{\text{avg}\{|\boldsymbol{\beta}|\}} \times 100 = 2.90\%$ $\frac{\sigma_{|\boldsymbol{\beta}|}}{255\sqrt{3}} \times 100 = 1.96\%$ $\text{avg}\{\widehat{\boldsymbol{\beta}_i, \boldsymbol{\beta}_j}\} = 0.22^\circ$ $\max\{\widehat{\boldsymbol{\beta}_i, \boldsymbol{\beta}_j}\} = 0.43^\circ$

Table 5.9. Results for robustness experiment 4-2 and OPAM: (a) TEAV, (b) TECA. In the tables, $\widehat{\boldsymbol{\alpha}}$ stands for the vector of estimates of $L_a^c A^c + \mathrm{E}_I[\mu_{dc}]A^c$ and $\widehat{\boldsymbol{\beta}}$ for the vector of estimates of $L_d^c A^c$, $c \in \{R, G, B\}$.

5.7 Conclusions

A total of four methods for estimating scene lighting parameters have been presented throughout this chapter. Two of them assume orthographic projection (OPAM) while the other two assume perspective projection (CPAM). Within every projection case, two algorithms for estimating the lighting orientation have been discussed and compared, TEAV and TECA, which differ in the way how the tilt of the illumination is determined.

The output of the estimators consists of the orientation of the illumination and the strengths of ambient and directional lighting. The availability of colour calibration images also allows determining the colour of both sources of light. As a limitation, the lighting strengths are estimated up to the factor including the F-number $(\pi/4)(d/f)^2$ and the exposure time T. If required, the particular value of the F-number, determined by the particular lens aperture, should be discounted from the output given by the algorithms. Accordingly, any change in the camera configuration would lead to the necessity of repeating the estimation procedure, even if just the camera gain is modified. In this sense, Automatic Gain Control (AGC) should be turned off if the lighting strengths are to be used. The lighting orientation is, however, unaffected by such a change in the camera configuration.

Despite the limitations of the lighting estimation methods proposed in this chapter as for the number of light sources, several advantages make them attractive in single-source cases against multi-source methods: (1) in general, its complexity is lower than the multi-source methods; (2) a previous geometric calibration of the camera is not required and the estimation of the surface normal vectors of the calibration sphere is embedded within the method itself; (3) all the image pixels belonging to the sphere projection are used in the estimation of the lighting parameters; (4) no threshold or parameter needs to be set up.

(a)

# image	$\widehat{s}$	$\widehat{\boldsymbol{\alpha}}$	$\widehat{\boldsymbol{\beta}}$
1	(-0.75,-0.13,-0.65)	(21.00,17.43,15.23)	(175.15,169.27,161.33)
2	(-0.75,0.02,-0.67)	(26.14,22.63,20.10)	(172.89,166.79,158.60)
3	(-0.76,0.09,-0.64)	(19.91,16.76,14.82)	(173.46,167.49,160.47)
4	(-0.72,0.23,-0.65)	(23.62,20.47,18.46)	(173.54,168.07,161.12)
5	(-0.74,-0.26,-0.62)	(26.03,22.22,19.40)	(175.48,169.61,161.99)
6	(-0.74,-0.30,-0.60)	(21.48,17.54,15.18)	(182.54,176.74,169.12)
7	(-0.74,-0.35,-0.57)	(24.22,19.93,17.11)	(185.64,180.22,173.28)
8	(-0.75,0.35,-0.56)	(21.26,18.07,15.83)	(168.51,162.50,155.69)

$\frac{\sigma_{|\boldsymbol{\alpha}|}}{\text{avg}\{|\boldsymbol{\alpha}|\}} \times 100 = 11.15\%$ $\frac{\sigma_{|\boldsymbol{\alpha}|}}{255\sqrt{3}} \times 100 = 0.87\%$ $\text{avg}\{\widehat{\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j}\} = 1.07^\circ$ $\max\{\widehat{\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j}\} = 2.36^\circ$

$\frac{\sigma_{|\boldsymbol{\beta}|}}{\text{avg}\{|\boldsymbol{\beta}|\}} \times 100 = 3.32\%$ $\frac{\sigma_{|\boldsymbol{\beta}|}}{255\sqrt{3}} \times 100 = 2.21\%$ $\text{avg}\{\widehat{\boldsymbol{\beta}_i, \boldsymbol{\beta}_j}\} = 0.16^\circ$ $\max\{\widehat{\boldsymbol{\beta}_i, \boldsymbol{\beta}_j}\} = 0.41^\circ$

(b)

# image	$\widehat{s}$	$\widehat{\boldsymbol{\alpha}}$	$\widehat{\boldsymbol{\beta}}$
1	(-0.74,-0.06,-0.67)	(18.17,14.83,12.79)	(177.61,171.45,163.29)
2	(-0.73,0.04,-0.68)	(23.11,19.74,17.50)	(175.75,169.50,160.70)
3	(-0.73,0.18,-0.66)	(15.83,12.86,11.27)	(177.06,170.87,163.16)
4	(-0.71,0.24,-0.66)	(22.94,19.85,17.88)	(173.93,168.36,161.37)
5	(-0.75,-0.19,-0.63)	(23.05,19.34,16.77)	(178.12,172.28,164.37)
6	(-0.74,-0.26,-0.61)	(19.95,16.10,13.84)	(183.17,177.30,169.58)
7	(-0.75,-0.31,-0.58)	(22.19,18.07,15.33)	(186.54,180.95,173.98)
8	(-0.72,0.40,-0.57)	(19.83,16.61,14.41)	(169.41,163.29,156.10)

$\frac{\sigma_{|\boldsymbol{\alpha}|}}{\text{avg}\{|\boldsymbol{\alpha}|\}} \times 100 = 14.07\%$ $\frac{\sigma_{|\boldsymbol{\alpha}|}}{255\sqrt{3}} \times 100 = 0.98\%$ $\text{avg}\{\widehat{\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j}\} = 1.22^\circ$ $\max\{\widehat{\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j}\} = 2.77^\circ$

$\frac{\sigma_{|\boldsymbol{\beta}|}}{\text{avg}\{|\boldsymbol{\beta}|\}} \times 100 = 3.14\%$ $\frac{\sigma_{|\boldsymbol{\beta}|}}{255\sqrt{3}} \times 100 = 2.11\%$ $\text{avg}\{\widehat{\boldsymbol{\beta}_i, \boldsymbol{\beta}_j}\} = 0.18^\circ$ $\max\{\widehat{\boldsymbol{\beta}_i, \boldsymbol{\beta}_j}\} = 0.47^\circ$

Table 5.10. Results for robustness experiment 4-2 and CPAM: (a) TEAV, (b) TECA. In the tables, $\widehat{\boldsymbol{\alpha}}$ stands for the vector of estimates of $L_a^c A^c + \mathrm{E}_I[\mu_{dc}]A^c$ and $\widehat{\boldsymbol{\beta}}$ for the vector of estimates of $L_d^c A^c$, $c \in \{R, G, B\}$.

In all cases, the corresponding estimation method has been developed in connection with the camera noise model presented in section 2.4.4, which has allowed identifying the noise sources corrupting digital pixel values in order to counteract their effects throughout the estimation process; furthermore, such a formulation has revealed that any estimation of the strength of ambient illumination always includes dark current, so that it must be estimated apart to get the real strength of ambient illumination, if needed.

An extensive battery of tests has been carried out in order to characterize as accurately as possible the scope of application of the lighting parameters estimation methods presented. The following conclusions can be drawn from the corresponding results:

- The different tests performed with synthetic images have shown OPAM and CPAM quite accurate when processing images generated under the corresponding projection. This accuracy has maintained at the same level whatever the tilt and slant of directional lighting has been.
- OPAM has resulted to be quite useful for estimating the lighting orientation even under perspective projection, as long as the sphere center is not too separated from the optical axis. Regarding the estimation of the lighting strengths, OPAM estimations tended to be biased when processing images generated under perspective projection. The bias seemed to be constant across lighting orientations and changes in the XY position of the sphere center, while the bias tended to decrease as the sphere was moved farther from the camera. CPAM produced errors below one tenth of degree in all cases.
- The initial contour of the projection sphere has not resulted determinant in the measurement of the lighting orientation when analyzing real calibration images. As for the lighting strengths, some oscillations less than 4 intensity levels in amplitude have been detected.

Table 5.11. Comparison results for synthetic spheres under orthographic projection. True and estimated lighting directions are given through the tilt τ and slant σ angles. L&R1: Original L&R method without ambient reflection term; L&R2: Original L&R method with ambient reflection term; L&R3: Modified L&R method without ambient reflection term; L&R4: Modified L&R method with ambient reflection term; Z&C1: Z&C method without ambient reflection term; Z&C2: Z&C method with ambient reflection term. Δ means the angle between true and estimated illumination directions. TEAV and TECA methods are used within OPAM. All the angles are given in degrees.

true		L&R1			L&R2			L&R3			L&R4		
τ	σ	$\widehat{\tau}$	$\widehat{\sigma}$	Δ	$\widehat{\tau}$	$\widehat{\sigma}$	Δ	$\widehat{\tau}$	$\widehat{\sigma}$	Δ	$\widehat{\tau}$	$\widehat{\sigma}$	Δ
0.00	112.50	0.00	113.50	1.00	0.00	135.50	23.00	0.00	112.50	0.00	0.00	126.50	14.00
30.00	112.50	30.23	113.50	1.02	30.23	135.50	23.00	30.23	112.50	0.21	30.23	126.50	14.00
60.00	112.50	59.77	113.50	1.02	59.77	135.50	23.00	59.77	112.50	0.21	59.77	126.50	14.00
90.00	112.50	90.00	113.50	1.00	90.00	135.50	23.00	90.00	112.50	0.00	90.00	126.50	14.00
0.00	135.00	0.00	135.50	0.50	0.00	152.50	17.50	0.00	135.50	0.50	0.00	144.50	9.50
30.00	135.00	30.38	135.50	0.57	30.38	152.50	17.50	30.38	135.50	0.57	30.38	144.50	9.50
60.00	135.00	59.62	135.50	0.57	59.62	152.50	17.50	59.62	135.50	0.57	59.62	144.50	9.50
90.00	135.00	90.00	135.50	0.50	90.00	152.50	17.50	90.00	135.50	0.50	90.00	144.50	9.50
0.00	157.50	0.00	158.50	1.00	0.00	180.00	22.50	0.00	157.50	0.00	0.00	166.50	9.00
30.00	157.50	29.87	158.50	1.00	0.00	180.00	22.50	29.87	157.50	0.05	29.87	166.50	9.00
60.00	157.50	60.13	158.50	1.00	0.00	180.00	22.50	60.13	157.50	0.05	60.13	166.50	9.00
90.00	157.50	90.00	158.50	1.00	0.00	180.00	22.50	90.00	157.50	0.00	90.00	166.50	9.00

true		Z&C1			Z&C2			TEAV			TECA		
τ	σ	$\widehat{\tau}$	$\widehat{\sigma}$	Δ	$\widehat{\tau}$	$\widehat{\sigma}$	Δ	$\widehat{\tau}$	$\widehat{\sigma}$	Δ	$\widehat{\tau}$	$\widehat{\sigma}$	Δ
0.00	112.50	0.00	105.50	7.00	0.00	114.50	2.00	0.00	112.47	0.03	0.00	112.58	0.08
30.00	112.50	30.11	105.50	7.00	30.11	114.50	2.00	30.00	112.43	0.07	30.01	112.47	0.03
60.00	112.50	59.89	105.50	7.00	59.89	114.50	2.00	60.00	112.43	0.07	59.99	112.47	0.03
90.00	112.50	90.00	105.50	7.00	90.00	114.50	2.00	90.00	112.47	0.03	90.00	112.58	0.08
0.00	135.00	0.00	120.50	14.50	0.00	128.50	6.50	0.00	135.04	0.04	0.00	135.00	0.00
30.00	135.00	30.00	120.50	14.50	30.00	128.50	6.50	30.02	134.98	0.02	30.02	135.07	0.07
60.00	135.00	60.00	120.50	14.50	60.00	128.50	6.50	59.98	134.98	0.02	59.98	135.07	0.07
90.00	135.00	90.00	120.50	14.50	90.00	128.50	6.50	90.00	135.04	0.04	90.00	135.00	0.00
0.00	157.50	0.00	142.50	15.00	0.00	149.50	8.00	0.00	157.49	0.01	0.00	157.49	0.01
30.00	157.50	30.15	142.50	15.00	30.15	149.50	8.00	30.05	157.60	0.10	30.00	157.57	0.07
60.00	157.50	59.85	142.50	15.00	59.85	149.50	8.00	59.95	157.60	0.10	60.00	157.57	0.07
90.00	157.50	90.00	142.50	15.00	90.00	149.50	8.00	90.00	157.49	0.01	90.00	157.49	0.01

- Both OPAM and CPAM have shown a significant robustness against zero-mean Gaussian noise of up to 15 intensity levels of standard deviation. Clearly TEAV has proved to be more robust than TECA for both OPAM and CPAM, specially as for the estimation of lighting strengths.
- Among all the variants, CPAM-TECA has exhibited the greatest constancy as for the determination of the lighting orientation against changes in the illumination strengths, with a difference of less than half a degree on average among estimations. In the remaining cases, the average difference did not exceed 1.5 degrees.
- The robustness in the estimation of illumination strengths against changes in the lighting direction has been determined to be high: on the one hand, no changes were appreciated over three real images rotated four times 90 degrees, either for OPAM and for CPAM; on the other hand, rotations and displacements of the camera led to deviations below 3 degrees among the estimates of the colour of the ambient illumination, and below half a degree for the directional illumination, while the variations in the norms of the strengths were about 1-2 % with regard to the largest possible norm.
- Although there are simpler methods suitable for the estimation of the lighting orientation from a calibration sphere (L&R [141] and Z&C [311]), the methods presented here clearly outperform them, producing errors around a tenth of degree in synthetic images, while the others yield impor-

Table 5.12. Comparison results for synthetic spheres under perspective projection. True and estimated lighting directions are given through the tilt τ and slant σ angles. L&R1: Original L&R method without ambient reflection term; L&R2: Original L&R method with ambient reflection term; L&R3: Modified L&R method without ambient reflection term; L&R4: Modified L&R method with ambient reflection term; Z&C1: Z&C method without ambient reflection term; Z&C2: Z&C method with ambient reflection term. Δ means the angle between true and estimated illumination directions. TEAV and TECA methods are used within CPAM. All the angles are given in degrees.

true		L&R1			L&R2			L&R3			L&R4		
τ	σ	$\widehat{\tau}$	$\widehat{\sigma}$	Δ	$\widehat{\tau}$	$\widehat{\sigma}$	Δ	$\widehat{\tau}$	$\widehat{\sigma}$	Δ	$\widehat{\tau}$	$\widehat{\sigma}$	Δ
0.00	112.50	0.00	112.50	0.00	0.00	134.50	22.00	0.00	112.50	0.00	0.00	126.50	14.00
30.00	112.50	29.43	112.50	0.52	29.43	134.50	22.00	30.23	112.50	0.21	30.23	126.50	14.00
60.00	112.50	60.57	112.50	0.52	60.57	134.50	22.00	59.77	112.50	0.21	59.77	126.50	14.00
90.00	112.50	90.00	112.50	0.00	90.00	134.50	22.00	90.00	112.50	0.00	90.00	126.50	14.00
0.00	135.00	0.00	135.50	0.50	0.00	152.50	17.50	0.00	135.50	0.50	0.00	144.50	9.50
30.00	135.00	29.30	135.50	0.70	29.30	152.50	17.50	30.38	135.50	0.57	30.38	144.50	9.50
60.00	135.00	60.70	135.50	0.70	60.70	152.50	17.50	59.62	135.50	0.57	59.62	144.50	9.50
90.00	135.00	90.00	135.50	0.50	90.00	152.50	17.50	90.00	135.50	0.50	90.00	144.50	9.50
0.00	157.50	0.00	159.50	2.00	0.00	180.00	22.50	0.00	157.50	0.00	0.00	166.50	9.00
30.00	157.50	30.32	159.50	2.00	0.00	180.00	22.50	29.87	157.50	0.05	29.87	166.50	9.00
60.00	157.50	59.68	159.50	2.00	0.00	180.00	22.50	60.13	157.50	0.05	60.13	166.50	9.00
90.00	157.50	90.00	159.50	2.00	0.00	180.00	22.50	90.00	157.50	0.00	90.00	166.50	9.00

true		Z&C1			Z&C2			TEAV			TECA		
τ	σ	$\widehat{\tau}$	$\widehat{\sigma}$	Δ	$\widehat{\tau}$	$\widehat{\sigma}$	Δ	$\widehat{\tau}$	$\widehat{\sigma}$	Δ	$\widehat{\tau}$	$\widehat{\sigma}$	Δ
0.00	112.50	0.00	105.50	7.00	0.00	114.50	2.00	0.00	112.35	0.15	0.00	112.55	0.05
30.00	112.50	30.08	105.50	7.00	30.08	114.50	2.00	30.00	112.39	0.11	29.99	112.44	0.06
60.00	112.50	59.92	105.50	7.00	59.92	114.50	2.00	60.00	112.39	0.11	60.01	112.44	0.06
90.00	112.50	90.00	105.50	7.00	90.00	114.50	2.00	90.00	112.35	0.15	90.00	112.55	0.05
0.00	135.00	0.00	120.50	14.50	0.00	128.50	6.50	0.00	135.01	0.01	0.00	135.04	0.04
30.00	135.00	29.97	120.50	14.50	29.97	128.50	6.50	30.04	135.05	0.06	29.96	135.06	0.06
60.00	135.00	60.03	120.50	14.50	60.03	128.50	6.50	59.96	135.05	0.06	60.04	135.06	0.06
90.00	135.00	90.00	120.50	14.50	90.00	128.50	6.50	90.00	135.01	0.01	90.00	135.04	0.04
0.00	157.50	0.00	142.50	15.00	0.00	149.50	8.00	0.00	157.43	0.07	0.00	157.45	0.05
30.00	157.50	30.22	142.50	15.00	30.22	149.50	8.00	30.17	157.40	0.12	29.98	157.41	0.09
60.00	157.50	59.78	142.50	15.00	59.78	149.50	8.00	59.83	157.40	0.12	60.02	157.41	0.09
90.00	157.50	90.00	142.50	15.00	90.00	149.50	8.00	90.00	157.43	0.07	90.00	157.45	0.05

tant errors, between 2 and 20 degrees in the different cases tested, mostly because of the presence of ambient lighting. The perspective projection has not led to important variations in the estimations of those alternative methods, however.

6

Radiometric Calibration of CCD Cameras

This chapter proposes an algorithm called R^2CIU (*Robust Radiometric Calibration and Intensity Uncertainty estimation*) for characterizing the noise sources affecting CCD performance. In other words, the purpose is to calibrate CCD cameras at the radiometric level. As will be seen, those techniques are intended to determine the distribution parameters of the noise sources discussed in section 2.4.4 and summarized in equation 2.44, namely the photo-response non-uniformity (PRNU), the dark current non-uniformity (DCNU) and the noise independent of the charge stored at collection sites.

The radiometric calibration of CCD sensors derive in two important benefits: on the one hand, it allows accounting for the corresponding noise within image processing algorithms by means of uncertainties associated to every digital intensity level; on the other hand, synthetic images can be generated according to the estimated noise model parameters, being the main interest of these images that they allow testing machine vision algorithms under more realistic noisy conditions, rather than just using standard models —Gaussian, salt and pepper, etc.— not taking into account the performance of the sensor producing the irradiance measurement. Besides, the determination of the distribution parameters of the DCNU will allow removing the term $\mathrm{E}_I[\mu_{dc}]A^c$ from α^c when estimating the lighting strengths after fitting points $(m_b, \mathrm{E}_{m_b}[\mu_{D^c}])$ by a straight line, as explained in section 5.1.

The rest of the chapter is organized as follows: section 6.1 presents a set of techniques for estimating the distribution parameters of the noise sources affecting CCD performance; section 6.2 discusses the two above-mentioned applications of the radiometric calibration of vision cameras; previous work is reviewed in section 6.3, while experimental results are provided in section 6.4; finally, conclusions appear in section 6.5.

6.1 Estimation of the Camera Noise Model Parameters

This section will present a set of techniques for estimating the parameters of the camera noise model expressed in equation 2.44. All of them are based on the use of constant reflectance matte calibration cards uniformly illuminated and covering the whole field of view of the camera.

Under those circumstances, equation 5.1 results again, which is reproduced next for convenience after a slight reordering of terms and the expansion of the noise term N^c:

$$\begin{aligned} D^c(i,j) &= K(i,j)\left[(L_a\rho_b)^c + m_b(i,j)(L_d\rho_b)^c\right]A^c + \mu_{dc}(i,j)A^c \\ &+ \overbrace{\underbrace{\left(N_S^c(i,j) + N_{dc}(i,j)\right)A^c}_{N_e^c(i,j)} + \underbrace{N_R A^c + N_Q}_{N_f^c}}^{N^c(i,j)} . \end{aligned} \tag{6.1}$$

In this last equation, if the the card is not bent anyway, $m_b(i,j)$ does not change across the image. Moreover, the light source uniformity assumption makes the whole expression $(L_a\rho_b)^c + m_b(L_d\rho_b)^c$ be constant.

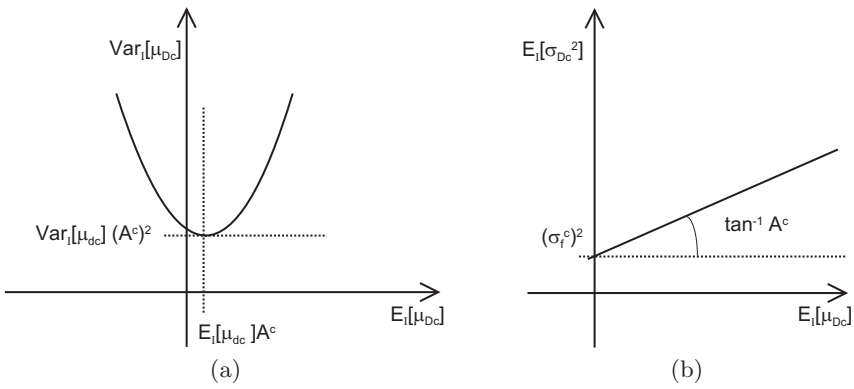


Fig. 6.1. Geometric relationships among camera noise model parameters: (a) equations 6.5 and (b) equation 6.8.

6.1.1 Estimation of the PRNU and DCNU distribution parameters

Let us assume that several images of the calibration card with the same camera and lighting parameters are taken and averaged pixel by pixel to produce image μ_{D^c}. As discussed in section 4.4.1, the non-spatial noise N^c tends to vanish in the average pixel values $\mu_{D^c}(i,j)$. If, in turn, the average across the whole image of the $\mu_{D^c}(i,j)$ values is taken, then equation 6.2 results:

$$\mathrm{E}_I[\mu_{D^c}] = \mathrm{E}_I[K]\left[(L_a\rho_b)^c + m_b(L_d\rho_b)^c\right]A^c + \mathrm{E}_I[\mu_{dc}]A^c\,, \tag{6.2}$$

which, from the fact that $\mathrm{E}_I[K]$ is typically 1, simplifies to equation 6.3:

$$\mathrm{E}_I[\mu_{D^c}] = \left[(L_a\rho_b)^c + m_b(L_d\rho_b)^c\right]A^c + \mathrm{E}_I[\mu_{dc}]A^c\,. \tag{6.3}$$

Next, given the independence between the PRNU and the DCNU and after the averaging of the $D^c(i,j)$ values, which reduces the influence of non-spatial noise in image pixels, the (spatial) variance of the $\mu_{D^c}(i,j)$ values is given by equation 6.4:

$$\mathrm{Var}_I\left[\mu_{D^c}\right] = \mathrm{Var}_I[K]\left[(L_a\rho_b)^c + m_b(L_d\rho_b)^c\right]^2\left(A^c\right)^2 + \mathrm{Var}_I[\mu_{dc}]\left(A^c\right)^2\,, \tag{6.4}$$

which, using equation 6.3, transforms into equation 6.5a:

$$\mathrm{Var}_I\left[\mu_{D^c}\right] = \mathrm{Var}_I[K]\left(\mathrm{E}_I[\mu_{D^c}] - \mathrm{E}_I[\mu_{dc}]A^c\right)^2 + \mathrm{Var}_I[\mu_{dc}]\left(A^c\right)^2\,, \tag{6.5a}$$

or, expanded:

$$\begin{aligned}\mathrm{Var}_I\left[\mu_{D^c}\right] = \mathrm{Var}_I[K]\left(\mathrm{E}_I[\mu_{D^c}]\right)^2 - 2\mathrm{Var}_I[K]\left(\mathrm{E}_I[\mu_{dc}]A^c\right)\mathrm{E}_I[\mu_{D^c}]\\ +\mathrm{Var}_I[K]\left(\mathrm{E}_I[\mu_{dc}]A^c\right)^2 + \mathrm{Var}_I[\mu_{dc}]\left(A^c\right)^2\,.\end{aligned} \tag{6.5b}$$

According to equations 6.5, points $(\mathrm{E}_I[\mu_{D^c}], \mathrm{Var}_I\left[\mu_{D^c}\right])$ lie in a parabola like the one shown in figure 6.1(a), from whose parameters the (spatial) variance of K and μ_{dc}, together with $\mathrm{E}_I[\mu_{dc}]$, can be estimated. Such points can be obtained from several sets of images of the calibration card taken at different values of $I^c = (L_a\rho_b)^c + m_b(L_d\rho_b)^c$. These images can be easily generated changing, from set to set, either $(L_a\rho_b)^c$, m_b, or $(L_d\rho_b)^c$, alone or in combination with one another. The m_b term can be altered just tilting the plane card in front of the camera, while the lighting-related terms change if neutral-density (ND) filters are placed in front of the camera or its optics aperture (i.e. the F-number) is varied, or, finally, if the reflectance of the card is varied uniformly. Both ND filters and modifications of the iris give rise to changes in the irradiance incident at all the CCD cells for all wavelengths, in a more uniform way in a spatial sense when ND filters are used. This affects the net values of the $(L_a\rho_b)^c$ and $(L_d\rho_b)^c$ terms equally. Changing the shutter speed (i.e. the integration time T of equation 2.40)

Pseudocode 6.1 Determination of the PRNU and DCNU distribution parameters.

(1) Take a number of sets of images of the calibration card so that, from set to set, $(L_a\rho_b)^c + m_b(L_d\rho_b)^c$ changes
(2) Compute the average image μ_{D^c} for every set
(3) Calculate $\mathrm{E}_I[\mu_{D^c}]$ and $\mathrm{Var}_I[\mu_{D^c}]$ for every set of images
(4) Fit the different pairs $(x, y) = (\mathrm{E}_I[\mu_{D^c}], \mathrm{Var}_I[\mu_{D^c}])$ by a parabola $y = ax^2 - 2bx + c$
(5) Estimate $\mathrm{Var}_I[K]$, $\mathrm{E}_I[\mu_{dc}]A^c$ and $\mathrm{Var}_I[\mu_{dc}](A^c)^2$ as follows:
$\mathrm{Var}_I[K] = a$, $\mathrm{E}_I[\mu_{dc}]A^c = \frac{b}{a}$ and $\mathrm{Var}_I[\mu_{dc}](A^c)^2 = c - \frac{b^2}{a}$

or the camera gain also gives rise to different pairs $(\mathrm{Var}_I\left[\mu_{D^c}\right], \mathrm{E}_I[\mu_{D^c}])$, although at the expense of modifying the behaviour of dark current, and, consequently, the values of $\mathrm{E}_I[\mu_{dc}]A^c$ and $\mathrm{Var}_I[\mu_{dc}](A^c)^2$. The resulting pairs would not necessarily satisfy, thus, the aforementioned parabolic relationship.

Pseudocode 6.1 summarizes the whole estimation procedure.

6.1.2 Estimation of camera gain and charge-independent non-spatial noise

Further manipulation of equation 6.1 allows estimating the camera gain A^c and the variance of the noise independent of the number of electrons stored at collection sites N_f^c. To perform this calculation, the (temporal) variance of the intensity level in a certain image cell between exposures, $\sigma^2_{D^c}(i, j)$, will be determined (i.e. several images of the calibration card under exactly the same imaging conditions are taken and the variance image $\sigma^2_{D^c}$ is considered). In such a case, the variation comes from the shot noises N_S^c and N_{dc}, and from N_f^c. On the basis of the independence between the different noise sources, $\sigma^2_{D^c}(i, j)$ is given by:

$$\sigma^2_{D^c}(i,j) = \left[K(i,j)\left((L_a\rho_b)^c + m_b(i,j)(L_d\rho_b)^c\right) + \mu_{dc}(i,j)\right](A^c)^2 + \left(\sigma_f^c\right)^2 , \tag{6.6}$$

where $(\sigma_f^c)^2$ is the variance of N_f^c. Now, given that $m_b(i, j)$ is constant throughout the image, the following equation results taking the (spatial) mean of the $\sigma^2_{D^c}(i, j)$ values:

$$\mathrm{E}_I[\sigma^2_{D^c}] = \left((L_a\rho_b)^c + m_b(L_d\rho_b)^c + \mathrm{E}_I[\mu_{dc}]\right)(A^c)^2 + \left(\sigma_f^c\right)^2 , \tag{6.7}$$

which, using equation 6.3, becomes:

$$\mathrm{E}_I[\sigma^2_{D^c}] = A^c\mathrm{E}_I[\mu_{D^c}] + \left(\sigma_f^c\right)^2 . \tag{6.8}$$

Therefore, according to equation 6.8, pairs $(\mathrm{E}_I[\mu_{D^c}], \mathrm{E}_I[\sigma^2_{D^c}])$ lie in a straight line whose slope and intercept with the $\mathrm{E}_I[\sigma^2_{D^c}]$ axis coincide with, respectively, A^c and $(\sigma_f^c)^2$ (see figure 6.1(b)). These pairs can be obtained taking several sets of images so that $I^c = (L_a\rho_b)^c + m_b(L_d\rho_b)^c$ changes from set to set, as previously discussed, and taking the (spatial) average of the pixel-by-pixel (temporal) mean and variance of the images within every set. Pseudocode 6.2 summarizes the estimation procedure.

Pseudocode 6.2 Determination of camera gain and non-spatial noise.

(1) Take a number of sets of images of the calibration card so that, from set to set, $(L_a\rho_b)^c + m_b(L_d\rho_b)^c$ changes
(2) Compute the average μ_{D^c} and variance $\sigma^2_{D^c}$ images for every set
(3) Calculate $\mathrm{E}_I[\mu_{D^c}]$ and $\mathrm{E}_I[\sigma^2_{D^c}]$ for every set of images
(4) Fit the different pairs $(x, y) = (\mathrm{E}_I[\mu_{D^c}], \mathrm{E}_I[\sigma^2_{D^c}])$ by a straight line $y = ax + b$
(5) Estimate A^c and $(\sigma_f^c)^2$ from, respectively, a and b

Pseudocode 6.3 Procedure for generating synthetic images with realistic noise. ($\mathcal{N}(x, y)$ represents a random number taken from a normal population of mean x and standard deviation y, while $\mathcal{P}(x)$ corresponds to a random number taken from a Poisson distribution of parameter x.)

for every image cell (i, j)
(1) $K(i, j) := \mathcal{N}(1, \sqrt{\mathrm{Var}_I[K]})$
(2) $\mu_{dc}(i, j) := \mathcal{N}(\mathrm{E}_I[\mu_{dc}], \sqrt{\mathrm{Var}_I[\mu_{dc}]})$
(3) **for** every colour channel c
(3.1) $[I^c(i, j)A^c] := [(L_a\rho_b(i, j))^c A^c] + m_b(i, j)\,[(L_d\rho_b(i, j))^c A^c] + m_i(i, j)\,[(L_d\rho_i(i, j))^c A^c]$
(3.2) $N_f^c := \mathcal{N}(0, \sigma_f^c)$
(3.3) $D^c(i, j) := \mathcal{P}\left(K(i, j)\frac{[I^c(i,j)A^c]}{A^c}\right) A^c + \mathcal{P}\left(\mu_{dc}(i, j)\right) A^c + N_f^c$
endfor
endfor

6.2 Applications of Radiometric Camera Calibration

The benefit of a camera model such as the one of section 2.4.4, together with the values of the corresponding parameters, estimated by means of the procedures proposed in section 6.1, is twofold:

- On the one hand, noise founded on the particular characteristics of the imaging sensor can be incorporated to synthetic images in order to test the robustness of computer vision algorithms, instead of adding random noise without any knowledge of its representativeness with regard to the real imaging conditions imposed by the sensor itself.
- On the other hand, the uncertainties in the measurements of image irradiance can be determined and therefore their knowledge introduced into a computer vision algorithm, for instance, to compute adaptive thresholds. Stokman suggested this idea in [262], although his camera noise model did not include PRNU nor charge-independent noises, nor distinguished between DCNU and dark current shot noise; moreover, although his method is based on the propagation of uncertainties along colour space transformations, the computation of the uncertainties is based on the unknown original irradiance value, not on the final digital value returned by the camera. The method presented in this section yields a table relating irradiance measurements directly with their uncertainty.

Both issues are developed in the following sections.

6.2.1 Incorporation of realistic noise into synthetic images

Pseudocode 6.3 proposes a method to obtain noisy images on the basis of the camera model introduced in section 2.4.4. To this end, this procedure makes use of estimates of $\mathrm{Var}_I[K]$, $\mathrm{E}_I[\mu_{dc}]$, $\mathrm{Var}_I[\mu_{dc}]$, A^c and σ_f^c and approximates the distributions of the PRNU, the DCNU and the charge-independent noise by Gaussian distributions.

Referring to pseudocode 6.3, steps (1) and (2) determine the behaviour of current cell (i, j) as for photo-response and dark current average. The next step, (3), computes the remaining terms, dependent on the colour channel considered. In this way, step (3.1) calculates the ideal intensity value corresponding to colour channel c and image cell (i, j) in the synthetic scene, $I^c(i, j)A^c$, on the basis of the reflectance and illumination values for cell (i, j) (observe this is a simplification of equation 4.1). Notice that the ambient, body and interface composite reflectances are assumed to include the corresponding camera gain, so that they are already values between 0 and 255; therefore, their sum, weighted by the geometrical terms m_b and m_i, directly gives rise to $I^c(i, j)A^c$. Next, step (3.2) gives value to N_f^c, the noise independent of the number of electrons stored in the collection site, by means of a zero-mean normal distribution with standard deviation σ_f^c. At last, in step (3.3), the final intensity value is generated through the sum of the non-charge-dependent noise N_f^c and two Poisson-distributed random values parameterized by the corresponding expected number of electrons at the collection site, $K(i, j)I^c(i, j)$ and $\mu_{dc}(i, j)$, to incorporate the uncertainty in the number of collected electrons.

Table 6.1. Contribution of the different noise sources to the uncertainty in digital intensity levels.

noise source	associated variance
PRNU	$\mathrm{Var}_I[K]\,(I^cA^c)^2$
DCNU	$\mathrm{Var}_I[\mu_{dc}](A^c)^2$
shot noise at image cell (i,j)	$K(i,j)I^c(A^c)^2$
dark current shot noise at image cell (i,j)	$\mu_{dc}(i,j)(A^c)^2$
readout and quantization noises	$\left(\sigma_f^c\right)^2$

Therefore, according to equation 2.44, the sum of the shot noise and dark current shot noise would correspond to:

$$N_e^c(i,j) = \mathcal{P}\left(K(i,j)\frac{[I^c(i,j)A^c]}{A^c}\right)A^c + \mathcal{P}\left(\mu_{dc}(i,j)\right)A^c - \left(K(i,j)[I^c(i,j)A^c] + \mu_{dc}(i,j)A^c\right) . \quad (6.9)$$

Taking into account that a Poisson distribution $\mathcal{P}(\mu)$ can be well approximated by a normal distribution $\mathcal{N}(\mu, \sqrt{\mu})$ when μ is large [272] and that the expected number of collected electrons μ will seldom be low, step (4) of pseudocode 6.3 can be substituted by $D^c(i,j) := \mathcal{N}\left(\mu_1, \sqrt{\mu_1}\right)A^c + \mathcal{N}\left(\mu_2, \sqrt{\mu_2}\right)A^c + N_f^c$, with $\mu_1 = K(i,j)\frac{[I^c(i,j)A^c]}{A^c}$ and $\mu_2 = \mu_{dc}(i,j)$ in order to ease the generation of the synthetic image. For instance, if $K(i,j) = 1$, $A^c = 0.01$ and $I^c(i,j)A^c = 1$ (i.e. very low intensity), μ_1 results to be 100, value for which the approximation is quite accurate.

6.2.2 Determination of uncertainties in the measurement of image irradiance

Generalizing equation 6.6, equation 6.10 expresses the uncertainty in the irradiance measurement for image cell (i,j):

$$\sigma_{D^c}^2(i,j) = \left[K(i,j)I^c(i,j) + \mu_{dc}(i,j)\right](A^c)^2 + \left(\sigma_f^c\right)^2 . \quad (6.10)$$

That is to say, given the precise value of K and μ_{dc} for cell (i,j), it is possible to determine how uncertain is D^c for a certain value $I^c(i,j) = (L_a\rho_a)^c + m_b(i,j)(L_d\rho_b)^c + m_i(i,j)(L_d\rho_i)^c$.

In case it is planned to use this information in computer vision algorithms, for a, let us say, typical 640×480 RGB camera providing 8-bit digital intensity levels, a data structure consisting of $640 \times 480 \times 256 \times 3$ uncertainty values would result. In case every uncertainty is represented as a 4-byte floating point value, the whole structure requires a total of 943,718.400 bytes, or 900 Mbytes, to be stored. Certainly, it is a huge data structure, probably overdetailed for most vision applications. Instead, in this section, it is proposed to determine the uncertainty corresponding to every digital intensity level I^cA^c irrespectively of the particular image cell (i,j) to which the irradiance measurement is related to. This strategy implies, however, to develop a expression which also accounts for the PRNU and the DCNU of the particular camera.[1]

To this end, it is assumed that a plane card of constant reflectance and uniformly illuminated is imaged by the camera. The goal is to quantify the variation which could be found in the resulting image with respect to the noiseless digital intensity value I^cA^c which would result in a perfect camera. Clearly, every noise source contributes to the final uncertainty value. These contributions have already been expressed somewhere within section 6.1. Table 6.1 puts them all together. As can be observed, shot noise and dark current shot noise are given for a particular image cell (i,j). This dependency can

[1] Obviously, it is also possible to calculate the uncertainty online during the execution of the vision algorithm or to precalculate an uncertainty image for every image to process. In the first case, only values $K(i,j)$ and $\mu_{dc}(i,j)$ should be stored, requiring $2 \times 640 \times 480 \times 4 = 2,457.600$ bytes if real values are represented using 4 bytes. In the second case, memory should also be allocated to the uncertainty image and, consequently, $3 \times 640 \times 480 \times 4 = 3,686.400$ bytes of storage would additionally be needed. Both solutions would, however, require a larger time of processing as well as more than the $3 \times 256 \times 4 = 3.072$ bytes needed by the solution proposed here.

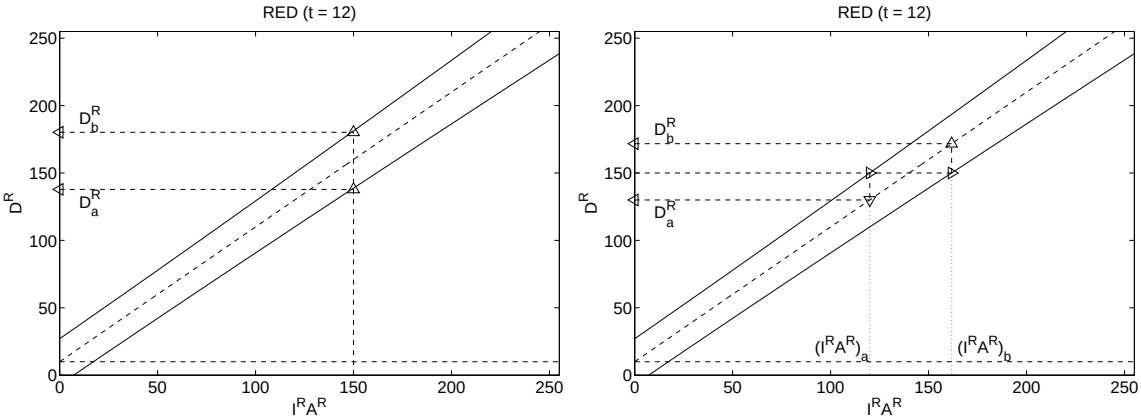


Fig. 6.2. [left] Uncertainty curves which derive from equation 6.11 after a proper calibration of the camera and a certain value of t, for the red colour channel. [right] Confidence interval $[D^R_a, D^R_b]$ corresponding to a certain measurement D^R.

be removed at the expense of the approximations $K(i,j) \approx 1$ and $\mu_{dc}(i,j) \approx \mathrm{E}_I[\mu_{dc}]$, also made in [93]. That is to say, $\mathrm{Var}_I[K]$ and $\mathrm{Var}_I[\mu_{dc}]$ are considered small so that the variation of the variances of both shot noise sources can be assumed approximately constant across image cells for a fixed I^c.

From the independence of the noise sources, the total uncertainty is given by the sum in quadrature of the different contributions [93]. In this way, equations 6.11 provide, in the form of *expected value* $\pm$ *uncertainty* [272], a confidence interval for the final digital value returned by the camera, D^c, for every possible digital intensity level, I^cA^c.

$$D^c = (I^cA^c + \mathrm{E}_I[\mu_{dc}]A^c) \pm t\sigma^c_I \tag{6.11a}$$

$$\sigma^c_I = \sqrt{\mathrm{Var}_I[K]\,(I^cA^c)^2 + \mathrm{Var}_I[\mu_{dc}]\,(A^c)^2 + (I^cA^c)\,A^c + \mathrm{E}_I[\mu_{dc}]\,(A^c)^2 + \left(\sigma^c_f\right)^2} \tag{6.11b}$$

For a given t, equations 6.11 relate every possible intensity value I^cA^c with an interval of values $[D^c_a, D^c_b]$ where the corresponding digital value returned by the camera can lie with a certain probability. Resorting to the Chebyshev inequality[2] [151], $t = 2, 3$ and 4 represent that the measurement D^c will lie inside $[D^c_a, D^c_b]$ with, respectively, probabilities of 75.00%, 88.89% and 93.75%. If, besides, the Poisson distributions associated to the shot noises are approximated by Gaussian distributions, and the remaining noise sources are assumed Gaussian, the noise distribution results to be normal, and, therefore, the previous probabilities for $t = 2, 3$ and 4 increase, respectively, up to 95.45%, 99.73% and 99.99%. By way of example, figure 6.2(left) shows uncertainty curves (solid lines) for the red colour channel of a hypothetical camera with the following set of radiometric parameters: $A^R = 0.002$, $(\sigma^R_f)^2 = 1.0$, $\mathrm{E}_I[\mu_{dc}]A^R = 10$, $\sqrt{\mathrm{Var}_I[\mu_{dc}]}A^R = 1.0$ and $\sqrt{\mathrm{Var}_I[K]} = 0.006$. In the figure, the lower curve corresponds to the '-' sign of equation 6.11a and therefore gives D^R_a for every I^RA^R, while the upper curve is for the '+' sign and D^R_b for every I^RA^R. In the example, an exaggerated value, 12, for t is used just for illustration purposes. Besides, the dashed horizontal straight line at the bottom corresponds to $\mathrm{E}_I[\mu_{dc}]A^R$, while the slanted dashed line between the two uncertainty curves is $D^R = I^RA^R + \mathrm{E}_I[\mu_{dc}]A^R$.

Note that, however, equations 6.11 go from the noiseless intensity values to the measurements produced by the camera, although only the latter are effectively available. Therefore, the relationship must be reversed to be useful: i.e. given a measure D^c, an interval of possible noiseless intensities, $[(I^cA^c)_a, (I^cA^c)_b]$, must be looked for (see figure 6.2(right)). $(I^cA^c)_a$ and $(I^cA^c)_b$ can be easily obtained reversing equation 6.11a to get equation 6.12, which involves a second degree polynomial in I^cA^c:

[2] For any constant $k > 0$, if a random variable X has mean μ and standard deviation σ, then $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$.

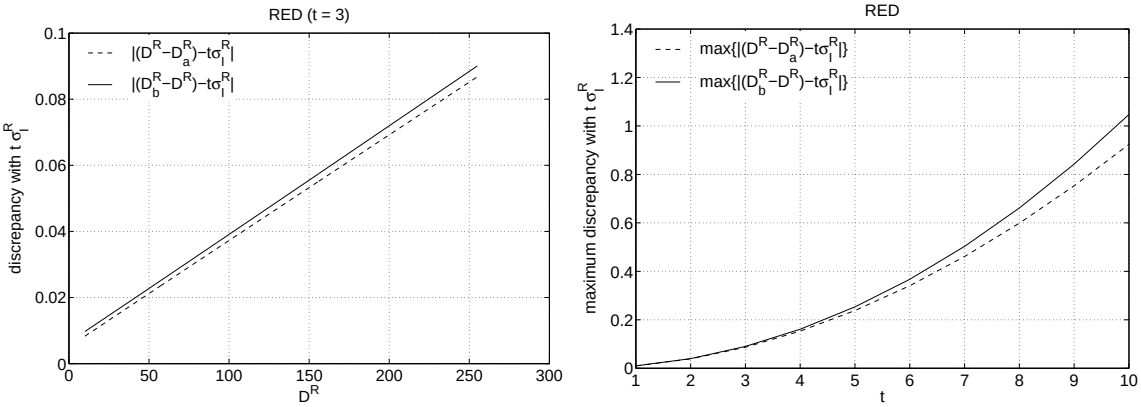


Fig. 6.3. [left] Discrepancy between $D^R - D^R_a$ and $t\sigma^R$, and between $D^R_b - D^R$ and $t\sigma^R$, for $t = 3$. [right] Maximum discrepancy between $D^R - D^R_a$ and $t\sigma^R$, and between $D^R_b - D^R$ and $t\sigma^R$, for $t = 1..10$.

$$\left(1 - t^2\mathrm{Var}_I[K]\right)(I^cA^c)^2 - (t^2A^c + 2(D^c - \mathrm{E}_I[\mu_{dc}]A^c))(I^cA^c) + \\ (D^c - \mathrm{E}_I[\mu_{dc}]A^c)^2 - t^2\left(\mathrm{Var}_I[\mu_{dc}](A^c)^2 + \mathrm{E}_I[\mu_{dc}](A^c)^2 + (\sigma^c_f)^2\right) = 0\,. \quad (6.12)$$

Once $(I^cA^c)_a$ and $(I^cA^c)_b$ are known for a certain D^c, a confidence interval $[D^c_a, D^c_b]$ can be calculated for D^c by means of $D^c_a = (I^cA^c)_a + \mathrm{E}_I[\mu_{dc}]A^c$ and $D^c_b = (I^cA^c)_b + \mathrm{E}_I[\mu_{dc}]A^c$ (see figure 6.2(right)).

Although this interval is not symmetric around D^c, a slight deviation has been observed for reasonable values of t and for the noise parameter values typically reported in the related literature [93, 262, 270] and the camera used in our experiments. Moreover, the differences $D^c - D^c_a$ and $D^c_b - D^c$ have been quite similar to $t\sigma^c_I$ (equation 6.11b) for low values of t. This fact is shown, for the red colour channel, in figure 6.3(left) for $t = 3$, while figure 6.3(right) plots the maximum discrepancy for several values of t, also for the red colour channel. In both plots, $A^R = 0.002$, $(\sigma^R_f)^2 = 1.0$, $\mathrm{E}_I[\mu_{dc}]A^R = 10$, $\sqrt{\mathrm{Var}_I[\mu_{dc}]}A^R = 1.0$ and $\sqrt{\mathrm{Var}_I[K]} = 0.006$, as before. As can be observed, up to approximately $t = 7$ the maximum discrepancy, attained for $D^R = 255$, lies below 0.5. Given that $t = 2, 3, 4$ are reasonable values for t by virtue of its statistical meaning, and given the later comments, it is clear that $D^c \pm t\sigma^c_I$ can result to be a simpler way for computing a meaningful confidence interval for D^c without solving equation 6.12, so that $\sigma^c_D \approx \sigma^c_I$.

6.3 Previous Work

In contrast with geometric calibration, for which lots of algorithms have been published, the radiometric calibration of vision cameras has been rarely studied within the vision research community. Healey and Kondepudy presented in [93] a monochrome camera noise model from where they derived several techniques based on plane cards to estimate the gain, the charge-independent noise, the DCNU and the PRNU. In their paper, the camera gain and the charge-independent noise were simultaneously estimated by linear regression using a set of difference images of a plane card taken with different neutral density filters placed in front of the camera. As for dark current, the authors propose to use a series of images taken in a dark environment (i.e. without removing the camera optics cap), while, regarding the PRNU, they suggest using a set of images of a plane card taken under different imaging configurations such that, for each configuration, every collection site images a different patch of the card illuminated by a different part of the source, i.e. the light source and the card are moved for each configuration. In this way, the effects of a spatially non-uniform illumination and surface reflectance are removed. The PRNU parameters are then estimated subtracting first an average dark image from every image of the plane card, to remove the dark current intensity, and dividing, next, the intensity at every point involved in the analysis by the average intensity over a 9×9 window of neighbours, which is supposed to be free of the effect of K.

After the paper by Healey and Kondepudy, Tarel proposed in [269, 270] several experiments for estimating the charge-independent noise, the dark current —using images taken under a dark environment again— and the joint effect of the fixed pattern array and the shadowing introduced by the camera optics due to vignetting. In all the cases, plane calibration cards were used. The calibration results were illustrated executing two common tasks in computer vision, edge detection and image segmentation, over the original and the corrected images.

Finally, Stokman makes use of a camera noise model consisting of camera gain, shot noise and dark current for error propagation inside an edge-detection framework [80, 262]. The author estimates the gain and the dark current variance using a series of images of a white reference card varying the lens aperture.

6.4 Experimental Results

Results for the radiometric calibration of the JAI CV-M70 camera used for assessing the performance of the lighting estimation methods are provided in this section, together with examples of the applications discussed in section 6.2. As before, halogen illumination and a COMET Matrox frame grabber complement the imaging equipment. White calibration cards are used to estimate the noise distribution parameters, although the particular reflectance is not relevant during the radiometric calibration provided that it is uniform.

Again, as a general rule throughout this section, uncertainties $\delta\widehat{x}$ corresponding to estimates $\widehat{x}$ are calculated through standard formulas when $\widehat{x}$ comes from a regression analysis, while, for averaged values, the Standard Deviation Of Mean, $\sigma_{\overline{x}} = \frac{\sigma_x}{\sqrt{N}}$, is used [272].

6.4.1 Estimation of noise distribution parameters

Both spatial and charge-independent non-spatial noise distribution parameters were measured by means of 12 sets of images of the calibration card, each set corresponding to a different lens aperture. Besides, each set of images consisted of 100 frames. For every set, images $\mu_{D^c}(i,j)$ and $\sigma^2_{D^c}(i,j)$ were calculated, which, in particular, taking into account table 4.2, allowed reducing the non-spatial noise in images $\mu_{D^c}(i,j)$ at about only 10% of their original magnitude, improving, thus, the calculation of $\mathrm{E}_I[\mu_{D^c}]$ and $\mathrm{Var}_I[\mu_{D^c}]$ (equations 6.5). Moreover, up to 12 different values for $I^c = (L_a\rho_b)^c + m_b(L_d\rho_b)^c$ could be generated to produce 12 pairs $(\mathrm{E}_I[\mu_{D^c}], \mathrm{Var}_I[\mu_{D^c}])$ and $(\mathrm{E}_I[\mu_{D^c}], \mathrm{E}_I[\sigma^2_{D^c}])$ for estimating the PRNU and the DCNU distribution parameters and the gain and charge-independent non-spatial noise of the camera under calibration. Finally, the estimation procedures were applied over the central 100×100 part of the images, in order to avoid the effects of nonuniform illumination and the distortion introduced by the camera optics (section 2.4.5).

With this set of images, camera gain A^c and the variance of the charge-independent non-spatial noise $(\sigma^c_f)^2$ were estimated for every colour channel c following pseudocode 6.2. Moreover, in order to force the parameters of the straight line to be positive, according to the signs of A^c and $(\sigma^c_f)^2$, non-negative least squares [138] was used. The resulting estimates were $(A^R, A^G, A^B) = (0.0046, 0.048, 0.074) \pm (0.0012, 0.0016, 0.0018)$ and $(\sigma^R_f, \sigma^G_f, \sigma^B_f) = (0.79, 0.59, 0.51) \pm (0.04, 0.06, 0.06)$, while the correlation coefficients for all three colour channels resulted to be (0.9634,0.9625,0.9761). By way of illustration, figure 6.4 shows the fittings corresponding to the different colour channels.

On the other hand, $\mathrm{Var}_I[K]$, $\mathrm{E}_I[\mu_{dc}]A^c$ and $\mathrm{Var}_I[\mu_{dc}](A^c)^2$ were measured following pseudocode 6.1. In order to improve the quality of the estimates, after the linear regression analysis performed to estimate the parameters of the parabola, a non-linear optimization refinement step based on equation 6.5b was applied next using as starting values for $\mathrm{Var}_I[K]$, $\mathrm{E}_I[\mu_{dc}]A^c$ and $\mathrm{Var}_I[\mu_{dc}](A^c)^2$ the ones corresponding to the output of the linear optimization step. In the linear step, non-negative least squares were used again. The measurements resulting for every colour channel were $(\sqrt{\mathrm{Var}_I[K]^R}, \sqrt{\mathrm{Var}_I[K]^G}, \sqrt{\mathrm{Var}_I[K]^B}) = (0.0056, 0.0061, 0.0056) \pm (0.0006, 0.0004, 0.0004)$, $(\mathrm{E}_I[\mu_{dc}]A^R$ $\mathrm{E}_I[\mu_{dc}]A^G, \mathrm{E}_I[\mu_{dc}]A^B) = (0.00, 0.00, 0.00) \pm (13.01, 6.83, 7.70)$ and $(\sqrt{\mathrm{Var}_I[\mu_{dc}]}A^R, \sqrt{\mathrm{Var}_I[\mu_{dc}]}A^G,$

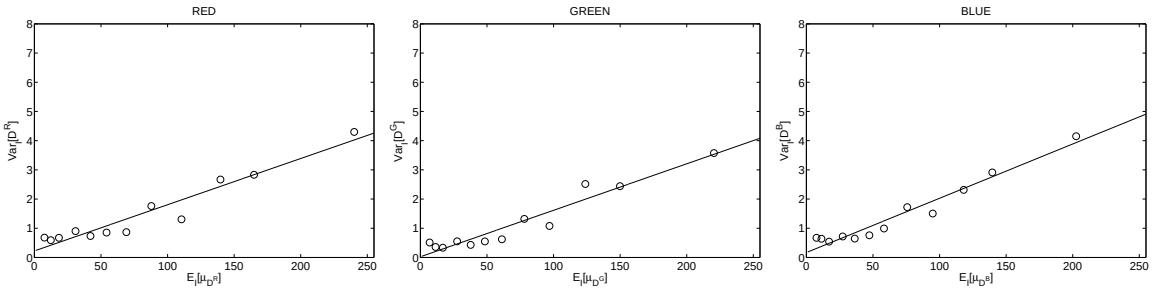


Fig. 6.4. Estimation of camera gain and charge-independent non-spatial noise.

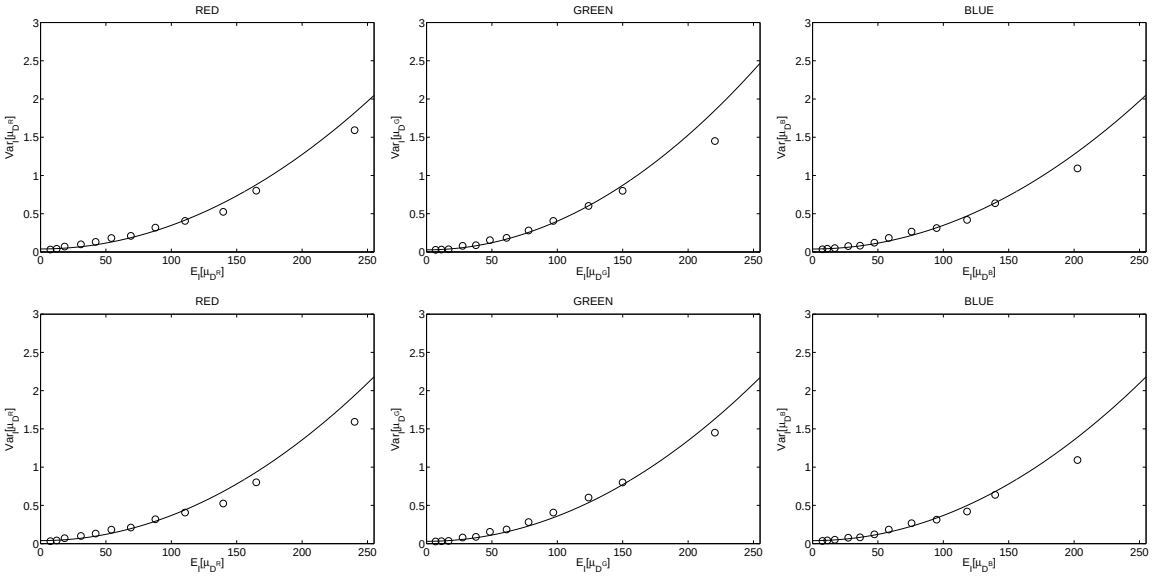


Fig. 6.5. Estimation of the PRNU and DCNU distribution parameters by means of different parabolas for every colour channel [upper row] and by means of centered parabolas sharing the first coefficient [lower row].

$\sqrt{\mathrm{Var}_I[\mu_{dc}]}A^B) = (0.20, 0.17, 0.20) \pm (0.03, 0.02, 0.02)$. The upper row of figure 6.5 shows the fitting parabola for every colour channel.

As can be observed, three slightly different values resulted for $\mathrm{Var}_I[K]$, one from every channel, while all three should coincide since the internal organization of the camera under calibration is based on one CCD. On the other hand, the confidence interval for the estimates of $\mathrm{E}_I[\mu_{dc}]A^c$ were too large so as to accept them; however, the intervals for the other parameters resulted to be quite acceptable. Because of this, a new regression analysis was performed with the intention of improving the calibration. On the one hand, the distribution parameters of the DCNU were measured averaging 100 images taken with the lens cap on (i.e. in a dark environment) and calculating, in turn, the mean and variance of the average image. As a result, a value of 0 resulted for both the spatial mean and variance. Since the mean calculated in this way coincided with the one from the regression analysis, and could be corroborated from the fact that intensity levels of 1 are also returned by the camera under calibration, this mean was accepted as a correct estimate. As for the variance, a value of 0 seemed quite suspicious, and with the intention of being more conservative with the estimations, it was rejected. Setting $\mathrm{E}_I[\mu_{dc}]A^c$ to 0 for all colour channels, pairs $(\mathrm{E}_I[\mu_{D^c}] - \mathrm{E}_I[\mu_{dc}]A^c, \mathrm{Var}_I[\mu_{D^c}]) = (\mathrm{E}_I[\mu_{D^c}], \mathrm{Var}_I[\mu_{D^c}])$ were fitted by centered parabolas, according to equation 6.5a, ensuring they all shared the same first coefficient $\mathrm{Var}_I[K]$ (see

Table 6.2. Comparison of the estimates resulting from the radiometric calibration.

			Healey and Kondepudy's	
	1st execution	2nd execution	1st execution	2nd execution
$\boldsymbol{A}$	(0.0046,0.0048,0.0074)	(0.0065,0.0053,0.0096)	(0.0031,0.0027,0.0047)	(0.0042,0.0028,0.0064)
$\boldsymbol{\sigma_f}$	(0.79,0.59,0.71)	(0.68,0.54,0.65)	(0.80,0.62,0.77)	(0.70,0.56,0.69)
$\sqrt{\mathrm{Var}_I[K]}$	0.0057	0.0059	0.0069	
$\mathrm{E}_I[\mu_{dc}]\boldsymbol{A}$	(0,0,0)	(0,0,0)	(0,0,0)	
$\sqrt{\mathrm{Var}_I[\mu_{dc}]}\boldsymbol{A}$	(0.20,0.17,0.20)	(0.17,0.16,0.18)	(0,0,0)	

appendix A for a description about the fitting method employed)[3]. In this case, the estimates were $\sqrt{\mathrm{Var}_I[K]} = 0.0057 \pm 0.0001$ and $(\sqrt{\mathrm{Var}_I[\mu_{dc}]}A^R, \sqrt{\mathrm{Var}_I[\mu_{dc}]}A^G, \sqrt{\mathrm{Var}_I[\mu_{dc}]}A^B) = (0.20, 0.17, 0.20) \pm (0.01, 0.01, 0.01)$, while the correlation coefficients resulted to be (0.9976,0.9939,0.9944). The lower row of figure 6.5 shows the fitting parabola for every colour channel. As can be noticed, the values obtained for the spatial variance of the DCNU are quite close to the values obtained in the previous regression analysis, while the spatial variance of the PRNU is among the three values obtained before.

The previous procedures were re-executed with a different set of images, but the same camera, in order to be able to observe the robustness of the calibration algorithm. The previous estimates and the new ones are provided in table 6.2 to make easier the comparison. Furthermore, in the same table, results for the Healey and Kondepudy's calibration algorithm (H&K) [93] are also provided. When calibrating the camera by means of H&K, the first and second images of every set were used to calculate the difference images needed by this algorithm to estimate A^c and $(\sigma_f^c)^2$. As for the determination of the distribution parameters of the PRNU and the DCNU, a different set of images was needed, which were taken using the instructions detailed in [93]; because of this, the part of the calibration related to the PRNU and the DCNU was executed only once for H&K.

As can be observed, some slight deviations between executions are observed in both methods as for the estimations of A^c and σ_f^c. On the other hand, H&K tends to yield lower values than R^2CIU for A^c (the slope of the straight line) but similar values as R^2CIU for σ_f^c (the intercept with the Y axis). This seems to be caused by the way how the fitting points are generated in H&K. In this algorithm, points $((\mathrm{E}_I[D_1^c] + \mathrm{E}_I[D_2^c])/2, \mathrm{Var}_I[D_1^c - D_2^c]/2)$ are used, where D_1^c and D_2^c are the two

[3] In fact, equations 6.5 can be transformed so that the estimation of the distribution parameters of both the PRNU and the DCNU are independent of the colour channel, dividing both sides by $(A^c)^2$ as shown next:

$$\frac{\mathrm{Var}_I\left[\mu_{D^c}\right]}{(A^c)^2} = \mathrm{Var}_I[K]\left(\frac{\mathrm{E}_I[\mu_{D^c}]}{A^c}\right)^2 - 2\mathrm{Var}_I[K]\mathrm{E}_I[\mu_{dc}]\left(\frac{\mathrm{E}_I[\mu_{D^c}]}{A^c}\right) + \mathrm{Var}_I[K]\left(\mathrm{E}_I[\mu_{dc}]\right)^2 + \mathrm{Var}_I[\mu_{dc}]\,. \tag{6.13}$$

In this way, the fitting should be performed over the pairs $(\mathrm{Var}_I\left[\mu_{D^c}\right]/(A^c)^2, \mathrm{E}_I[\mu_{D^c}]/A^c)$. This fact implies, however, to estimate previously A^c and then use the estimation to generate the fitting points. This is a way of action which, if possible, should be avoided in order to prevent the uncertainty in the A^c value from affecting the estimation of the distribution parameters of the PRNU and the DCNU.

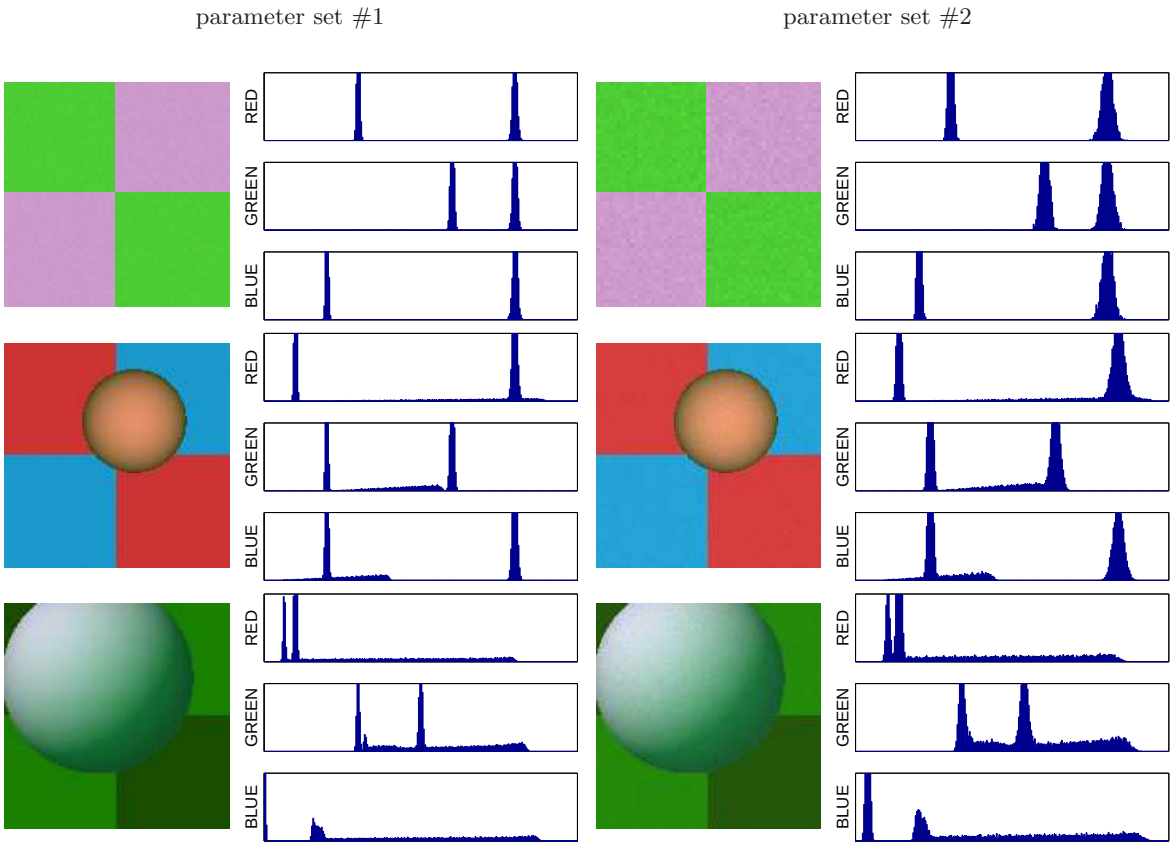


Fig. 6.6. Examples of synthetic noisy images generated by pseudocode 6.3.

images mentioned above. These points intend to play the same role as the ones required by R^2CIU, although, while $(E_I[D_1^c]+E_I[D_2^c])/2$ (used in H&K) and $E_I[\mu_{D^c}]$ (used in R^2CIU) tend to take similar values, $\mathrm{Var}_I[D_1^c - D_2^c]/2$ (used in H&K) tends to be more optimistic (i.e. lower variance) than $E_I[\sigma^2_{D^c}]$ (used in R^2CIU). As a result, lower slope straight lines are returned by H&K with regard to R^2CIU.

As for the distribution parameters of the PRNU and the DCNU, R^2CIU gives rise to almost identical estimates in both executions of the camera calibration. On the other hand, R^2CIU and H&K return similar or identical values for $\mathrm{Var}_I[K]$ and $E_I[\mu_{dc}]A^c$, while a value of 0 is returned by H&K for $\mathrm{Var}_I[\mu_{dc}](A^c)^2$. This is because H&K uses images taken in a dark environment to estimate both $E_I[\mu_{dc}]A^c$ and $\mathrm{Var}_I[\mu_{dc}](A^c)^2$.

6.4.2 Generation of noisy synthetic images

Just by way of illustration, figure 6.6 presents some synthetic images generated using the pseudocode 6.3 and two sets of radiometric performance parameters. The first one coincides with the first execution of R^2CIU (section 6.4.1). The second set of parameters corresponds to a noisier imaging device, for which $\boldsymbol{A} = (0.0020, 0.0020, 0.0020)$, $\boldsymbol{\sigma_f} = (1.00, 1.00, 1.00)$, $\mathrm{Var}_I[K] = 0.0200$, $E_I[\mu_{dc}]\boldsymbol{A} = (10.0, 10.0, 10.0)$ and $\sqrt{\mathrm{Var}_I[\mu_{dc}]}\boldsymbol{A} = (1.00, 1.00, 1.00)$. A histogram for every colour channel appears next to the corresponding image to be able to appreciate the amount of noise introduced for every parameter set.

6.4.3 Determination of irradiance measurement uncertainties

Figure 6.7 contains several plots regarding the computation of intensity uncertainties according to equations 6.11 and 6.12. The averaged parameters between the two executions of R^2CIU have been

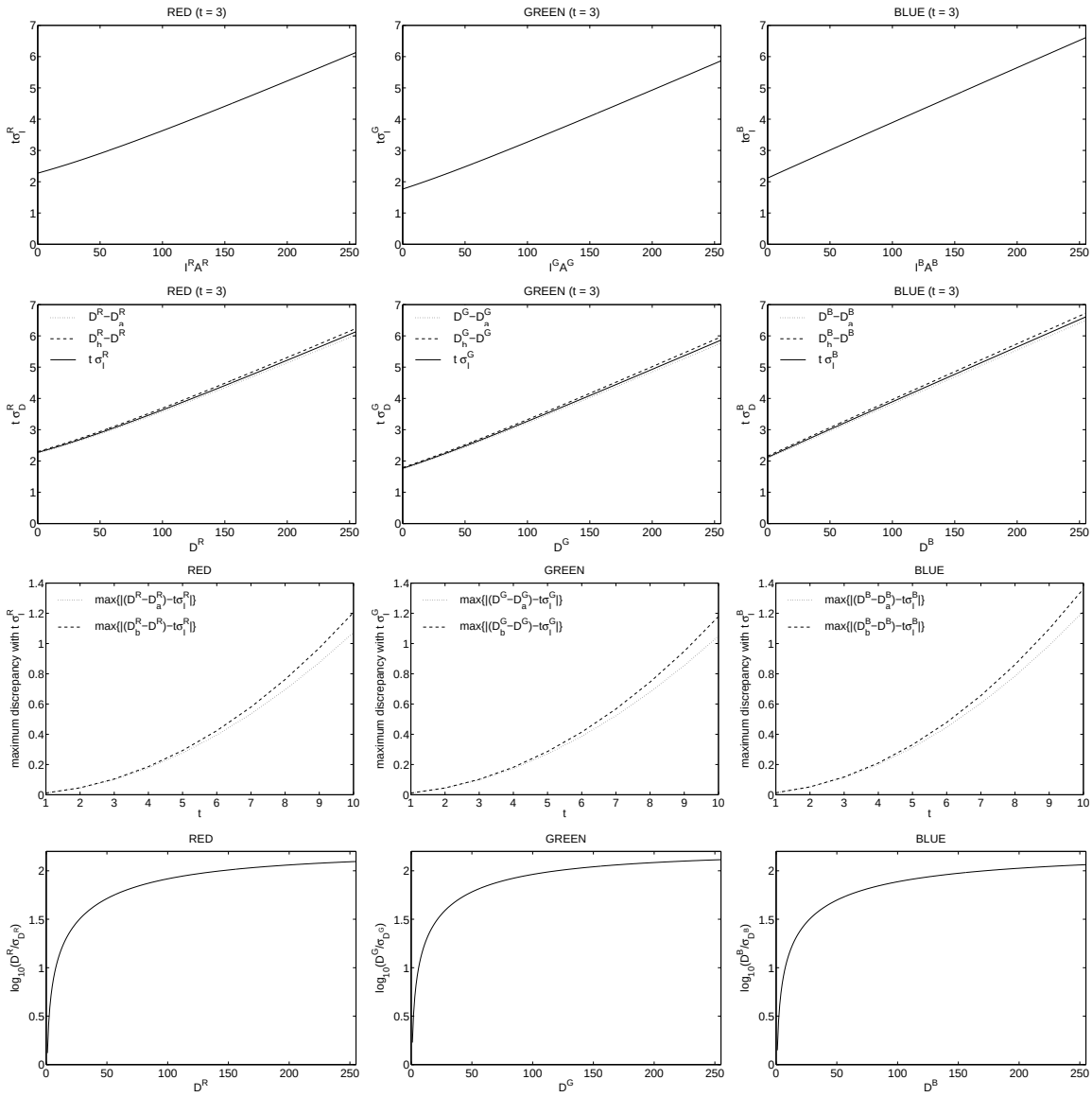


Fig. 6.7. Estimation of intensity uncertainties for every colour channel: [1st row] Value of $t\sigma_I^c$ of equation 6.11 for $t = 3$. [2nd row] Value of $D^c - D_a^c$ and $D_b^c - D^c$ for $t = 3$. [3rd row] Maximum discrepancy between $D^c - D_a^c$, $D_b^c - D^c$ and $t\sigma_I^c$ for $t = 1..10$. [4th row] Corresponding Signal-to-Uncertainty Ratio, computed as $\mathrm{SUR} = \log_{10}\left(\frac{D^c}{\sigma_D^c}\right)$.

used in all cases. On the one hand, the first row plots $t\sigma_I^c$ of equation 6.11b for $t = 3$ against I^cA^c for every colour channel. The following row corresponds to D_a^c and D_b^c using equation 6.12 and adding $\mathrm{E}_I[\mu_{dc}]A^c$ to the $(I^cA^c)_a$ and $(I^cA^c)_b$ resulting values. Observe that $t\sigma_I^c$ is also plotted to compare with D_a^c and D_b^c. The third row shows the extent to which the approximation σ_I^c can serve as a valid approximation of σ_D^c. As can be noticed, for the parameters obtained from the calibration of the camera, up to $t = 6$ the maximum discrepancy is below 0.5. In the same figure, the fourth row plots a sort of Signal-to-Noise Ratio for every digital value D^c which is called Signal-to-Uncertainty Ratio and that is calculated as $\mathrm{SUR} = \log_{10}(D^c/\sigma_D^c)$. Finally, table 6.3 details the values of σ_D^c and $D^c \pm \sigma_D^c$ for a certain range of values of D^c, while figure 6.8 compares predicted uncertainties with the real uncertainties found in real images of plane cards taken with different lens apertures. In this last figure,

Table 6.3. Excerpt of the uncertainty table for the red, green and blue colour channels.

D^c	D_a^R	D_b^R	σ_{D^R}	D_a^G	D_b^G	σ_{D^G}	D_a^B	D_b^B	σ_{D^B}
0	-0.76	0.76	0.76	-0.59	0.59	0.59	-0.71	0.71	0.71
1	0.24	1.76	0.76	0.41	1.59	0.59	0.29	1.71	0.71
2	1.23	2.77	0.77	1.40	2.60	0.60	1.28	2.72	0.72
3	2.23	3.77	0.77	2.40	3.60	0.60	2.28	3.72	0.72
4	3.23	4.77	0.77	3.39	4.61	0.61	3.27	4.73	0.73
5	4.22	5.78	0.78	4.39	5.61	0.61	4.26	5.74	0.74
6	5.22	6.78	0.78	5.39	6.61	0.61	5.26	6.74	0.74
7	6.22	7.78	0.78	6.38	7.62	0.62	6.25	7.75	0.75
8	7.21	8.79	0.79	7.38	8.62	0.62	7.25	8.75	0.75
9	8.21	9.79	0.79	8.37	9.63	0.63	8.24	9.76	0.76
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
110	108.74	111.26	1.26	108.86	111.14	1.14	108.64	111.36	1.36
111	109.73	112.27	1.27	109.85	112.15	1.15	109.64	112.36	1.36
112	110.73	113.27	1.27	110.84	113.16	1.16	110.63	113.37	1.37
113	111.72	114.28	1.28	111.84	114.16	1.16	111.63	114.37	1.37
114	112.72	115.28	1.28	112.83	115.17	1.17	112.62	115.38	1.38
115	113.71	116.29	1.29	113.83	116.17	1.17	113.61	116.39	1.39
116	114.71	117.29	1.29	114.82	117.18	1.18	114.61	117.39	1.39
117	115.70	118.30	1.30	115.82	118.18	1.18	115.60	118.40	1.40
118	116.70	119.30	1.30	116.81	119.19	1.19	116.60	119.40	1.40
119	117.69	120.31	1.31	117.81	120.19	1.19	117.59	120.41	1.41
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
246	244.01	247.99	1.99	244.10	247.90	1.90	243.85	248.15	2.15
247	245.00	249.00	2.00	245.09	248.91	1.91	244.84	249.16	2.16
248	246.00	250.00	2.00	246.08	249.92	1.92	245.84	250.16	2.16
249	246.99	251.01	2.01	247.08	250.92	1.92	246.83	251.17	2.17
250	247.98	252.02	2.02	248.07	251.93	1.93	247.83	252.17	2.17
251	248.98	253.02	2.02	249.07	252.93	1.93	248.82	253.18	2.18
252	249.97	254.03	2.03	250.06	253.94	1.94	249.81	254.19	2.19
253	250.97	255.03	2.03	251.06	254.94	1.94	250.81	255.19	2.19
254	251.96	256.04	2.04	252.05	255.95	1.95	251.80	256.20	2.20
255	252.96	257.04	2.04	253.05	256.95	1.95	252.80	257.20	2.20

real uncertainties for two different sets of images have been plotted, labelled as *real-1* and *real-2*. As can be observed, real and predicted uncertainties compare well, with, perhaps, a larger difference for the blue channel.

6.4.4 An example on the use of intensity uncertainties

To illustrate the use of the uncertainty estimates produced by R^2CIU, an edge detection task based on detecting LOG zero-crossings is next presented. In this case, uncertainties are used to devise a strategy for adaptive thresholding which will be called UTLOG (*Uncertainty-based Thresholding for LOG zero-crossings*). In particular, a detected LOG zero-crossing is classified as relevant if the positive and negative peak LOG values along the direction of detection are larger than t times the respective uncertainties. Those uncertainties are computed using standard uncertainty propagation rules, by which, if the output of the LOG operator is calculated as $f = \sum_x D^k(x)m(x)$, where $m(x)$ would be the

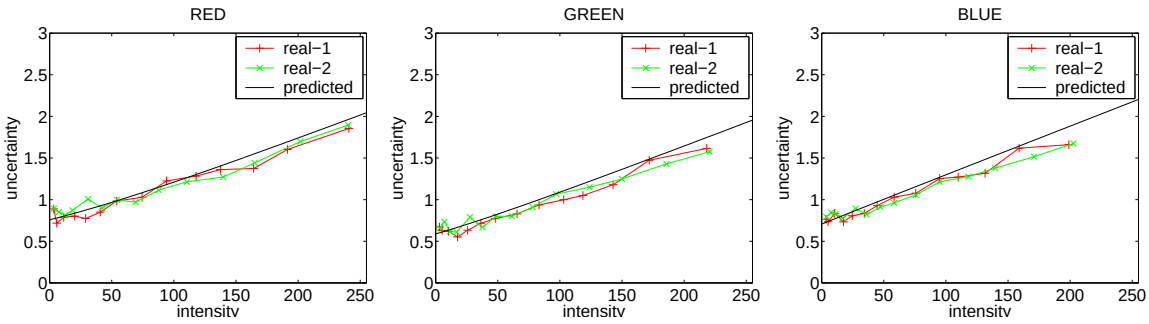


Fig. 6.8. Predicted *versus* real uncertainties for all colour channels.

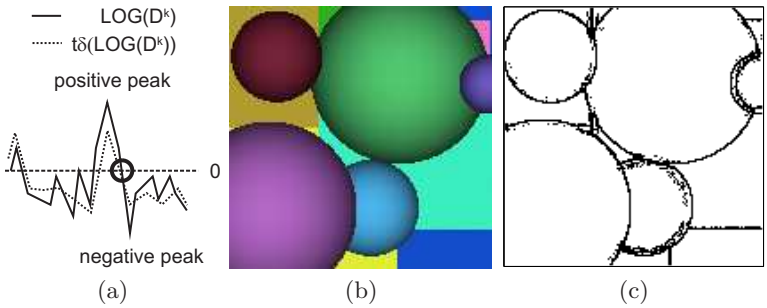


Fig. 6.9. Illustration of UTLOG: (a) only the LOG zero-crossing behind the circle is considered relevant by UTLOG; (b) example of noisy synthetic image; (c) resultant edge map for $t = 3$.

LOG mask constants, then $\delta(f) = \sqrt{\sum_x \delta(D^k(x))^2 m^2(x)}$ [272]. By way of example, see figure 6.9(a), where only the zero-crossing under the circle would be considered relevant. Figure 6.9(c) shows the edge map corresponding to the synthetic image on its left.

In order to study the performance of UTLOG, and, by extension, the usefulness of the estimated uncertainties, UTLOG was compared with a non-adaptive method which will be called NATLOG (*Non-Adaptive Thresholding for LOG zero-crossings*) in which LOG zero-crossings was selected by requiring negative and positive LOG peaks above a global threshold independent of the intensities involved. The goal of the experiment consisted in determining the optimum value of the only parameter of each algorithm (t for UTLOG and the aforementioned threshold for NATLOG) for each image of a series of noisy synthetic images. A lower standard deviation of the optimum values found would mean, in this experiment, a larger easiness for finding the proper value and, thus, a larger adaptivity. Furthermore, the standard deviation of the LOG operator was set to 1.0 in all cases. The details of the experiment are provided next:

- The comparison was performed over two sets of 100 synthetic images to which noise was added according to the camera noise parameters estimated in section 6.4.1 following pseudocode 6.3. One set corresponded to scenes involving spheres and planes of different reflectances (see figure 6.9(b) for an example), while the other set only imaged planes, and, therefore, implied, *a priori*, less difficulty for detecting the correct edges for a strategy based on LOG zero-crossings.
- Besides, up to four empirical discrepancy evaluation measures were used to find the optimum values: the often-cited Pratt's Figure of Merit (FOM) [232] with (1) $\alpha = 1/9$ and (2) $\alpha = 1$; (3) the discrepancy percentage (D) [239, 306]; and (4) the Baddeley measure [6]. In this way, the result of the experiment did not depend on the optimization strategy employed.
- As for the range of values considered for the respective parameters, for UTLOG, $t \in [1..15]$, while, for NATLOG, global integer thresholds between 1 and 30 were considered. As can be observed, the range of values is sufficiently large so as to allow the optimization strategy to find the global maximum (for FOM) or minimum (for D and Baddeley's measures).

The quantitative results of the experiment appear in figure 6.10. All these quantities come from a matrix $M_{100\times N}$ containing the evaluation measures, where every row of M corresponds to one image of the set, and every column is a value of the respective parameter ($N = 15$ for UTLOG and $N = 30$ for NATLOG). The following quantities are computed from M: (In the following, $\mathrm{E}[\cdot]$ and $\mathrm{Var}[\cdot]$ stand for, respectively, mean and variance, while *arg best* corresponds to *arg max* for FOM and to *arg min* for the rest of evaluation measures.)

- $avgP$ and $stdP$: respectively, average and standard deviation of the parameter values for which the best evaluation measure is attained for every image,

$$avgP = \underset{i=1..100}{\mathrm{E}}\left[\underset{j=1..N}{\arg\,\mathrm{best}}\left\{M(i,j)\right\}\right] \tag{6.14}$$

$$stdP = \sqrt{\underset{i=1..100}{\mathrm{Var}}\left[\underset{j=1..N}{\arg\,\mathrm{best}}\left\{M(i,j)\right\}\right]} \tag{6.15}$$

- $avgM1$: average of the best measure values among all the images,

$$avgM1 = \underset{\substack{i\,=\,1..100\\ j\,=\,\underset{k=1..N}{\arg\,\mathrm{best}}\{M(i,k)\}}}{\mathrm{E}}\left[M(i,j)\right] \tag{6.16}$$

- $avgM2$: minimum and maximum average of the evaluation values attained for every value of the parameter,

$$avgM2 = \left[\min_{j=1..N}\left\{\underset{i=1..100}{\mathrm{E}}\left[M(i,j)\right]\right\},\ \max_{j=1..N}\left\{\underset{i=1..100}{\mathrm{E}}\left[M(i,j)\right]\right\}\right] \tag{6.17}$$

- $range$: range of evaluation values achieved throughout the experiment,

$$range = \left[\min_{\substack{i\,=\,1..100\\ j\,=\,1..N}}\left\{M(i,j)\right\},\ \max_{\substack{i\,=\,1..100\\ j\,=\,1..N}}\left\{M(i,j)\right\}\right] \tag{6.18}$$

Besides, in the tables, the value of the parameter for which the best evaluation value was attained is indicated in parenthesis (columns $avgM2$ and $range$).

As can be observed in the tables of figure 6.10, the standard deviation of the parameter values (column $stdP$, in bold face in the tables) is quite lower for UTLOG than for NATLOG in all cases. Besides, for UTLOG, $t \in [3..5]$ seem to work well most times, while for NATLOG the global threshold interval seems to widen so as to cover values between 5-6 and 11-12, or more, confirming, thus, the higher adaptivity of UTLOG thanks to the incorporation of intensity uncertainties in the thresholding. As for the evaluation measures attained, at the optima, they are almost identical for both cases, but the other end of the intervals are always better for UTLOG than for NATLOG (columns $avgM2$ and $range$).

6.5 Conclusions

An algorithm called R^2CIU (*Robust Radiometric Calibration and Intensity Uncertainty estimation*) for radiometric camera calibration has been proposed. It consists of pseudocodes 6.1 and 6.2 and allows estimating the distribution parameters of the noise sources of the camera noise model proposed by Healey and Kondepudy in [93], described in section 2.4.4. Besides, two applications of the radiometric calibration of CCD cameras have also been discussed, namely synthesis of noisy images for testing purposes, and estimation of intensity uncertainties, both according to the radiometric performance of a particular camera, or, in other words, a particular set of camera noise model parameters.

More precisely, R^2CIU is based on uniform reflectance calibration cards. Besides, the fact that all the pixels of the captured image can participate in the estimation makes R^2CIU a robust estimator

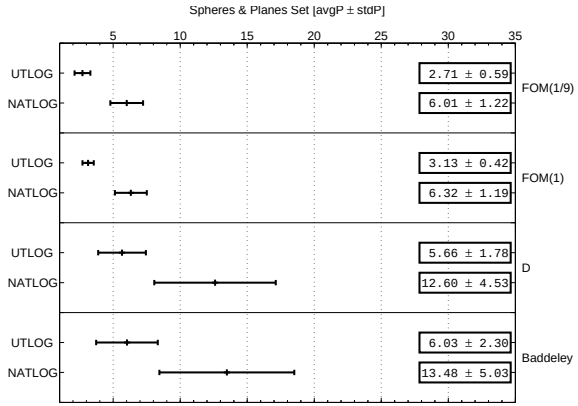
(a) Spheres & Planes set

measure	UTLOG					NATLOG				
	avgP	**stdP**	avgM1	avgM2	range	avgP	**stdP**	avgM1	avgM2	range
FOM(1/9)	2.71	**0.59**	0.901	[0.550,0.881(3)]	[0.333,0.974(4)]	6.01	**1.22**	0.913	[0.328,0.889(5)]	[0.198,0.964(8)]
FOM(1)	3.13	**0.42**	0.806	[0.363,0.801(3)]	[0.225,0.921(4)]	6.32	**1.19**	0.811	[0.185,0.990(6)]	[0.100,0.910(6)]
D	5.66	**1.78**	0.025	[0.026(5),0.182]	[0.009(4),0.236]	12.60	**4.53**	0.026	[0.027(11),0.470]	[0.010(8),0.549]
Baddeley	6.03	**2.30**	0.347	[0.360(5),2.979]	[0.205(4),3.419]	13.48	**5.03**	0.359	[0.379(11),3.884]	[0.209(9),4.231]

(b) Only-Planes set

measure	UTLOG					NATLOG				
	avgP	**stdP**	avgM1	avgM2	range	avgP	**stdP**	avgM1	avgM2	range
FOM(1/9)	2.47	**1.17**	0.963	[0.567,0.959(2)]	[0.329,1.000(1)]	6.44	**3.13**	0.967	[0.391,0.959(6)]	[0.211,1.000(1)]
FOM(1)	2.67	**1.22**	0.934	[0.430,0.925(2)]	[0.220,1.000(1)]	6.64	**3.13**	0.936	[0.262,0.929(7)]	[0.112,1.000(1)]
D	4.28	**2.15**	0.010	[0.011(3),0.160]	[0.000(2),0.242]	10.33	**4.54**	0.010	[0.011(8),0.505]	[0.000(6),0.710]
Baddeley	7.01	**3.67**	0.206	[0.236(3),3.084]	[0.000(2),4.051]	17.77	**8.08**	0.200	[0.236(10),3.957]	[0.000(6),4.729]

(c) Spheres & Planes set (only avgP and stdP)



(d) Only-Planes set (only avgP and stdP)

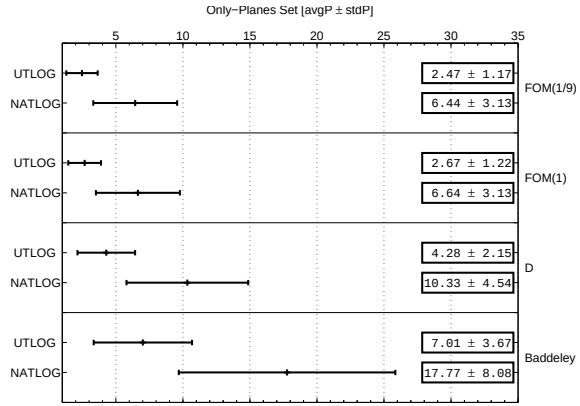


Fig. 6.10. Quantitative results of the comparison between UTLOG and NATLOG.

of the noise model parameters. Nevertheless, special care should be taken when calibrating cameras with an only CCD since the demosaicking process can introduce a certain level of correlation between pixel values, what could compromise the accuracy of the estimates. In this case, using the pattern of the colour filter array to select the image cells corresponding to each colour channel and avoid the interpolated pixels can solve the problem (see figure 2.11). Another solution could be to use raw pixel values in case the camera is able to produce images in RAW format [236].

R^2CIU has been compared against itself using two sets of calibration images taken at different time instants with slight deviations in the parameters estimated for both sets. It also compares favourably with the algorithm by Healey and Kondepudy [93], showing a similar behaviour but a somewhat simpler use.

Finally, examples of the applications of the radiometric calibration of CCD cameras have been provided. In particular, an uncertainty table has been built which predicts an uncertainty of approximately ± 2 intensity levels at maximum intensity level.

7

Image Segmentation by Scene Reconstruction

Making use of equation 4.7, a region-growing approach called IS^{2}R (*Image Segmentation by Scene Reconstruction*) is presented in this chapter. It is based on estimating the surface reflectance and using it as a homogeneity criterion to join pixels to the currently grown region. To this end, the scene shape is (partially) reconstructed from the image data. The term *quantitative* can also be used for this approach because it heavily relies on the numerical aspects of the formation of the image. That is to say, reflectance, light distribution and scene geometry numerical estimations are essential in the process of image partitioning.

Section 7.1 depicts the model of image formation assumed by IS^{2}R. Next, section 7.2 overviews the general method, and the details are given in sections 7.3, 7.4, 7.5 and 7.6. Some segmentation anomalies are discussed in section 7.7. Later, section 7.8 discusses how IS^{2}R could handle some extensions to the model of image formation. Next, section 7.9 presents the experiments performed, and, finally, section 7.10 concludes the chapter.

7.1 Model of Image Formation

The image formation model assumed by IS^{2}R is simpler than the one presented in chapter 4. This is because of the quantitative nature of the method itself; more precisely, the technique used to reconstruct the shape of scene objects, which is essential as will be seen later, requires some constraints to be met about the formation of the image.

The features of the simplified model are summarized next:

(1) Only matte surfaces are considered. Therefore, specularities are not expected, and the corresponding terms can be removed from equation 4.1. A further implication of this simplification is the fact that ρ_a, which is usually considered as a linear combination of ρ_b and ρ_i, turns out to be $\rho_a = \rho_b$, and only ρ_b needs, thus, to be included in the new model of image formation.

(2) The lighting distribution consists of: (a) one point light source located at infinity, whose strength is uniform throughout the scene; and (b) ambient light, also of uniform strength. As the scene is illuminated only by one directional light source, $ls = 1$ in equation 4.1, and, accordingly, k can be removed from $\boldsymbol{s}_k(i,j)$ in equation 4.2. On the other hand, if such directional light source can be considered a point and it is located at infinity, then $\boldsymbol{s}(i,j)$ is independent of the image location (i,j). That is to say, illumination is considered parallel (i.e. collimated) throughout the scene.

(3) The focal length of the camera is assumed large enough so that parallel projection can be considered a good approximation of perspective. Being this the case, scene locations $p = (\mathrm{x}, \mathrm{y}, \mathrm{z})$ are such that $\mathrm{x} = i$ and $\mathrm{y} = j$.

(4) As for the camera, only gray-level images are included in the model. Therefore, the term $\tau^c(\lambda)$ corresponding to colour filter c is also removed from the image formation equation.

The resulting image formation model is expressed in equation 7.1:

$$D(i,j) = q_0 \int_\Lambda L_a(\lambda)\rho_b(i,j;\lambda)s(\lambda)\,d\lambda + q_0\,(\boldsymbol{n}(i,j)\cdot\boldsymbol{s}) \int_\Lambda L_d(\lambda)\rho_b(i,j;\lambda)s(\lambda)\,d\lambda \tag{7.1}$$

where D refers to the gray-level nature of the image.

It is common practice to introduce a further simplification for every material $\mathcal{M}$ in the scene, consisting in the following approximations (see [304] and more recently [28, 55, 105, 154, 299], among many others):

$$\int_\Lambda L_a(\lambda)\rho_b^{\mathcal{M}}(i,j;\lambda)s(\lambda)\,d\lambda \approx L_a\rho_b^{\mathcal{M}}(i,j) \tag{7.2}$$

$$\int_\Lambda L_d(\lambda)\rho_b^{\mathcal{M}}(i,j;\lambda)s(\lambda)\,d\lambda \approx L_d\rho_b^{\mathcal{M}}(i,j) \tag{7.3}$$

That is to say, the non-geometric parts of the ambient lighting and body reflection terms are approximated by a product of two factors independent of wavelength and related to, respectively, surface reflectance at location (i,j), $\rho_b(i,j)$, and to illumination, L_a and L_d.

Note that those approximations, though traditionally not explicitly stated, implicitly assume that, either (1) and/or (2) below take place: (In the expressions below, the location coordinates (i,j) have been removed to make the notation easier.)

(1) For every material $\mathcal{M}$ in the scene:

$$\frac{\int_\Lambda L_a(\lambda)\rho_b^{\mathcal{M}}(\lambda)s(\lambda)\,d\lambda}{\int_\Lambda L_a(\lambda)s(\lambda)\,d\lambda} \approx \widetilde{\rho}_b^{\mathcal{M}} \tag{7.4}$$

$$\frac{\int_\Lambda L_d(\lambda)\rho_b^{\mathcal{M}}(\lambda)s(\lambda)\,d\lambda}{\int_\Lambda L_d(\lambda)s(\lambda)\,d\lambda} \approx \widetilde{\rho}_b^{\mathcal{M}} \tag{7.5}$$

so that:

$$\int_\Lambda L_a(\lambda)\rho_b^{\mathcal{M}}(\lambda)s(\lambda)\,d\lambda \approx \left(\int_\Lambda L_a(\lambda)s(\lambda)\,d\lambda\right)\widetilde{\rho}_b^{\mathcal{M}} = L_a\widetilde{\rho}_b^{\mathcal{M}} \tag{7.6}$$

$$\int_\Lambda L_d(\lambda)\rho_b^{\mathcal{M}}(\lambda)s(\lambda)\,d\lambda \approx \left(\int_\Lambda L_d(\lambda)s(\lambda)\,d\lambda\right)\widetilde{\rho}_b^{\mathcal{M}} = L_d\widetilde{\rho}_b^{\mathcal{M}} \tag{7.7}$$

Note that $\rho_b(\lambda) = \rho_b\,, \forall\lambda$, is a particular case for which equations 7.4 and 7.5 are simultaneously satisfied. Finally, the validity of the approximation would depend on the following sum:

$$V = \sum_{\mathcal{M}} \left(\frac{\int_\Lambda L_a(\lambda)\rho_b^{\mathcal{M}}(\lambda)s(\lambda)\,d\lambda}{\int_\Lambda L_a(\lambda)s(\lambda)\,d\lambda} - \frac{\int_\Lambda L_d(\lambda)\rho_b^{\mathcal{M}}(\lambda)s(\lambda)\,d\lambda}{\int_\Lambda L_d(\lambda)s(\lambda)\,d\lambda} \right)^2 \tag{7.8}$$

The larger V, the less valid would be the approximation.

(2) For every material $\mathcal{M}$ in the scene:

$$\frac{\int_\Lambda L_a(\lambda)\rho_b^{\mathcal{M}}(\lambda)s(\lambda)\,d\lambda}{\int_\Lambda \rho_b^{\mathcal{M}}(\lambda)s(\lambda)\,d\lambda} \approx \widetilde{L}_a \tag{7.9}$$

$$\frac{\int_\Lambda L_d(\lambda)\rho_b^{\mathcal{M}}(\lambda)s(\lambda)\,d\lambda}{\int_\Lambda \rho_b^{\mathcal{M}}(\lambda)s(\lambda)\,d\lambda} \approx \widetilde{L}_d \tag{7.10}$$

so that:

$$\int_\Lambda L_a(\lambda)\rho_b^{\mathcal{M}}(\lambda)s(\lambda)\,d\lambda \approx \widetilde{L}_a\left(\int_\Lambda \rho_b^{\mathcal{M}}(\lambda)s(\lambda)\,d\lambda\right) = \widetilde{L}_a\rho_b^{\mathcal{M}} \tag{7.11}$$

$$\int_\Lambda L_d(\lambda)\rho_b^{\mathcal{M}}(\lambda)s(\lambda)\,d\lambda \approx \widetilde{L}_d\left(\int_\Lambda \rho_b^{\mathcal{M}}(\lambda)s(\lambda)\,d\lambda\right) = \widetilde{L}_d\rho_b^{\mathcal{M}} \tag{7.12}$$

Note that white illumination, i.e. $L_a(\lambda) = L_a$ and $L_d(\lambda) = L_d\,, \forall\lambda$, is a particular case for which equations 7.9 and 7.10 are simultaneously satisfied. In this case, the validity of the approximation would depend on the following two sums:

$$V_1 = \sum_{\mathcal{M}_1 \neq \mathcal{M}_2} \left(\frac{\int_\Lambda L_a(\lambda)\rho_b^{\mathcal{M}_1}(\lambda)s(\lambda)\,d\lambda}{\int_\Lambda \rho_b^{\mathcal{M}_1}(\lambda)s(\lambda)\,d\lambda} - \frac{\int_\Lambda L_a(\lambda)\rho_b^{\mathcal{M}_2}(\lambda)s(\lambda)\,d\lambda}{\int_\Lambda \rho_b^{\mathcal{M}_2}(\lambda)s(\lambda)\,d\lambda} \right)^2 \tag{7.13}$$

$$V_2 = \sum_{\mathcal{M}_1 \neq \mathcal{M}_2} \left(\frac{\int_\Lambda L_d(\lambda)\rho_b^{\mathcal{M}_1}(\lambda)s(\lambda)\,d\lambda}{\int_\Lambda \rho_b^{\mathcal{M}_1}(\lambda)s(\lambda)\,d\lambda} - \frac{\int_\Lambda L_d(\lambda)\rho_b^{\mathcal{M}_2}(\lambda)s(\lambda)\,d\lambda}{\int_\Lambda \rho_b^{\mathcal{M}_2}(\lambda)s(\lambda)\,d\lambda} \right)^2 \tag{7.14}$$

The larger $V_1 + V_2$, the less valid would be the approximation.

Irrespective of the case, the satisfaction of equations 7.2 and 7.3 will be assumed from now on. Furthermore, factor q_0 from equation 7.1 will be considered embedded in either $\rho_b(i,j)$, L_a or L_d, so that the definitive model of image formation will be given by equation 7.15:

$$D(i,j) = (L_a\rho_b(i,j)) + (\boldsymbol{n}(i,j)\cdot\boldsymbol{s})\,(L_d\rho_b(i,j)) \tag{7.15}$$

7.2 Overview of IS^{2}R

7.2.1 Basic strategy

According to equation 4.7, surface reflectance can be determined if the lighting distribution and the scene geometry are known. If the image formation model corresponds to equation 7.15, then function $\mathcal{F}$ can be reversed (i.e. an expression for $\mathcal{F}^{-1}$ can be obtained), and surface reflectance can be recovered:

$$\rho_b(i,j) = \mathcal{F}^{-1}(D(i,j), \boldsymbol{n}(i,j), \boldsymbol{s}, L_a, L_d) = \frac{D(i,j)}{L_a + (\boldsymbol{n}(i,j)\cdot\boldsymbol{s})\,L_d} \tag{7.16}$$

In this way, the segmentation process just consists in joining pixels with similar reflectance.

In order to apply equation 7.16, L_a, L_d, $\boldsymbol{s}$ and $\boldsymbol{n}(i,j)$ must be known in advance. In particular, knowing $\boldsymbol{n}(i,j)$ for every image pixel implies having determined the shape of the scene —the estimation need not be complete, it suffices with the surface normal vectors of the visible points of the scene. To this end, a procedure capable of producing shape information from images must thus be embedded within the segmentation process. Accordingly, *shape from X* methods are briefly revised in the following sections.

7.2.2 Shape from X methods

In the computer vision literature, the term *shape from X* refers to a family of methods aiming at obtaining a shape description of the imaged scene, where X represents the cue which, mostly inspired in human vision, is used to compute the shape information. Among the different methods (see [119,217] for a survey), some are based on the information contained in a single image, while others require a special image acquisition infrastructure in order to take several images of the same scene, but varying one of the parameters of the imaging system in a controlled way.

Shape from shading and *shape from texture* are examples of the first group. The former yields shape information by analyzing the brightness variations in the image due to the scene objects curvature (i.e. their shading in the image). In the latter, the shape information is produced from the distortion suffered by individual *textons*[1] during the imaging process.

Shape from stereo, *shape from motion* and *shape from photometric stereo* are examples of the second group. In the first case, disparities between images seen from slightly different viewpoints (by,

[1] The term *texton* was introduced by Julesz in [114] as the supposed unit of pre-attentive human texture perception. They are also known as *texels*.

perhaps, a binocular pair of cameras) are exploited to estimate scene shape. In the second case, a time image sequence is used in conjunction with knowledge about the corresponding trajectory of the camera; a moving camera is therefore needed. Finally, the third method is based on taking images with different illumination conditions and using the brightness variations together with knowledge about the lighting in every case. In this third example, a set of light sources with proper mechanisms to turn each one on and off as required is thus needed.

Part of these methods are discarded regarding their use in the segmentation process by their very nature. This is the case of shape from texture and shape from stereo: both need textured surfaces to produce accurate results, case in which a segmentation according to surface reflectance does not make much sense. Among the remaining methods, shape from shading seems the most attractive for a segmentation strategy in the sense that no special infrastructure needs to be arranged to obtain shape information, and this is the reason why it is used in the segmentation method which is proposed.

7.2.3 Shape from Shading

The *shape from shading* (SFS) problem was formerly formulated by Horn in the 1970s in his doctoral thesis as *"given an intensity image of a continuous surface with constant, known reflectance and illumination, recover the shape of the surface"* [100]. The SFS problem is usually expressed as solving for $\boldsymbol{n}(i,j)$ in the so-called brightness equation $E(i,j) = R(\boldsymbol{n}(i,j))$, where $E(i,j)$ is the total irradiance at image point (i,j), which would be substituted by $D(i,j)$ since the intensity is a well-behaved function of brightness (equation 2.32), and $R(\boldsymbol{n}(i,j))$ is the reflectance map, a mapping between $\boldsymbol{n}$ and the irradiance values E, already discussed in section 2.3.1.

Since Horn's seminal work, several algorithms have been proposed (see [17, 24, 38, 141, 225, 283, 300, 311], among others), a comprehensible survey of which, together with a performance evaluation of the algorithms surveyed, can be found in [302–304] and also in [119]. Most SFS algorithms make some assumptions in order to obtain a solution of the brightness equation [119]:

- the light distribution is uniform throughout the scene and comes from a distant point light source with both unit direction $\boldsymbol{s}$ and radiance L_d known;
- the scene just contains one object, or at least, if there is more than one, they can be distinguished in the image by means of a previously computed segmentation;
- the surface reflection properties of scene objects are known, and in particular they obey the Lambert's Cosine Law $R(\boldsymbol{n}(i,j)) = \cos\theta(i,j) = \boldsymbol{n}(i,j) \cdot \boldsymbol{s}$;
- scene objects behave as perfect reflectors and do not absorb any radiation for any wavelength, or, if they do, the fraction of reflected radiation against incoming radiation, the reflectance, ρ_b, is known for every image point, and can be discounted from the intensity, $\widetilde{D}(i,j) = D(i,j)/\rho_b(i,j)$, before solving the brightness equation; and
- the sensor is linear and produces images which can be approximated by orthographic projection.

Some of these constraints are included in the reduced model of image formation introduced in section 7.1. The remaining ones (i.e. one object in the image with known reflectance and one point light source with known parameters) will be accounted for in the next point.

7.2.4 Algorithmic description of IS^{2}R

Classical SFS techniques assume changes in intensity are only due to variations in the angle $\theta = \cos^{-1}(\boldsymbol{n}(i,j) \cdot \boldsymbol{s})$ between the surface normal and the light source direction. Therefore, when a material boundary is crossed, the change in intensity due to the reflectance variation is considered as an abrupt change in the angle θ, and, therefore, a variation in the surface normal $\boldsymbol{n}$. As a result, the heights of the corresponding points appear higher or lower than they actually are. If, in addition to the reflectance changes, the image is affected by ambient lighting, the intensity values are then displaced to higher values depending on the material reflectance (equation 7.15), which also affects the subsequent height computations.

However, if the SFS algorithm is guided so that material boundaries are not crossed and ambient illumination contribution is estimated for every material and discounted from the intensity values, the

Pseudocode 7.1 Iterative growth of a region $\mathcal{R}$ of uniform reflectance in IS^{2}R. (Estimates $\widehat{L}_a$, $\widehat{L}_d$ and $\widehat{\boldsymbol{s}}$ are assumed available.)

(0.1) select a small set of image pixels as $\mathcal{R}^0$
(0.2) estimate $\widehat{\rho}_b$ from $\mathcal{R}^0$

repeat for k = 1,2,...
(k.1) estimate shape of $\mathcal{R}^{k-1}$ through SFS discounting first $(\widehat{L}_a\widehat{\rho}_b)$ and $(\widehat{L}_d\widehat{\rho}_b)$ from pixels in $\mathcal{R}^{k-1}$
(k.2) determine the set $\mathcal{C}\left(\mathcal{R}^{k-1}\right)$, consisting of all the pixels adjacent to the border of $\mathcal{R}^{k-1}$
(k.3) $\mathcal{W}\left(\mathcal{R}^{k-1}\right) \leftarrow \emptyset$
(k.4) **for every** pixel $(i,j) \in \mathcal{C}\left(\mathcal{R}^{k-1}\right)$, hypothesize its reflectance is $\widehat{\rho}_b$ and verify whether it is consistent with the image:
(k.4.1) extrapolate the shape found for $\mathcal{R}^{k-1}$ to a neighbourhood $\mathcal{N}(i,j)$ of pixel (i,j), including pixel (i,j)
(k.4.2) obtain an estimate $\widehat{\theta}(u,v)$ for pixels $(u,v) \in \mathcal{N}(i,j)$, using the shape extrapolation of (k.4.1)
(k.4.3) determine whether $(\widehat{L}_a\widehat{\rho}_b)$, $(\widehat{L}_d\widehat{\rho}_b)$ and $\widehat{\theta}(u,v)$ are consistent with $D(u,v)$ values throughout neighbourhood $\mathcal{N}(i,j)$;
if they are consistent, add pixel (i,j) to the set $\mathcal{W}\left(\mathcal{R}^{k-1}\right)$
endfor
(k.5) $\mathcal{R}^k \leftarrow \mathcal{R}^{k-1} \bigcup \mathcal{W}\left(\mathcal{R}^{k-1}\right)$
until $\mathcal{W}\left(\mathcal{R}^{k-1}\right) = \emptyset$

resultant heights will not be affected by the aforementioned unwanted behaviour. Nevertheless, to avoid crossing material boundaries, the image must be first segmented in regions of uniform reflectance so as to apply SFS to every region separately. Therefore, a circular dependence is found: a segmentation is needed to properly apply the SFS algorithm, but shape (i.e. the output of SFS) is needed to determine the material reflectance according to equation 7.16. This situation suggests a recurrent procedure where regions are iteratively grown through pixels whose reflectance has been determined to be the same as the common reflectance of pixels already belonging to the region currently being grown, until no more growth can take place.

Pseudocode 7.1 makes use of all the aforementioned to iteratively grow a region $\mathcal{R}$ through a series of steps $\mathcal{R}^0 \subseteq \mathcal{R}^1 \subseteq \cdots \subseteq \mathcal{R}^N$. The process is stopped as soon as region $\mathcal{R}$ does not grow in step k (in the pseudocode, this fact corresponds to an empty resultant $\mathcal{W}\left(\mathcal{R}^{k-1}\right)$ after sub-step $(k.4)$). Next, another region is grown, and pseudocode 7.1 applies again, and so on until no more regions can be created.

The main details of the procedure for growing a region are given below:

- First, step 0 starts the growing process by selecting a group of seed pixels [sub-step (0.1)] from where to derive an estimate $\widehat{\rho}_b$ of ρ_b [sub-step (0.2)]
- Next, step k grows region $\mathcal{R}$ from $\mathcal{R}^{k-1}$ to $\mathcal{R}^k$. To this end:
  a) The shape of $\mathcal{R}^{k-1}$, which is ensured to group only pixels coming from the same material, is computed through SFS, discounting first $(\widehat{L}_a\widehat{\rho}_b)$ and $(\widehat{L}_d\widehat{\rho}_b)$ to every pixel, so that only $\cos\widehat{\theta}(i,j)$ is passed to the SFS algorithm [sub-step $(k.1)$]:

$$\cos\widehat{\theta}(i,j) = \frac{D(i,j) - (\widehat{L}_a\widehat{\rho}_b)}{(\widehat{L}_d\widehat{\rho}_b)} \,. \tag{7.17}$$

  b) Pixels which are adjacent to the border of region $\mathcal{R}^{k-1}$ are then considered as candidates to join region $\mathcal{R}$ and are grouped into set $\mathcal{C}(\mathcal{R}^{k-1})$ [sub-step $(k.2)$, see figure 7.1].
  c) Next, every pixel $(i,j) \in \mathcal{C}(\mathcal{R}^{k-1})$ is considered for joining region $\mathcal{R}$: first, pixel (i,j) is hypothesized to have been generated by reflection over a material with reflectance $\widehat{\rho}_b$ —the same as pixels already belonging to region $\mathcal{R}$—, and then the hypothesis is checked.

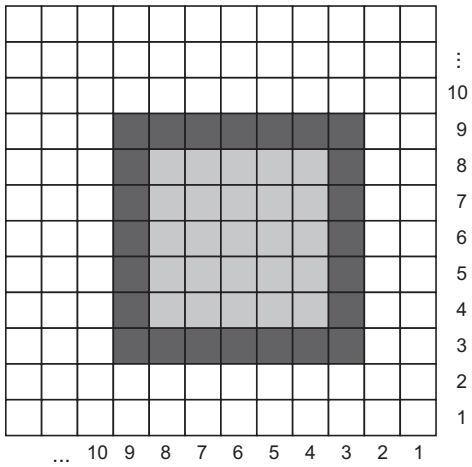


Fig. 7.1. Step (k.2) of IS^{2}R region growing procedure: pixels in light gray correspond to $\mathcal{R}^{k-1}$, while pixels in dark gray are candidates to join region $\mathcal{R}$ and therefore belong to the set $\mathcal{C}(\mathcal{R}^{k-1})$. (In the example, 8-connectivity is used, though it is not essential.)

To check the hypothesis:

i. shape information produced for region $\mathcal{R}$ is extrapolated to a neighbourhood $\mathcal{N}(i,j)$ around pixel (i,j), by means of surface fitting, so that an analytical surface is obtained [sub-step (k.4.1), see figure 7.2];

ii. with this analytical surface and knowledge about the point light source direction $\widehat{\boldsymbol{s}}$, an estimate $\widehat{\theta}(u,v)$ of $\theta(u,v)$ is obtained for every pixel $(u,v) \in \mathcal{N}(i,j)$ [sub-step (k.4.2), see figure 7.2];

iii. finally, it is evaluated whether $D(u,v)$ could have been generated from $\widehat{\theta}(u,v)$, $(\widehat{L}_a\widehat{\rho}_b)$ and $(\widehat{L}_d\widehat{\rho}_b)$ for every $(u,v) \in \mathcal{N}(i,j)$; that is to say, expression 7.18 is computed for every (u,v):

$$D(u,v) - \left[\left(\widehat{L}_a\widehat{\rho}_b\right) + \cos\widehat{\theta}(u,v)\left(\widehat{L}_d\widehat{\rho}_b\right)\right] , \tag{7.18}$$

and from the whole set of differences a decision is made about whether the image values are consistent with the hypothesis; in such a case, pixel (i,j) can join region $\mathcal{R}$ and is added to the set $\mathcal{W}(\mathcal{R}^{k-1})$ [sub-step (k.4.3)].

d) Finally, every pixel $(i,j) \in \mathcal{C}(\mathcal{R}^{k-1})$ to which reflectance $\widehat{\rho}_b$ can be assigned consistently with the image (i.e. $(i,j) \in \mathcal{W}(\mathcal{R}^{k-1}) \subseteq \mathcal{C}(\mathcal{R}^{k-1})$) is aggregated to region $\mathcal{R}$, giving rise to $\mathcal{R}^k$ [sub-step (k.5)].

It is worth noting the fact that no assumption has been made about the SFS algorithm to use, so that, theoretically, anyone proposed in the related literature can be used. In the particular implementation of this method, one of them has been selected taking into account performance and efficiency reasons, but no limitation is imposed *a priori* by the method (the SFS technique used is described in section 7.4). Likewise, the segmentation method is also independent of the surface fitting technique applied, although some techniques yield better results than others (the particular fitting technique used is described in section 7.5). Remaining details related with the different sub-steps of the recurrent procedure can be found in the following sections.

Finally, contrary to what one can initially think about the method proposed, it cannot be considered as a range segmentation algorithm[2]. In effect, in range segmentation, shape information is available for all the image points right from the beginning, while, in the proposed method, shape is available for a given image pixel only if it has been ensured it belongs to the region which is currently being grown.

[2] See [99, 112].

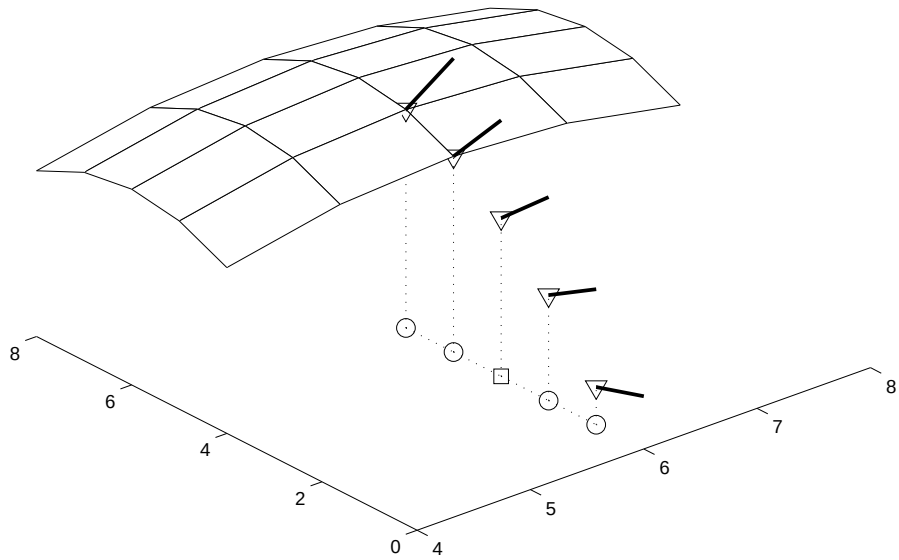


Fig. 7.2. Sub-steps (k.4.1) and (k.4.2) of IS^{2}R region growing procedure. After having applied SFS to region $\mathcal{R}^{k-1}$ (figure 7.1, pixels in light gray), pixel (3,6), labelled with a square $\square$, is considered to be added to region $\mathcal{R}$. In the example, its neighbourhood (set $\mathcal{N}(3,6)$) includes itself and pixels labelled with a circle ($\bigcirc$). Points labelled with a triangle (∇) have been extrapolated using the analytical surface resulting from fitting the shape information obtained from the shading. The thicker lines represent surface normal vectors from where angle θ would be estimated

7.3 Initialization of Regions

In order for a region $\mathcal{R}$ to grow through pseudocode 7.1, it is necessary to have a set of pixels, say $\mathcal{R}^0$, acting as a sort of seed from which to obtain shape information which allows the growing process to start. Since an SFS algorithm is going to provide such shape information and this sort of algorithms essentially solve the brightness equation, pseudocode 7.1 assumes estimates for L_a, L_d and $\boldsymbol{s}$ are available, as well as for ρ_b, which would be deduced from $\mathcal{R}^0$. Furthermore, observe that ρ_b, L_a and L_d are not really needed separately for applying pseudocode 7.1 nor for the SFS algorithm. What is truly necessary are the two composite reflectances $D_a = (L_a\rho_b)$ and $D_b = (L_d\rho_b)$. This fact simplifies enormously the problem, since multiple values of ρ_b, L_a and L_d give rise to the same D_a and D_b values.

Remember now from chapter 5 that equation 7.16 predicts that the pairs $(D, m_b = \boldsymbol{n}\cdot\boldsymbol{s})$ lie on a straight line whose parameters are D_a and D_b. If enough pairs (D, m_b) were available for every seed $\mathcal{R}^0$ of every image region $\mathcal{R}$, then a least-squares-based procedure could provide the composite reflectances. Unfortunately, this is not typically the case and an alternative method must be looked for. Although several methods of composite reflectance estimation using just the information contained in a single image can be found in the physics-based vision literature [140, 141, 226, 311][3], however:

(1) none of these methods was initially devised to incorporate ambient lighting, and,

[3] So far, several techniques have been published with regard to the estimation of image reflectance values from a single image. The final goal of such techniques was mostly to use the estimates in shape from shading calculations. As it will be seen, none of them take into account ambient lighting, so that their image formation model is reduced to equation 7.19:

$$D(i,j) = (\boldsymbol{n}(i,j)\cdot\boldsymbol{s})\,(L_d\rho_b(i,j))\ . \tag{7.19}$$

(2) because of their respective formulations, only the method by Pentland [226] could be adapted to tolerate ambient lighting since, due to being mostly based on first and second image derivatives, the ambient lighting term vanishes during the derivation of the main results.

Besides, even for Pentland's method, the estimates of composite reflectances obtained were not accurate enough so as to be used within IS^{2}R. A method based on the composite reflectances of a calibration object was devised instead and is presented in the following.

Let us assume now that any of the lighting parameters estimation methods of chapter 5 has been applied using a calibration object whose body reflectance is $\rho_{b,C}$, so that estimates of the respective composites reflectances $\widehat{D}_{a,C} = (\widehat{L}_a\widehat{\rho}_{b,C})$ and $\widehat{D}_{b,C} = (\widehat{L}_d\widehat{\rho}_{b,C})$ are available[4].

Let us also assume that for region $\mathcal{R}$ a *singular point*, i.e. a point such that the surface normal coincides with the light source direction and, consequently, $m_b = 1$, has been located. For the corresponding pixel, $D_{sp} = (L_a\rho_b) + m_b(L_d\rho_b) = (L_a\rho_b) + (L_d\rho_b)$, which, according to the discussion of section 7.1, is $D_{sp} \approx (L_a + L_d)\rho_b$ and:

$$\frac{D_{sp}}{(L_a\rho_{b,C}) + (L_d\rho_{b,C})} = \frac{D_{sp}}{(L_a + L_d)\rho_{b,C}} = \frac{(L_a + L_d)\rho_b}{(L_a + L_d)\rho_{b,C}} = \frac{\rho_b}{\rho_C} = \widetilde{\rho}_b\,, \tag{7.20}$$

where $\widetilde{\rho}_b$ would be a sort of reflectance for $\mathcal{R}$ relative to $\rho_{b,C}$.

Now, $\widehat{D}_{a,C}$ and $\widehat{D}_{b,C}$ can be used as it is indicated below:

$$\widehat{D}_{a,C}\widetilde{\rho}_b = L_a\rho_{b,C}\frac{\rho_b}{\rho_{b,C}} = L_a\rho_b \tag{7.21}$$

$$\widehat{D}_{b,C}\widetilde{\rho}_b = L_d\rho_{b,C}\frac{\rho_b}{\rho_{b,C}} = L_d\rho_b \tag{7.22}$$

which are, precisely, the values to discount to image intensities for region $\mathcal{R}$ before applying the SFS algorithm. Therefore, it is at singular points where $(L_a\rho_b)$ and $(L_d\rho_b)$ can be estimated without previous knowledge about the shape of the scene, and, consequently, it is around singular points where seeds $\mathcal{R}^0$ must be created[5].

Regarding the composition of $\mathcal{R}^0$, it has been chosen to consist of either pixels not differing more than a fixed number of intensity levels with the singular point intensity value, and being within a

On the one hand, two statistical approaches for estimating the composite reflectance term $(L_d\rho_b)$ were reported by Lee and Rosenfeld [141] and Zheng and Chellappa [311]. Both methods were based on the distribution of surface normal vectors all over the image. The main difference between them precisely lie in their assumptions about this distribution. While the former assumed locally spherical surface patches, the latter considered them as locally flat. Using these premises both derived the needed distribution function. The two methods, however, tried to estimate a constant reflectance value for all the pixels in the image, independently of the number of existing materials.

A different method was proposed by Lee and Rosenfeld [140]. They computed the composite reflectance value for each pixel P using only the intensity information of P and two pixels on opposite sides of P, provided they did not come from coplanar points in the scene. The estimated values were then collected in a histogram for the whole image. The peaks of such histogram corresponded then to the different materials in the scene.

Pentland also proved that, for umbilical surface points (i.e. points with equal principal surface curvatures), first and second image derivatives allowed estimating the composite reflectance for those points and provided the corresponding formulation [226].

[4] Remember that the lighting estimation methods proposed in chapter 5 provide estimations of the composite reflectances $\widehat{D}_{a,C} = (\widehat{L}_a\widehat{\rho}_{b,C})$ and $\widehat{D}_{b,C} = (\widehat{L}_d\widehat{\rho}_{b,C})$, which, if the calibration object has been adequately chosen (i.e. $\rho_{b,C} = 1$), allow obtaining the estimates $\widehat{L}_a = \widehat{D}_{a,C}$ and $\widehat{L}_d = \widehat{D}_{b,C}$ of, respectively, L_a and L_d.

[5] As a further result, observe that, from a theoretical point of view, $\widetilde{\rho}_b$ is invariant against illumination changes, so that it could be used for object recognition purposes. In effect, if light strengths change from session (1) to session (2), whatever $L_a^{(2)}$ and $L_d^{(2)}$ are, then:

$$\widetilde{\rho}_b^{\,(2)} = \frac{D_{sp}^{(2)}}{(L_a^{(2)}\rho_{b,C}) + (L_d^{(2)}\rho_{b,C})} = \frac{L_a^{(2)}\rho_b + L_d^{(2)}\rho_b}{L_a^{(2)}\rho_{b,C} + L_d^{(2)}\rho_{b,C}} = \frac{\rho_b}{\rho_{b,C}} = \widetilde{\rho}_b^{\,(1)}\,. \tag{7.23}$$

square area centered around the singular point. Parameters $\Delta_{\mathcal{R}^0}$ and $R_{\mathcal{R}^0}$ determine, respectively, the difference in intensity with the singular point and the side of the square area.

To grow regions around singular points, the latter must be located first. As $\theta = 0$ for those points, they correspond to pixels of local maximum brightness. Therefore, looking for pixels in the image which have the largest intensity in a neighbourhood seems to be enough[6]. This is a low-cost strategy for locating singular points, which has shown to work well along the different experiments performed[7]. Note the previous process can be applied to every pixel in the image a priori, producing a *singular points map* which can be consulted when looking for singular points. In the algorithm, the size of the 2D neighbourhood is parameterized through the length of the side of a square area whose center is the current image point (parameter R_{sp}).

The fact that regions grow around singular points implies that, at the beginning, the growth of regions is only along decreasing intensity paths. However, after some iterations, it is possible that increasing intensity paths have to be explored (i.e. think of the case of a multiple singular point object consisting of one material). Although, theoretically, pseudocode 7.1 is insensitive to the sort of intensity path followed, experimental results have shown that both the shape extrapolation and the consistency test behave more reliably if only intensity decreasing paths are used. As a consequence, however, objects with multiple singular points are over-segmented. How to deal with it, together with other segmentation anomalies, will be discussed in section 7.7.

To finish, regions are grown from the darkest to the brightest one (i.e. from the region with the darkest singular point to the region with the brightest singular point). In this way, lower-reflectance regions are grown first and, accordingly, the corresponding object pixels are labelled first, so that adjacent higher-reflectance regions will not even consider them for growing the current region through decreasing intensity paths.

7.4 Region Shape Estimation

Recovering the correct shape of an object from the shading of the corresponding image has proven to be a difficult mathematical problem, even under the standard idealizing assumptions of Lambertian reflectance, one light source and orthographic projection. Under these conditions, the brightness at a given point fixes only the projection of the surface normal onto the incident light vector and, hence, a unique normal vector cannot be easily associated to each point by means of just the brightness value. Therefore, some constraints must be imposed to cut down the huge number of solutions consistent with the image and to strongly couple normal vectors across the surface.

Since the formulation of the shape from shading problem by Horn, several approaches and constraining strategies have been proposed in order to ensure the accuracy of the estimated surface [303]: while *local approaches* impose restrictions on the local shape of surfaces, which give rise to closed-form solutions (for instance, [141, 224] assume surfaces are locally spherical), *global minimization methods* formulate the problem as a minimization of a functional regarding the relation between the shape and the image, and *propagation methods* develop the object surface propagating shape information from points with known height or surface orientation[8].

Global minimization and propagation approaches often solve the problem in an iterative way, so that the convergence of the algorithm and the convergence to the correct solution are issues of prime importance. As the former issue depends on the algorithm itself and on the particular brightness pattern (i.e. is there a surface giving rise to the image?), the latter mostly depends on the initial values given to the algorithm. In order to enforce proper convergence, researchers imposed regularization constraints to guarantee physically plausible surfaces (mostly smoothing and integrability), while shape information

[6] For instance, in [193], where it is also needed to locate singular points, the image is assumed to be generated according to the brightness equation and, thus, locating singular points is just a matter of thresholding the image by a value close to 1.

[7] Another property of singular points is that the image surface at the corresponding pixel has a normal vector equal to $(0, 0, -1)$, because it is a local extreme.

[8] Zhang et al. [303] distinguish a further class, which they call *linear approaches*, as they linearize the reflectance map.

at occluding boundaries was used as initial values (many of the related papers are collected in [103]). Later, Oliensis [190, 191] proved that the complete shape information at an occluding boundary does not well-determine the surface reconstruction, but height at singular points corresponding to local mimima or maxima of height (Local Minimum/Maximum Singular Point or LMSP, according to the terminology of Oliensis and Dupuis [193]) does without the need of any regularization[9]; Oliensis and Dupuis then proposed a propagation strategy based on the singular points constraint and optimal control theory [192]. Afterwards, on the basis of Oliensis' idea, Bichsel and Pentland presented a simpler propagation approach based on a minimum downhill principle [17], while Rouy and Tourin [242] proposed a solution based on viscosity solutions to partial differential equations.

Oliensis not only proved that an LMSP uniquely determines the surface, but he also showed that this happens just within some image neighbourhood. As a consequence, if height is provided for every LMSP of the surface (e.g. from stereo), a correct global reconstruction can be achieved[10]. Otherwise, assigning a height to only one singular point of the surface, let us say 0, provides a correct reconstruction but just in the above mentioned neighbourhood, called the *domain of attraction* of the singular point by Oliensis and Dupuis. In those cases of lack of information about singular point heights, a global algorithm is needed (see [23, 117, 193]), in contrast with the previous algorithms, which would be considered local since only locally is the surface correctly reconstructed. Those algorithms use at their heart a local algorithm and then patch the local solutions together by first determining the relative height of LMSP whose domains of attraction touch.

Therefore, at first sight, it seems that either height information should be provided at LMSP or a global shape from shading algorithm should be applied in sub-step (k.1) of the recurrent procedure of pseudocode 7.1. However, to obtain the segmentation of the image using the above procedure, it is just necessary that the brightness equation is fulfilled by the surface recovered, irrespectively of its resemblance to the original surface. In effect, changes in reflectance are detected by means of the consistency test, which, in turn, checks whether assigning the reflectance of the currently growing region to the considered pixel is coherent with the image on the basis of the angle θ estimated through extrapolation of the recovered surface. Therefore, if the recovered surface satisfies the brightness equation and the extrapolation is correct, just the reflectance determines the result of the consistency test. Furthermore, take into account that, since regions growth is restricted to decreasing intensity paths around singular points, the resultant regions will mostly agree with the domains of attraction of Oliensis and Dupuis [193], where shape is correctly reconstructed. (Note the reconstruction would not be correct for saddle singular points, but the brightness equation would indeed be satisfied.)

Among the different SFS algorithms cited above, Bichsel and Pentland proposal [17] (B&P algorithm from now on) has shown to satisfy the brightness equation and has proven to be a low-cost way of estimating shape, which is crucial in this application due to the huge number of times the SFS procedure is called (experiments performed have shown that between 5 and 10 iterations are sufficient in most cases)[11]. Essentially, the B&P algorithm uses a variant of the method of characteristic strips to obtain height maps of the scene. The known unstability of the original method by Horn [100] is overcome by an iterative scheme where the local surface height is updated according to a minimum downhill principle.

In the B&P algorithm, the heights of local maxima singular points have to be given as an initial height map in order to obtain a correct reconstruction. If those heights are truly available, the segmentation algorithm can be provided with them. Otherwise, an initial height 0 can be used for every singular point. Due to the way the B&P algorithm derives the surface, using such a height for a local maximum gives rise to a locally correct surface, while for a local minimum, a convex solution

[9] Maximum brightness is attained at singular points. Therefore, singular points correspond to local minima of height, local maxima of height or saddle points with respect to the light source direction. However, only LMSP determine the surface reconstruction, although up to a concave-convex ambiguity (see, for instance, [226]).

[10] In fact, height information for every non-saddle singular point is not needed, since some algorithms propagate shape information from local maximum singular points [17], while others do it from local minimum singular points [118]. Some algorithms even propose two versions, one for each case [192].

[11] Another low-cost solution for estimating shape from shading which could have been adopted is the algorithm by Kimmel and Sethian [118].

is obtained instead of the expected concave surface. As for saddle points, the recovered surface also tends to be concave-shaped. In all cases, however, the brightness equation is satisfied.

As for the particular implementation of the B&P algorithm within the segmentation strategy, it has been observed that, if this algorithm is restricted to work with a light source at infinity and over the optical axis (i.e. the light source orientation vector is $(0, 0, -1)^T$ according to the axis notation depicted in figure 4.1), several benefits can be obtained: first, the computations for estimating shape are considerably simplified; secondly, Bichsel and Pentland suggest in the paper an image rotation according to the tilt of $\boldsymbol{s}$ before applying the algorithm in order to improve results, which now is no longer needed; finally, a better behaviour of the B&P algorithm has been observed as for convergence under side lighting.

As a consequence of using a vertical light source, at the end of the reconstruction, the surface normal vectors of the estimated shape appear rotated according to the coordinate transformation from vector $(0, 0, -1)^T$ to the true light source direction. If R is the rotation matrix which transforms the light source direction from $\boldsymbol{s}$ to $(0, 0, -1)^T$, the brightness equation can also be put as indicated in equation 7.24, since $R^T\, R$ is the identity matrix because R is a unit matrix:

$$E = (\boldsymbol{n} \cdot \boldsymbol{s}) = \left(\boldsymbol{n}^T\, \boldsymbol{s}\right) = \left(\boldsymbol{n}^T\, R^T\, R\, \boldsymbol{s}\right) = \left[(R\,\boldsymbol{n})^T\ (R\,\boldsymbol{s})\right] = \left[(R\,\boldsymbol{n})^T \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}\right] \tag{7.24}$$

Thereby, since the recovered surface satisfies the brightness equation for $\boldsymbol{s}' = (0, 0, -1)^T$, then the surface normal vectors of the recovered surface are $\boldsymbol{n}' = R\,\boldsymbol{n}$. As the surface normal is used in the surface reconstruction to determine the final height values z, the recovered surface points (x, y, z) are rotated accordingly [118][12]. Note that the rotation back is not needed within the segmentation algorithm since the shape extrapolation and the consistency test can both be equally done in the light coordinate frame.

7.5 Shape Extrapolation

The shape extrapolation sub-step is carried out through an analytical surface which is fitted to sets of image points whose heights are available. In this way, θ can be estimated for every image pixel involved in the consistency test. Clearly, the sort of surface model used in the fitting determines the quality of the extrapolation. At first, it seems the surface model would depend on the sort of local surface shapes which are expected in the scene. However, due to the peculiar way the shape estimation is carried out, recovered surfaces, although satisfying the brightness equation, need not be similar to local shapes in the scene. To achieve such a performance, heights at singular points must be given, and only LMSP should be used for growing regions (i.e. saddle singular points should be avoided). Therefore, in principle, it is not clear the relationship between the shapes in the scene and the surface model.

Another criterion for choosing the surface model is the time involved in the fitting. At this point, it is worth noting the fitting procedure can be executed thousands of times for a typical 200×200 image (at a minimum, almost once per pixel). Therefore, a surface model for which a low-cost fitting algorithm exists would be an interesting option.

Taking into account the locality of the fitting and the computational cost limitation, the family of quadric surfaces has been considered a compromise solution to adopt at this point. On the one

[12] Matrix R relating points in the world coordinates frame and the light coordinate frame is given by:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = R \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cos\tau & -\sin\tau & 0 \\ \sin\tau & \cos\tau & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \cos\sigma & 0 & -\sin\sigma \\ 0 & 1 & 0 \\ \sin\sigma & 0 & \cos\sigma \end{pmatrix} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} \tag{7.25}$$

where $\sigma = \cos^{-1} s_z$ and $\tau = \tan^{-1} \frac{s_y}{s_x}$ are, respectively, the slant and tilt angles of the light source direction $\boldsymbol{s} = (s_x, s_y, s_z)$. In the above expression, primes denote the light coordinate frame.

hand, patches of quadric curves and surfaces have found widespread use in modeling and recognition of objects of interest in computer vision. On the other hand, the fitting technique by Taubin [271] provides an approximation to the orthogonal distance for implicit surfaces at the cost of solving an eigenvalue problem of the same dimension as the number of parameters of the surface. Surely, more complex analytical surfaces (e.g. bivariate polynomials of higher order, superquadrics, surface splines, etc...) could yield better results, but at the expense of a higher computational cost. Finally, the fact that only decreasing intensity paths are allowed when growing a region avoids strange surfaces (i.e. as only one singular point is included in the region, no more than one local minimum, local maximum or saddle point can be found in the recovered surface), making thus quadric surfaces more suitable for extrapolation purposes.

The shape extrapolation sub-step is executed in different ways throughout the analysis of the image. Once the singular points map has been obtained, regions around singular points start being grown using all the image points belonging to the region and their respective heights; a global fitting strategy is thus used. When $\mathcal{R}^k$ can no longer be fitted by a quadric because the fitting error is too high, then its growth is stopped, and the next singular point is considered. Once all singular points have been processed, region-growing is resumed for every region, but now using a local fitting strategy: for a given region $\mathcal{R}^k$, a sub-region of $\mathcal{R}^k$ around the pixel considered to be added to $\mathcal{R}^k$, and the set of respective heights, are used in the fitting. The whole shape extrapolation strategy intends, thus, to improve shape extrapolation at the beginning, when regions are incipient. The length of the side of the square area used in the local fitting stage is another parameter of the method (R_{lf}).

In order to detect when the surface model is not compatible with the object surface, two tests must be passed by the quadric resulting from a fitting:

(1) On the one hand, it is required, for the fitted points, the fulfillment of the following relationship between the maximum square Taubin distance $\tau_M(S)^2$ and the mean square Taubin distance $\bar{\tau}(S)^2$:

$$\tau_M(S)^2 \leq \epsilon_f \bar{\tau}(S)^2 \tag{7.26}$$

where $\epsilon_f > 1$ is a parameter, $S = 0$ is the implicit surface resulting from the fitting and

$$\tau(S, x_i) = \frac{S(x_i)}{\nabla S(x_i)} \tag{7.27}$$

$$\bar{\tau}(S)^2 = \frac{1}{n} \sum_{i=1}^{n} \tau(S, x_i)^2 \tag{7.28}$$

$$\tau_M(S)^2 = \max_{1 \leq i \leq n} \tau(S, x_i)^2 \tag{7.29}$$

where $S(x_i)$ is the algebraic distance of point x_i to the surface S (i.e. the deviation of $S(x_i)$ from 0). This is the adaptation of the second test used in Taubin's planar curves segmentation algorithm [271].

(2) On the other hand, as a quadric generally shows two leafs, a set of points is considered well approximated by the surface if a single leaf fits all of them.

To finish this section, the validity of the shape extrapolation is discussed next. If the surface is smooth at the point where the extrapolation is performed and the surface approximates well the points already belonging to the region, the extrapolation error typically depends on the distance (over the image plane) between the extrapolated points and the points from which the extrapolation has been computed; generally speaking, the nearer the extrapolated points are (over the image plane), the lower the error. Therefore, if the consistency test is applied in a small neighbourhood of the pixel considered to join the currently grown region, the extrapolation error is expected to be low.

7.6 Consistency Test

The consistency test verifies whether the hypothesis made in sub-step (k.4) is coherent with the image. That is to say, whether using the extrapolated shape and the reflectance of the region currently being

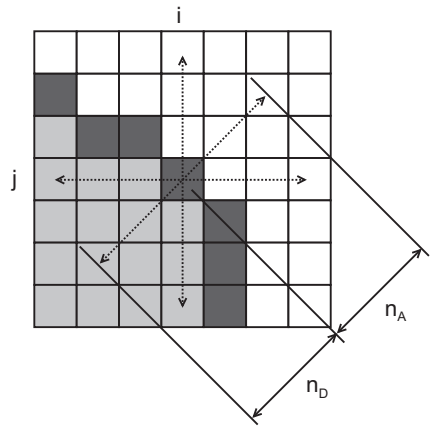


Fig. 7.3. 1D neighbourhoods used in the consistency test. Pixels within the region (light gray pixels) are used to compute $\bar{D}$ for every direction (equation 7.31). Pixels outside the region (white pixels) and pixel (i, j) are all submitted to the consistency test (equation 7.31).

grown, and, thus, the corresponding composite reflectances $(L_a\rho_b)$ and $(L_d\rho_b)$, the intensity values predicted by the equation defining the image formation model (equation 7.15) are similar to the ones found in the image. That is to say, for pixel (i, j), the consistency test would be formulated as:

$$| D(i, j) - [(L_a\rho_b) + \cos\theta(i, j)(L_d\rho_b)] | \approx 0 \tag{7.30}$$

where $\theta(i, j)$ would be determined from the shape extrapolation at pixel (i, j).

The performance of the consistency test proposed in equation 7.30 is directly related with, in particular, the quality of the estimate of $\theta(i, j)$. Admittedly, several sources of noise can condition the reliability of the consistency test formulated in equation 7.30. For a start, the imperfections of the sensor itself first and the subsequent discretization next give rise to deviations between the true irradiance at the sensor cell and the intensity value found in the image array. Next, the SFS step, when reconstructing a surface compatible with those intensity values, can be affected not only by the noise of the image but also by the proficiency of the SFS algorithm in producing a surface fulfilling the brightness equation. Furthermore, $\theta(i, j)$ is computed from a fitting stage at which a surface model has been imposed and where, again, the performance of the algorithm chosen for this step also affects the final value for $\theta(i, j)$. Finally, the camera aliasing makes intensity edges corresponding to scene reflectance transitions appear rounded in the image, what can significantly reduce the gap expected between predicted and actual intensity values at those locations.

Because of all this, the consistency test is modified as it is indicated in the following in order to improve its performance:

(1) On the one hand, in order to counteract the effects of camera aliasing, several pixels in the nearness of pixel (i, j) are tested instead of just pixel (i, j), so that pixel (i, j) is rejected if any of them does not pass the test. In order to improve the localization of region borders, several 1D neighbourhoods of pixel (i, j) are used, taking as orientations for the 1D neighbourhoods the region outward directions at pixel (i, j) (see figure 7.3). The test must be satisfied in at least one of those directions. The number of pixels ahead of pixel (i, j) which are used, n_A, is a parameter.

(2) On the other hand, instead of comparing the difference between actual and predicted intensity values with just 0, a certain deviation between predicted and actual intensity values is accepted. In the new form of the consistency test, the deviation tolerated is expressed by means of a first-order model incorporating either the goodness of the estimation of θ and the uncertainty of the intensity values involved, as it is expressed in equation 7.31:

$$| D(i', j') - [(L_a\rho_b) + \cos\theta(i', j')(L_d\rho_b)] | \leq \epsilon_1\overline{\Delta} + \epsilon_2\delta(D(i', j')) + \epsilon_3 \,. \tag{7.31}$$

In equation 7.31, i' and j' are the image co-ordinates along the outward part of the 1D neighbourhood considered, $\overline{\Delta}$ is the deviation average measured along the inward part of the same 1D

neighbourhood, $\delta(D(i', j'))$ is the uncertainty of the intensity value $D(i', j')$, and ϵ_1, ϵ_2 and ϵ_3 are weights for balancing the importance given to the different sources of information regarding the deviations between predicted and actual intensity values. The number of pixels used to compute $\overline{\Delta}$, n_D (see figure 7.3), is also a parameter.

7.7 Discussion on Segmentation Anomalies

Due to the way how IS^{2}R works, some anomalies can appear in the resultant segmentation:

(1) As has been discussed before, the surface model imposed by the surface fitting stage can lead to higher fitting errors if the real surface is not locally compatible with the analytical surface. In this case, the extrapolation is not performed and the point is not considered for joining the region, so that the growth of the region towards a certain direction can stop without being at an object boundary. As a consequence, a region in the image corresponding to the same material can be split up into several regions.
On the other hand, due to the fact that only non-ascending intensity paths are used for growing regions, objects with several singular points are also divided into as many regions as singular points possess the object.
Both situations can lead to an over-segmentation of the image caused by one or more sets of adjacent regions of similar reflectance. A further stage of region merging should thus be considered for adoption in those cases.

(2) A segmentation can be incomplete when either the scene contains objects that do not show singular points (e.g. due to occlusions), or if, for a certain object, some pixels are labelled because they are in an intensity descending path from a singular point, while others are not because there is no singular point leading to them through an intensity decreasing path.
In case this happens, the segmentation can be completed by following a similar way of action as the one suggested in the consistency test section: (1) those non-labelled pixels adjacent to labelled pixels are considered candidates to grow new regions; (2) among them, the non-labelled lowest intensity one is used to grow a region using decreasing intensity paths; (3) as no reflectance can be estimated for the region, since no singular point can be found, no predictions are made and the region just grows towards lower or identical intensity pixels until no more growth can take place; (4) the process is repeated for the following non-labelled higher intensity pixel until all the pixels have been labelled.
Note that the previously outlined strategy does not guarantee that the new regions satisfy the reflectance uniformity assumption of SFS algorithms. Figure 7.4 presents a 1D case in which this requirement would likely be violated. Regions X and Y are regions grown around singular points so that they follow the normal procedure. However, at points A and B the consistency test stops the growth and, therefore, the pixels between A and B leave unlabelled. In the posterior stage, A would be selected first to grow a new region, but the growth would not stop until point E was reached. The same would apply from point B. At points C and D, no prediction could be done, since no reflectance was known for the newly created regions, so that the growth could not be stopped, and regions V and W finally would result.
Some of the new regions can be added to the previous ones (i.e. the regions for which the reflectance is known) using continuity criteria along the borders of adjacent regions. In effect, if the image surface at the border between two regions is continuous then no change in reflectance can take place and therefore the newly created region can inherit the reflectance of the former region.
As a result of the previous process, several regions would be merged and a new partition would result. In the end, those regions for which the reflectance was known would be suitable for estimating shape from image shading.

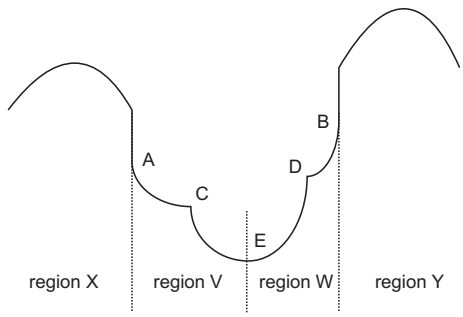


Fig. 7.4. 1D example of a segmentation where a region can violate the reflectance-uniformity requirement.

7.8 Extension of the Model of Image Formation

This section discusses how IS^{2}R can be enhanced so as to be able to segment images coming from scenes not satisfying all the assumptions made in the simple image formation model of section 7.1. The different extensions considered are discussed in the following sections.

7.8.1 Segmentation of colour images

In order for IS^{2}R to be able to segment colour images, some of the steps of pseudocode 7.1 must be generalized to work with several colour planes instead of an only gray-level plane. The corresponding modifications are described as follows:

1. Now, estimates $(\widehat{L}_a\widehat{\rho}_{b,C})^k$ and $(\widehat{L}_d\widehat{\rho}_{b,C})^k$, $k \in \{R, G, B\}$, are required, what can be obtained from the lighting estimation algorithms of chapter 5. Accordingly, $(L_a\rho_b)^k$ and $(L_d\rho_b)^k$ must be determined from singular points, using equations 7.21 and 7.22, and removed from the respective pixels before estimating shape.
2. As for the localization of singular points, colour pixel values are transformed to gray-level in order to have an only image surface from where to determine local maxima of brightness. The conversion follows the NTSC video standard, which determines how a colour television signal is rendered on a black and white television:
$$D = 0.2989\, D^R + 0.5870\, D^G + 0.1140\, D^B\,, \tag{7.32}$$
where D represents the gray-level value.
3. The order in which regions are grown agrees with the modulus of the colour of the corresponding singular point $\boldsymbol{D_{sp}} = (D^R_{sp}, D^G_{sp}, D^B_{sp})$: the lower $\|\boldsymbol{D_{sp}}\|$ is, the sooner the singular point is visited.
4. In general, SFS algorithms, and in particular the B&P algorithm, require an only array of values corresponding to the brightness equation, which now will have to be generated from the three colour planes usually available from colour cameras. Since the body geometrical term m_b affects all the colour channels in the same way, any of the colour planes could be used from a theoretical point of view. However, since images contain discrete values of intensity, the colour plane for which the reflectance is the largest turns out to be the best election. In effect, consider a region which, for the red channel, $(L_d\rho_b)^R = 255$ and, for the green channel, $(L_d\rho_b)^G = 10$, and $(L_a\rho_b)^R = 0$ and $(L_a\rho_b)^G = 0$ for simplicity. Then, the array of values used by the SFS algorithm for the red colour channel will be constituted by elements of the set:
$$\frac{D^R - (L_a\rho_b)^R}{(L_d\rho_b)^R} = \{0/255, 1/255, 2/255, 3/255, \ldots, 255/255\}\,, \tag{7.33}$$
while in the case of the green colour channel this set will consist of just:
$$\frac{D^G - (L_a\rho_b)^G}{(L_d\rho_b)^G} = \{0/5, 1/5, 2/5, 3/5, 4/5, 5/5\}\,. \tag{7.34}$$
Clearly, in the first case, the accuracy of the shape reconstruction must necessarily be better.

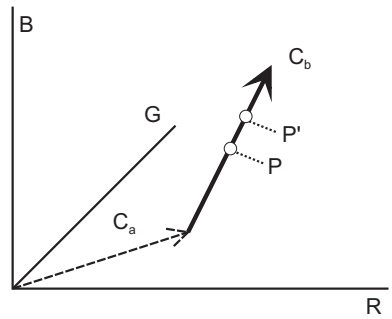


Fig. 7.5. Illustration of the concept of intensity descending path in colour images.

5. The consistency test must now take into account that three intensity levels per pixel are available. Therefore, to incorporate a new pixel to the currently grown region, the difference between predicted and actual intensity values for all three colour channels must be required to be in accordance with the discrepancy model of equation 7.31.
6. Finally, the concept of descending intensity path introduced for gray-level images must be adapted to colour images. In a noiseless colour image and within a uniformly coloured object, pixel values found along paths departing from a singular point lie in colour space in the straight line defined by vector $\boldsymbol{C_b} = ((L_d\rho_b)^R, (L_d\rho_b)^G, (L_d\rho_b)^B)^T$ and point $\boldsymbol{C_a} = ((L_a\rho_b)^R, (L_a\rho_b)^G, (L_a\rho_b)^B)^T$. Therefore, in this context, a pixel P with values $(D^R, D^G, D^B)^T$ can be said "to be below" a pixel P' with values $(D^{R\prime}, D^{G\prime}, D^{B\prime})^T$ along an intensity descending path, if the projection of the former onto the straight line defined by vector $\boldsymbol{C_b}$ and point $\boldsymbol{C_a}$ is lower than the projection of the latter (see figure 7.5). In other words:

$$(\boldsymbol{P} - \boldsymbol{C_a})^T \frac{\boldsymbol{C_b}}{\|\boldsymbol{C_b}\|} < (\boldsymbol{P'} - \boldsymbol{C_a})^T \frac{\boldsymbol{C_b}}{\|\boldsymbol{C_b}\|} \tag{7.35}$$

Besides, pixel P is required to be close to the mentioned straight line, or, in other words, the orthogonal distance in colour space from pixel P to the straight line defined by vector $\boldsymbol{C_b}$ and point $\boldsymbol{C_a}$ is bounded.

7.8.2 Tolerance to shadows

IS^{2}R attempts to grow regions around all the singular points which have been detected in the image. Therefore, if a singular point is found inside a shadow, a region, let us say R, will be created from the intensity descending paths departing from it. Accordingly, IS^{2}R will assign composite reflectances $\boldsymbol{C_a}$ and $\boldsymbol{C_b}$ to region R using equations 7.21 and 7.22. Remember that, as was already commented at the end of section 2.3.4, pixels from shadows have $m_b = 0$ and, therefore, lie exactly at $\boldsymbol{C_a}'$, the composite reflectance of the main region, let us say R'. Therefore, IS^{2}R erroneously splits $\boldsymbol{D_{sp}} = (D^R_{sp}, D^G_{sp}, D^B_{sp})$ between $\boldsymbol{C_a}$ and $\boldsymbol{C_b}$ through equations 7.21 and 7.22, since, a priori, it does not have knowledge about whether R is a shadow or not. However, $\boldsymbol{D_{sp}} = \boldsymbol{C_a} + \boldsymbol{C_b} = \boldsymbol{C_a}'$, so that, if adjacent regions R and R' are discovered to have $\boldsymbol{C_a} + \boldsymbol{C_b} = \boldsymbol{C_a}'$, then R can be inferred to be a shadow of R' and, therefore, merged.

Clearly, this is a simplified model for shadows which considers them as areas of low constant brightness. This is particularly true when illumination is assumed to come from a point, but not for general lighting distributions, which usually give rise to areas of penumbra next to full shadows [83, p. 92].

7.8.3 About other extensions

The incorporation of further extensions to the image formation model greatly depends on the capabilities of the part of the segmentation algorithm devoted to computing shape information, which in the

case of IS^{2}R refers to the SFS method embedded in the algorithm. Although most SFS methods require the satisfaction of all the model constraints enumerated in section 7.1, some authors have proposed newer algorithms able to cope with more general models of image formation. These are the cases of, among others, Lee and Kuo [146] or Samaras and Metaxas [247], who mainly allow non-Lambertian surfaces in the scenes so that, from a theoretical point of view, specularities are tolerated by the proposed SFS methods. Unfortunately, according to the results published in the respective papers, the accuracy of the shape recovered is clearly not adequate so as to be used within IS^{2}R, not even for simple shapes. In fact, along the experiments, Lee and Kuo have to resort to photometric stereo to improve the quality of the shape recovered and, even under the enhanced operating conditions, the improvement achieved is acceptable only in some cases and for very simple shapes. Some other authors have proposed the removal of specularities from images before proceeding to estimating shape information, what could be an alternative for IS^{2}R (see [45, 179, 223, 249, 250, 268] among others).

There are however other enhancements of the image formation model which IS^{2}R could easily incorporate. This is, for instance, the case of the improvements in the lighting distribution model proposed by Tian et al. [277–279]. More precisely, they determine reflectance maps for scenes illuminated by multiple rectangular light sources and for scenes involving spherical and cylindrical light sources. Besides, they use the B&P SFS algorithm to test the new reflectance maps.

7.9 Experimental Results

Results for IS^{2}R corresponding to several segmentation and surface reconstruction experiments with synthetic and real images under different lighting conditions and different reflectance combinations are provided in this section. Although many of the experiments have been designed so as to satisfy scrupulously the reduced image formation model expressed through equation 7.15, some others will involve images incorporating the extensions to the image formation model of section 7.8. As will be seen, among the different parameters of IS^{2}R, some of them remained fixed throughout almost all the experiments while others needed slight variations. In general, the former were set to values that seemed reasonable, while the values for the latter are the ones giving rise to the best segmentations. Table 7.1 summarizes all the parameters of IS^{2}R.

As suggested in section 7.7, a further stage of region merging is incorporated into IS^{2}R to alleviate the incidence of over-segmentation in cases where objects of the scene exhibit several singular points. To this end a very simple region merging algorithm based on reflectance similarity has been adopted along the experiments of the following sections. Furthermore, isolated pixels that result unlabelled because they did not pass the consistency test on any occasion are finally added to the adjacent region closer in colour space. This last step would correspond to the results post-processing box of figures 4.2 and 4.7.

As will be seen later, IS^{2}R was faced against synthetic images, first without noise and next with noise. In this way, it was intended to show the performance of the segmentation algorithm under the most controlled conditions. When processing noiseless images, intensity uncertainties were removed from expression 7.31 when performing the consistency tests (i.e. $\epsilon_2 = 0$). They were, however, used in the experiments with noisy images. Besides, when gray-level images were involved in the experiments, intensity uncertainties were recalculated using the coefficients appearing in equation 7.32[13]. That is to say:

$$\delta(D) = 0.2989\,\delta(D^R) + 0.5870\,\delta(D^G) + 0.1140\,\delta(D^B) \tag{7.36}$$

Furthermore, in this section, IS^{2}R will be compared with the following segmentation methods: the physics-based algorithms by Klinker et al. [129] (KLN from now on) and Gevers [68] (GEV from now on), together with the non-physics-based algorithm by Comaniciu and Meer published in [29][14] (C&M from now on). During the comparison, parameters for KLN, GEV and IS^{2}R are set up according to the requirements of the particular experiment, in order to obtain the best performance; accordingly,

[13] Remember that the method for computing intensity uncertainties of chapter 6 produces intensity uncertainties for every colour channel.

[14] The code is publicly available at http://www.caip.rutgers.edu/~comanici/segm_images.html

Table 7.1. Summary of parameters of IS^{2}R.

parameter	description
$2R_{sp}+1$	Side of the square area used to determine whether an image point corresponds to a singular point
$\Delta_{\mathcal{R}^0}$	Maximum number of intensity levels between the singular point and the pixels of $\mathcal{R}^0$
$2R_{\mathcal{R}^0}+1$	Side of the square area used to grow $\mathcal{R}^0$
$2R_{lf}+1$	Side of the square area used for local fitting
ϵ_f	Multiplicative factor related to the fitting error (equation 7.26)
n_D	Number of pixels used to estimate the mean deviation between predicted and actual intensity in the consistency test
n_A	Number of pixels used to compare predicted and actual intensity in the consistency test, to discard or accept a given pixel
ϵ_1, ϵ_2, ϵ_3	Parameters of the deviation model used in the consistency test (equation 7.31)

they are explicitly indicated in every case. The $C_1C_2C_3$ space has been selected among the three colour spaces suggested by Gevers in [68] as the working space for the GEV algorithm, since it provides invariance to objects curvature. On the other hand, C&M selects automatically the algorithm parameters according to the type of result required: under-segmentation, over-segmentation and quantization. Given that non-physics-based segmentation algorithms tend to over-segment the image, the under-segmentation option is chosen in all the experiments where C&M is involved.

Finally, as suggested in section 4.4.2, segmentation performance will be measured regarding both the classification and the contour localization aspects. The corresponding quantitative results will be provided in a section devoted to comparing IS^{2}R with the above-mentioned segmentation algorithms. On the other hand, unless otherwise stated, the CG(90), OS(90) and US(90) measures defined in section 4.4.2 will be used for the classification aspect of image segmentation, while the contour localization facet will be measured mostly by means of the Baddeley's measure (the (90) will be removed from CG, OS and US for the sake of simplicity in the following sections). Regarding the performance on shape reconstruction, since this part of IS^{2}R is executed by the B&P algorithm, it is suggested to refer to the original paper or to the surveys by Zhang et al. [302–304].

7.9.1 A detailed example

Figure 7.6 details the stages and intermediate results produced by IS^{2}R for a simple synthetic colour image. Throughout all the process, unlabelled pixels are indicated in black. In the figure, step (1) shows the regions $\mathcal{R}^0$ found, and, consequently, the location of the respective singular points. Next, step (2) presents the result of the incorporation of new pixels to every region by means of the application of the consistency tests over global surface fitting (i.e. the whole region is fitted by a quadric). In step (3), regions are considered for new extensions through the consistency test again, but now by means of local surface fitting (i.e. small areas around the pixel to be added are fitted by a quadric). Next, in step (4), regions with similar reflectance are merged. Finally, after adding unlabelled pixels to most similar adjacent regions, step (5) is reached with the final segmentation result.

7.9.2 Segmentation results for noiseless synthetic images

All the experiments of this section were performed over synthetic scenes consisting of surfaces of increasing difficulty and different reflectance combinations. Figure 7.7 shows the shape of those scenes.

The scene of figure 7.7(a) consists of a hemisphere over a plane background. It was generated using the formula:

$$Z(x,y) = \sqrt{30^2 - (x-65)^2 - (y-65)^2} \tag{7.37}$$

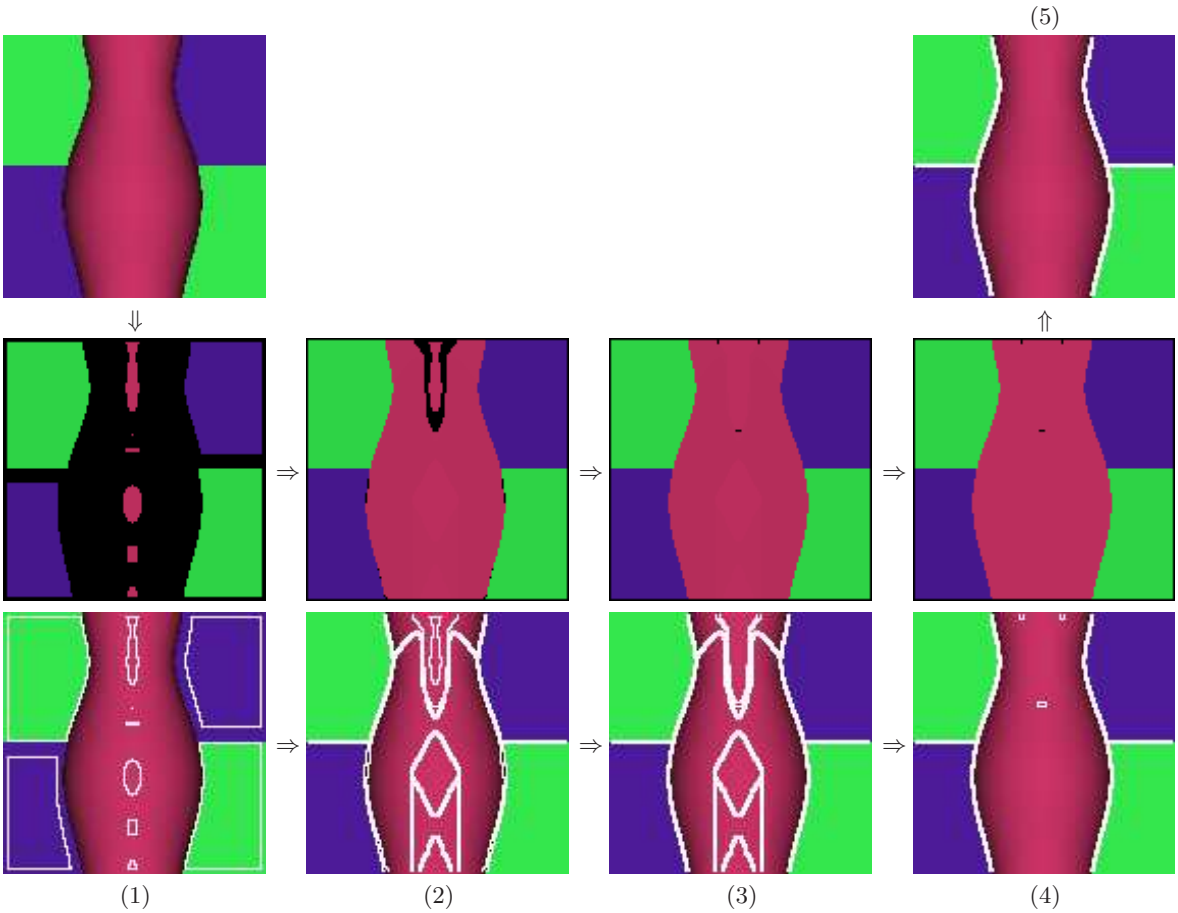


Fig. 7.6. A detailed example of how IS^{2}R works. (The row in the middle shows the progression of the segmentation, while the lower row presents the corresponding region contours.)

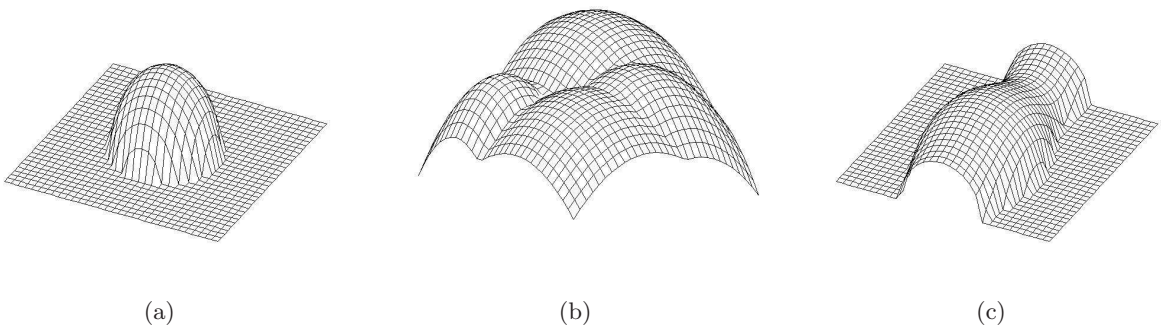


Fig. 7.7. Synthetic scenes used to test IS^{2}R: (a) SPHERE scene; (b) PARABOLOIDS scene; (c) VASE scene.

Table 7.2. Parameters used in the SPHERE, PARABOLOIDS and VASE scenes.

R_{sp}	$\Delta_{\mathcal{R}^0}$	$R_{\mathcal{R}^0}$	R_{lf}	ϵ_f	n_D	n_A	ϵ_1	ϵ_3
7	10	15	5	100	6	1	2	1

Table 7.3. Results for the SPHERE, PARABOLOIDS and VASE scene cases.

scene	lighting	reflectances recovered
(a) SPHERE	(0,0,-1)	0.902, 0.600, 0.302
	(1,0,-1)	0.898, 0.600, 0.302
	(1,1,-1)	0.898, 0.600, 0.302
(b) PARABOLOIDS	(0,0,-1)	1.000, 0.800, 0.498
	(-1,0,-1)	0.996, 0.796, 0.498
	(-1,1,-1)	1.000, 0.796, 0.498
(c) VASE	(0,0,-1)	0.898, 0.400, 0.302
	(-1,0,-1)	0.898, 0.400, 0.302
	(-1,-1,-1)	0.843, 0.400, 0.302

The second synthetic scene considered is shown in figure 7.7(b) and is a combination of four paraboloids, as equation 7.38 indicates:

$$\begin{aligned} Z(x,y) &= \max\{z_1(x,y), z_2(x,y), z_3(x,y), z_4(x,y)\} \\ z_1(x,y) &= 20 - \frac{(x+15)^2}{50} - \frac{(y-15)^2}{50} \\ z_2(x,y) &= 30 - \frac{(x-15)^2}{40} - \frac{(y-15)^2}{30} \\ z_3(x,y) &= 50 - \frac{(x-30)^2}{40} - \frac{(y+30)^2}{60} \\ z_4(x,y) &= 20 - \frac{(x+30)^2}{20} - \frac{(y+30)^2}{20} \\ & -60 \leq x \leq 60, \quad -60 \leq y \leq 60 \end{aligned} \tag{7.38}$$

Unlike the SPHERE scene, the PARABOLOIDS scene contains several curved singular points and the corresponding curved surfaces overlap. Finally, the third synthetic scene, shown in figure 7.7(c), consists of a synthetic vase over a plane background, which is typical from SFS performance evaluation studies [303]. The corresponding surface can be expressed as follows:

$$\begin{aligned} Z(x,y) &= \sqrt{f(y)^2 - x^2} \\ f(y) &= 0.15 - 0.1\, y\, (6y+1)^2 (y-1)^2 (3y-2) \\ & -0.5 \leq x \leq 0.5, \quad 0.0 \leq y \leq 1.0 \end{aligned} \tag{7.39}$$

Contrary to the previous scenes, this one includes a surface consisting of several curved singular points and it is not a quadric itself. That is to say, the actual surface does not coincide with the surface model of the fitting stage of IS^{2}R.

In all the experiments, white illumination was simulated, being the ambient lighting strength set to $L_a = 0.08 \times 255 = 20.4$ for all channels and the point light source strength $L_d = 0.92 \times 255 = 234.6$, also for all channels. On the other hand, once an image was segmented, all the scenes were reconstructed providing the B&P algorithm with true heights at singular points. Finally, the parameter values indicated in table 7.2 were used in all cases. As can be observed, n_A remained set to 1 in all cases, which means the consistency test just needed the information of one point to make a decision.

Figure 7.8 presents segmentation and reconstruction results for gray-level images of the SPHERE scene and three lighting orientations: $(0,0,-1)$, $(1,0,-1)$ and $(1,1,-1)$. In the experiment, the reflectance of the hemisphere was set to 0.9, while the plane at the background is a sort of checker board,

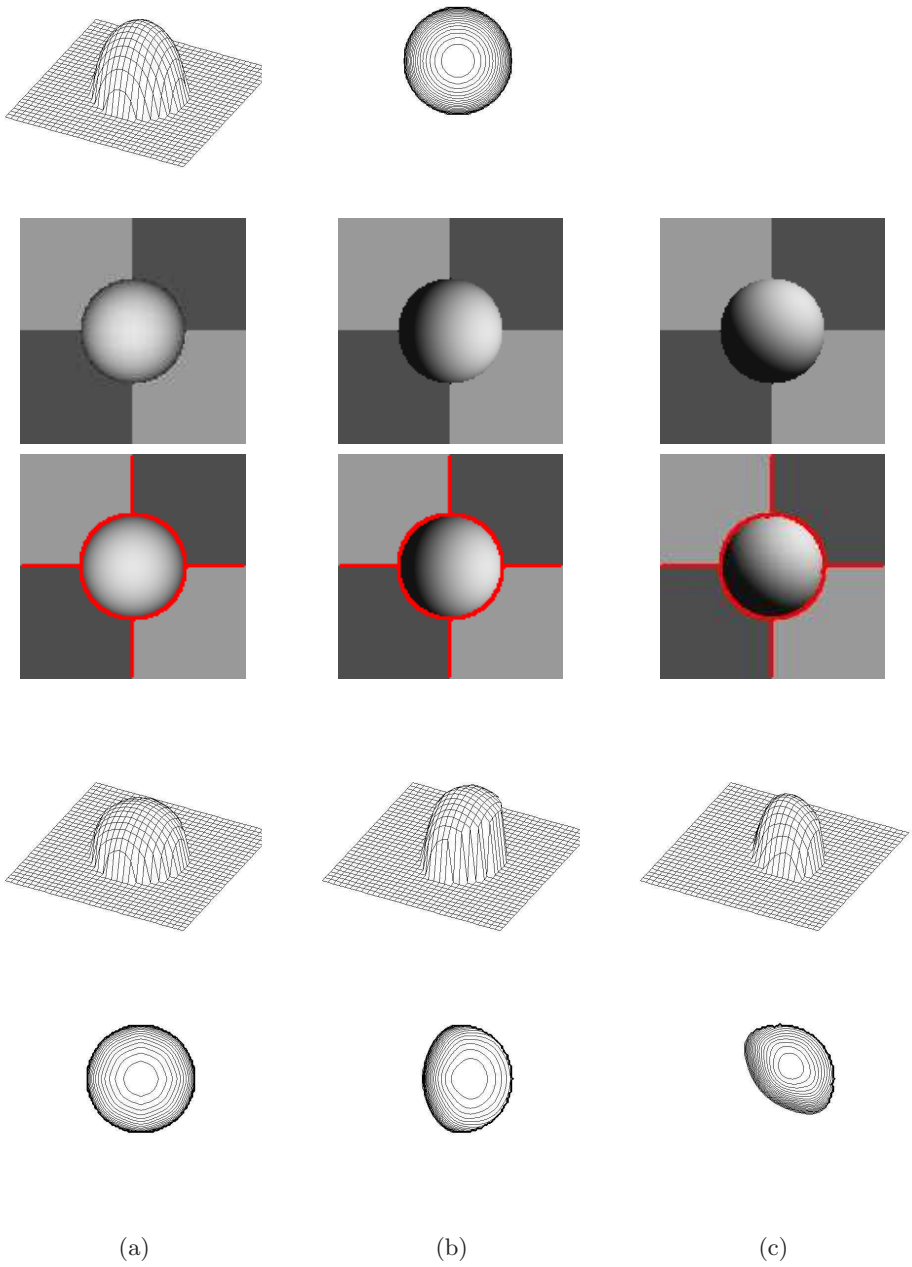


Fig. 7.8. Segmentation of the gray-level SPHERE images and shape reconstruction by means of IS^{2}R. Light source directions: (a) $(0, 0, -1)$, (b) $(1, 0, -1)$, (c) $(1, 1, -1)$. The original scene and its contour plot appears in the upper row while the contour plots of the recovered scene for (a), (b) and (c) appear in the lower row. The upper corner of each scene corresponds to the origin of the image coordinates. Region contours are superimposed over original images.

with reflectances 0.6 and 0.3. Table 7.3(a) shows the reflectances recovered. The slight deviations which can be observed are due to the discretization of intensity. On the other hand, none of the image pixels was incorrectly classified by IS^{2}R.

Figure 7.9 presents segmentation and reconstruction results for gray-level images of the PARABOLOIDS scene and three lighting orientations: $(0, 0, -1)$, $(-1, 0, -1)$ and $(-1, 1, -1)$. As for the scene materials, the reflectance of neighbouring paraboloids z_2 and z_3 was set to 0.8, while for z_1 and z_4 they were, respectively, 1 and 0.5. Table 7.3(b) shows the reflectances estimated. Again none of the image pixels was incorrectly classified and the same sort of slight deviations in the recovered reflectances is observed.

Figure 7.10 presents segmentation and reconstruction results for gray-level images of the VASE scene and three lighting orientations: $(0, 0, -1)$, $(-1, 0, -1)$ and $(-1, -1, -1)$. As well as for the SPHERE scene, the plane background was a sort of checker board, with reflectances 0.4 and 0.3, while the reflectance of the vase was 0.9. Table 7.3(c) shows the recovered reflectances. Contrary to the previous scenes, not all the reflectances are correctly recovered. This is the case of the vase, and the error comes from a bad election of the singular point. The segmentations are, nevertheless, perfect again.

As a final part of the experiment with noiseless synthetic images, IS^{2}R was faced against the same set of scenes and lighting conditions, but now coloured materials were considered. The new images and the corresponding results are provided in figure 7.11. This time no reconstruction results are given since they have already been supplied for the black and white images. As for the estimated reflectances, again slight deviations took place. Finally, once more, the segmentations produced show no misclassified pixels.

7.9.3 Experiments with real images

As for the experiments with real images, three sets of experiments were performed. The first set included three scenes constituted by high-curvature objects representable by quadric surfaces and three lighting configurations. In the second set, several scenes with the same lighting configuration and involving non-quadric real surfaces were considered. In both image sets, a special care was taken in order not to introduce objects whose surface materials could violate the image formation model assumed by IS^{2}R. To this end, the experiments were performed over pieces of clothing, plushes, cork objects and other objects painted in matte paint. Besides, all scenes were arranged over a piece of black fabric in order to avoid disturbances from the background. Finally, the third set of images consisted of several scenes with objects violating the image formation model required by IS^{2}R to check its dependence on this aspect.

On the other hand, in order to prevent the process of image formation from being affected by camera clipping, the camera iris was adjusted to control the amount of light entering the camera. The scenes were illuminated by an halogen lamp in different configurations to simulate different light source orientations. In this way, ambient lighting appeared in the form of inter-reflections on the walls and other objects. Finally, images were pre-smoothed because improvements on the determination of singular points and on the behaviour of the SFS algorithm were detected under those conditions. This pre-processing of the image was intended to counteract the non-inclusion of intensity uncertainties in both the localization of singular point localization and the computation of shape information.

Figure 7.12 shows the images for estimating the lighting conditions for every lighting configuration of the first set, together with the estimation results. As can be seen, the light comes from the side in configurations (b) and (c), while in case (a) the light source direction is nearly parallel to the optical axis.

Segmentation and reconstruction results for the three scenes can be found in figures 7.13, 7.14 and 7.15. The corresponding parameters appear in table 7.4, while table 7.5 list the reflectances recovered for each case.

As can be seen, good segmentations are obtained in all three lighting configurations and for the three scenes. As for the parameters, they were very stable among scenes and lighting configurations, as can be observed. It is worth noting how the side of the square area used to compute the singular points map was increased to a value of $2 \times 13 + 1$ pixels. In fact, it is not needed in all cases, but, generally speaking, a larger side improves significantly the localization of singular points across the image.

Regarding the relative reflectance constancy across lighting configurations which was theoretically hypothesized in section 7.3, equation 7.23, the reflectances recovered for the same object in each case

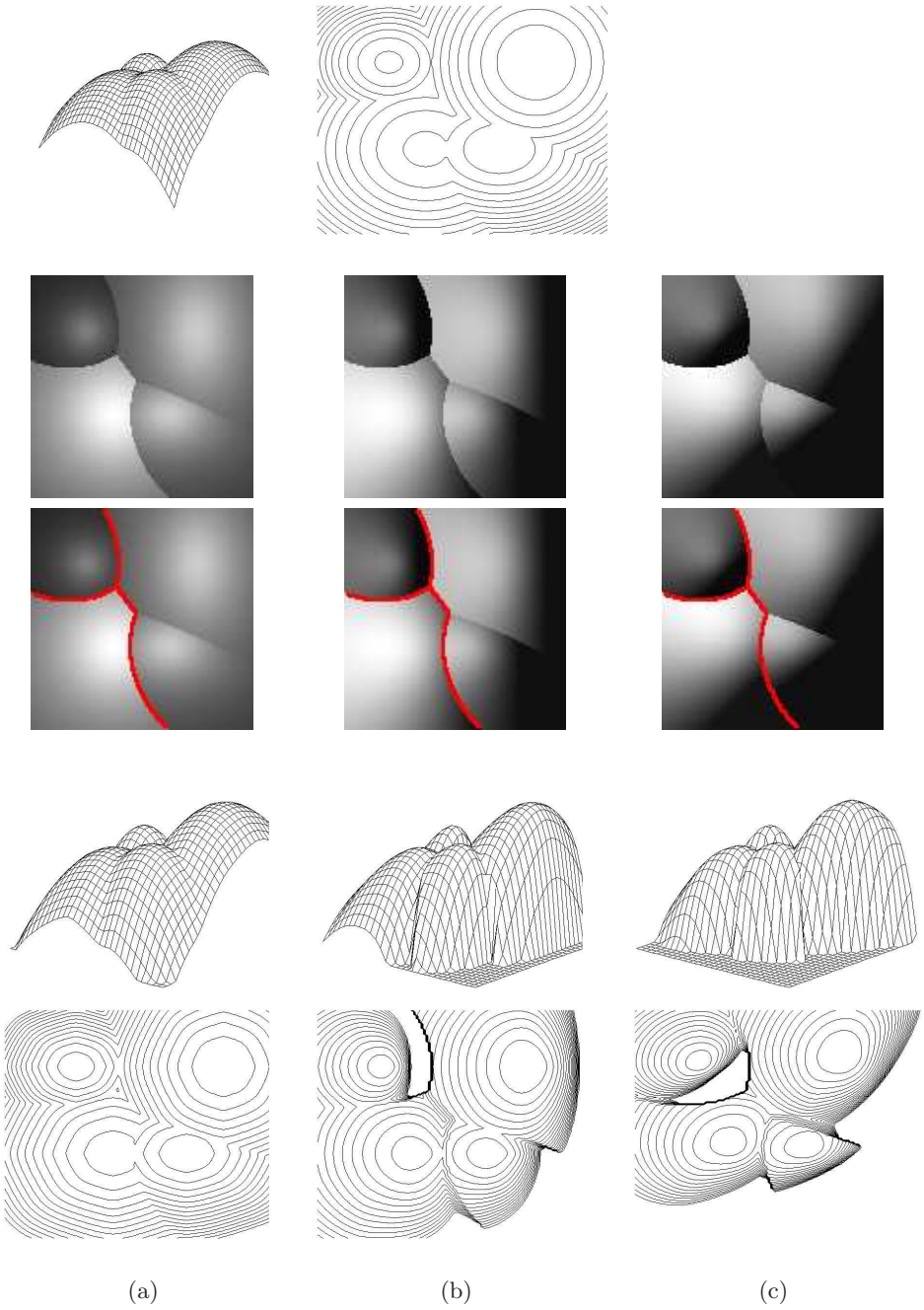


Fig. 7.9. Segmentation of the gray-level PARABOLOIDS images and shape reconstruction by means of IS^2R. Light source directions: (a) $(0, 0, -1)$, (b) $(-1, 0, -1)$, (c) $(-1, 1, -1)$. The original scene and its contour plot appears in the upper row while the contour plots of the recovered scene for (a), (b) and (c) appear in the lower row. The upper corner of each scene corresponds to the origin of the image coordinates. Region contours are superimposed over original images.

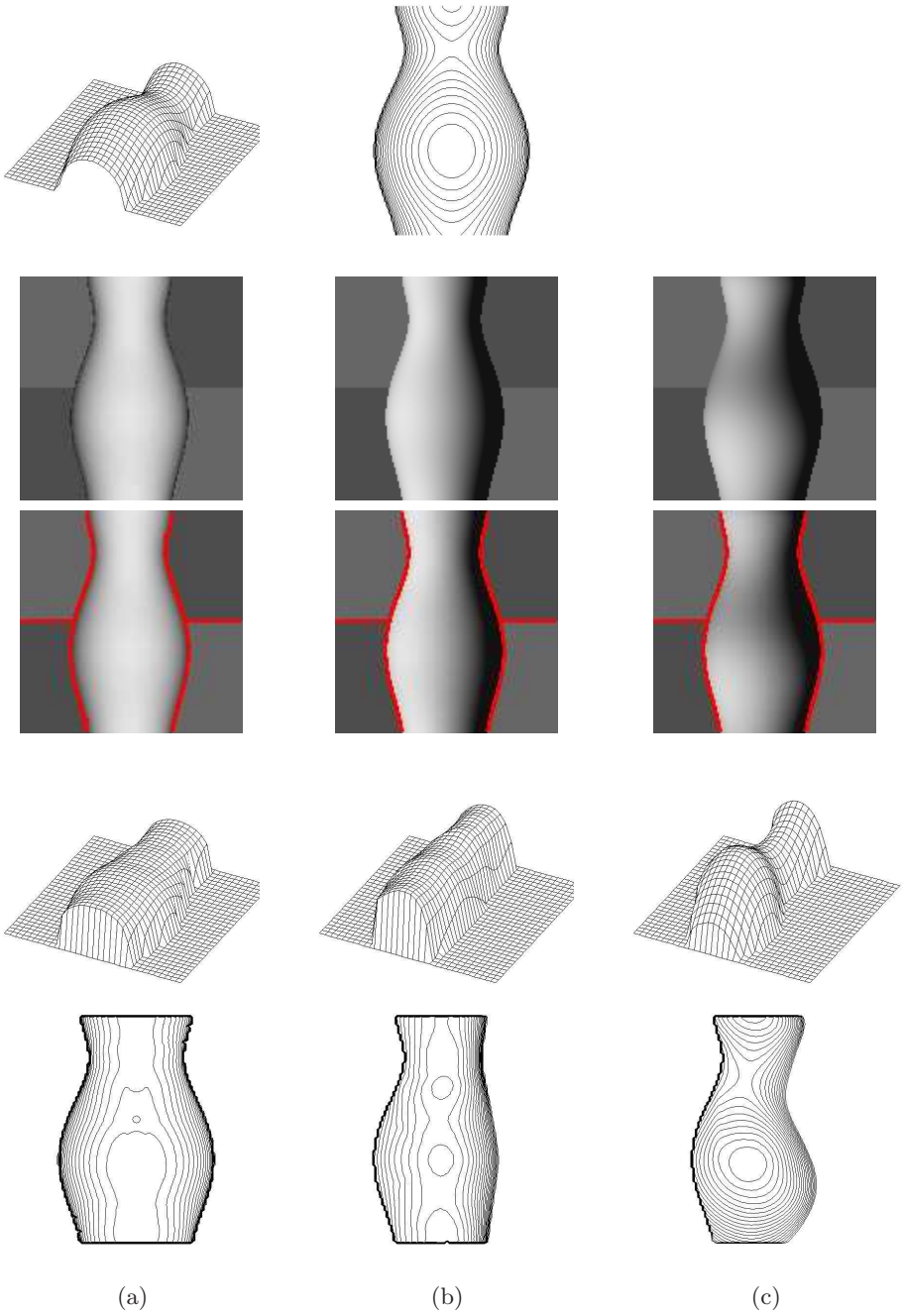


Fig. 7.10. Segmentation of the gray-level VASE images and shape reconstruction by means of IS^{2}R. Light source directions: (a) $(0, 0, -1)$, (b) $(-1, 0, -1)$, (c) $(-1, -1, -1)$. The original scene and its contour plot appears in the upper row while the contour plots of the recovered scene for (a), (b) and (c) appear in the lower row. The upper corner of each scene corresponds to the origin of the image coordinates. Region contours are superimposed over original images.

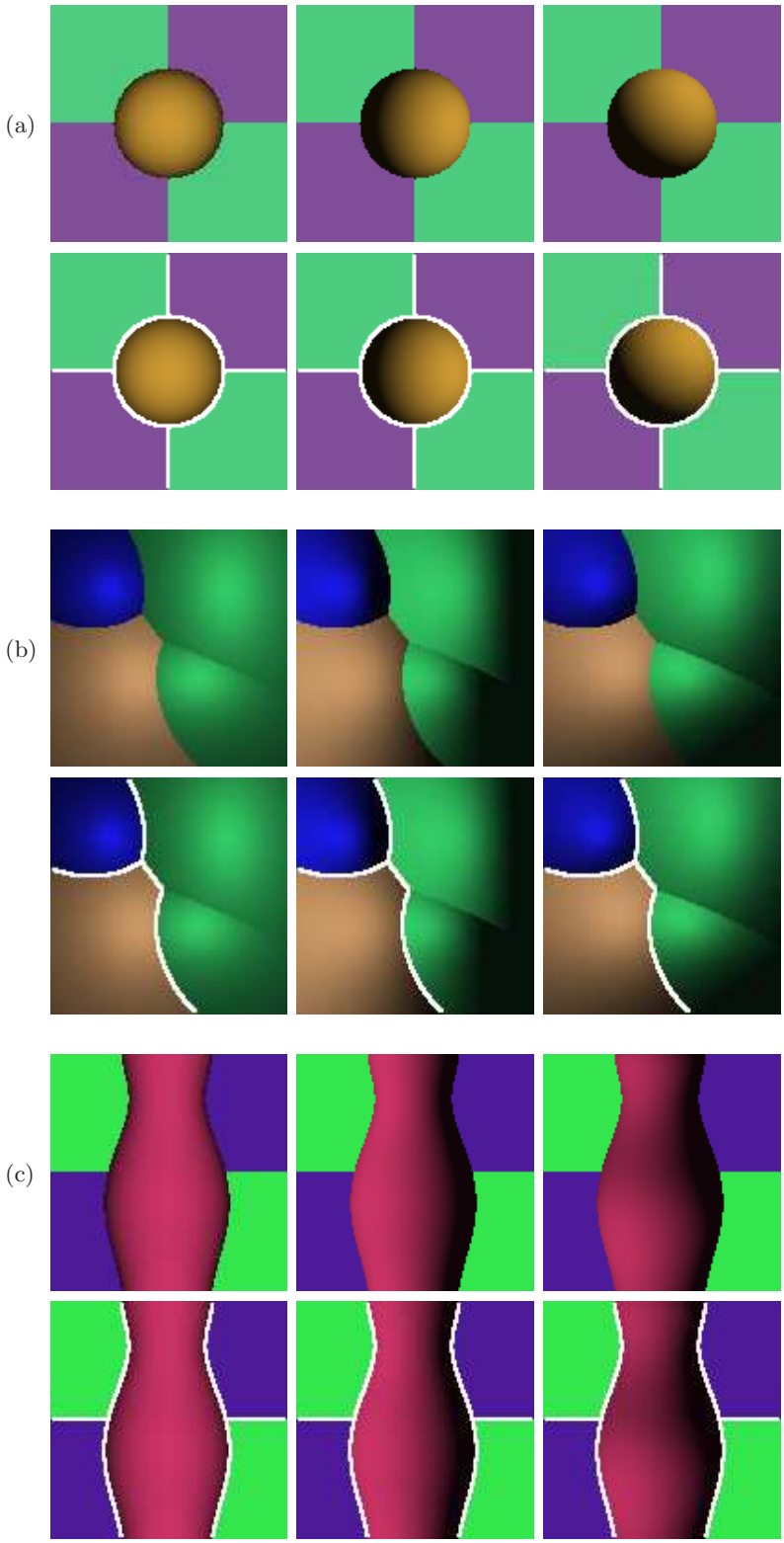


Fig. 7.11. Segmentation of the SPHERE, PARABOLOIDS and VASE colour images by IS^{2}R. For every group of images, the first row is for the original images and the second row shows the same image with region contours superimposed after segmentation.

$\boldsymbol{s} = (0.16, 0.16, -0.97)$ $L_a\rho_{b,C} = 53.01$ $L_d\rho_{b,C} = 143.24$ (a)

$\boldsymbol{s} = (-0.02, 0.51, -0.86)$ $L_a\rho_{b,C} = 33.23$ $L_d\rho_{b,C} = 119.23$ (b)

$\boldsymbol{s} = (0.55, 0.41, -0.72)$ $L_a\rho_{b,C} = 16.93$ $L_d\rho_{b,C} = 169.52$ (c)

Fig. 7.12. Lighting conditions for IS^{2}R corresponding to the first set of experiments with real images.

Table 7.4. Parameters used by IS^{2}R in the three scenes of the first set of real images.

scene	lighting	R_{sp}	$\Delta_{\mathcal{R}^0}$	$R_{\mathcal{R}^0}$	R_{lf}	ϵ_f	n_D	n_A	ϵ_1	ϵ_2	ϵ_3
1st	(a)	13	10	15	5	100	6	2	2	3	3
	(b)	13	10	15	5	100	6	2	2	3	3
	(c)	13	10	15	5	100	6	2	2	3	3
2nd	(a)	13	10	15	5	100	6	2	2	3	3
	(b)	13	10	15	5	100	6	2	2	3	3
	(c)	13	10	15	5	100	6	2	2	3	3
3rd	(a)	13	10	15	5	100	6	2	2	3	3
	(b)	13	10	15	5	100	6	2	2	3	3
	(c)	13	10	15	5	100	6	2	2	3	3

Table 7.5. Reflectances recovered by IS^{2}R for the three scenes of the first set of real images. (The reflectance of black areas is 0.000)

scene	lighting	reflectances recovered
1st	(a)	0.999, 0.535, 0.438
	(b)	1.076, 0.558, 0.394
	(c)	0.649, 0.370, 0.349
2nd	(a)	1.126, 0.530, 0.479
	(b)	1.036, 0.479, 0.387
	(c)	0.821, 0.359, 0.252
3rd	(a)	1.090, 0.948, 0.851, 0.428, 0.311
	(b)	1.253, 1.148, 0.945, 0.531, 0.315
	(c)	0.928, 0.778, 0.697, 0.349, 0.236

are relatively similar, although care should be taken if a higher accuracy was needed. The differences which are observed can be attributed to either the noise of the images and some specularities which could not be avoided (in glossy objects, specularities can be easily confounded with singular points as they are also local maxima of brightness). This could also be the cause of the reflectance having value above 1 in some cases. Nonetheless, those discrepancies were not enough so as to compromise the computation of good segmentations.

Figure 7.17 provides segmentation results for colour images of the previous scenes. As can be observed once more, quite reasonable segmentation results are obtained. The images for estimating the lighting conditions, together with the respective estimations, appear in figure 7.16. The parameter values are the same as for the black and white images (table 7.4).

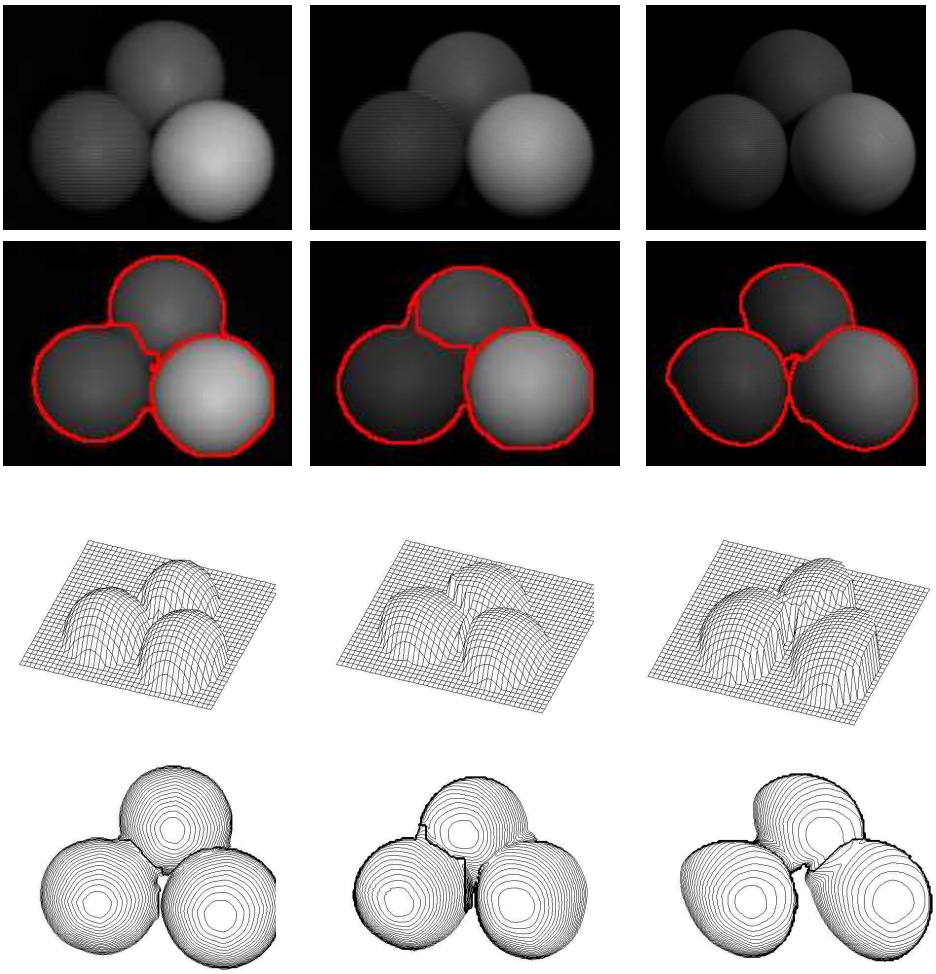


Fig. 7.13. Results of IS^{2}R for the first scene of the first set of images: [1st row] scenes (images have been enhanced to improve visualization and printing); [2nd row] segmentation results (region contours are superimposed over original images); [3rd row] height maps after rotation; [4th row] contour plots. The columns order corresponds to figure 7.12.

Figure 7.18 shows lighting estimation results for the second set of real scenes, while figures 7.19 and 7.20 present the corresponding segmentation and reconstruction results. The parameter values are indicated in table 7.6. This new set of real images consists of plushes (figure 7.19) and pieces of clothing (figure 7.20), which bring more complex surfaces than the ones of the first set. This time most part of the images are well segmented but, in some specific cases, misclassification takes place. For instance, this is the case of the duck image, figure 7.19(a), where the beak is merged with a shadow in the left side of the head's duck because of a very smooth transition from the beak to the shadow. Small details, such as the eyes of the different toys, are also wrongly segmented because the corresponding singular points are not detected due to the value given to R_{sp}. Consequently, no region was grown and the respective pixels were added to the surrounding region. As can be observed in the values of the parameters shown in table 7.6, some fine tuning was needed for this set of images. Of particular relevance is the fact that the side of the square area used to build the singular points map had to be adapted to almost every image in order to locate the largest amount of singular points. It is also noteworthy that the parameters related with the consistency test need not be modified from image to image.

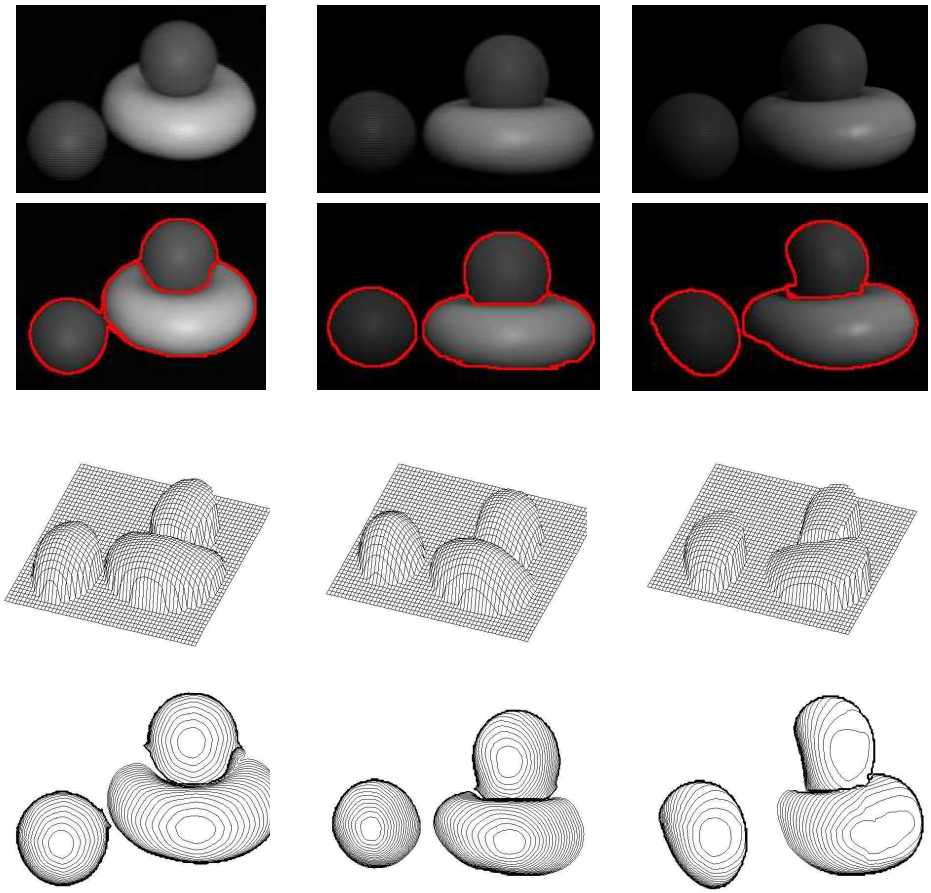


Fig. 7.14. Results of IS^{2}R for the second scene of the first set of images: [1st row] scenes (images have been enhanced to improve visualization and printing); [2nd row] segmentation results (region contours are superimposed over original images); [3rd row] height maps after rotation; [4th row] contour plots. The columns order corresponds to figure 7.12.

Table 7.6. Parameters used by IS^{2}R in the different scenes of the second set of real images.

image	R_{sp}	$\Delta_{\mathcal{R}^0}$	$R_{\mathcal{R}^0}$	R_{lf}	ϵ_f	n_D	n_A	ϵ_1	ϵ_2	ϵ_3
7.19(a)	11	10	15	5	100	6	2	2	3	3
7.19(b)	11	10	15	5	100	6	2	2	3	3
7.19(c)	5	2	5	5	100	6	2	2	3	3
7.19(d)	7	5	5	5	100	6	2	2	3	3
7.20(a)	11	10	15	5	100	6	2	2	3	3
7.20(b)	7	10	15	5	100	6	2	2	3	3
7.20(c)	13	10	15	5	100	6	2	2	3	3
7.20(d)	19	10	15	5	100	6	2	2	3	3

Segmentation results for colour images of the previous scenes are provided in figure 7.22. The behaviour exhibited by IS^{2}R in the case is the same as for the gray-level images. Lighting conditions data is given in figure 7.21. The parameter values are the same as for the black and white images (table 7.6).

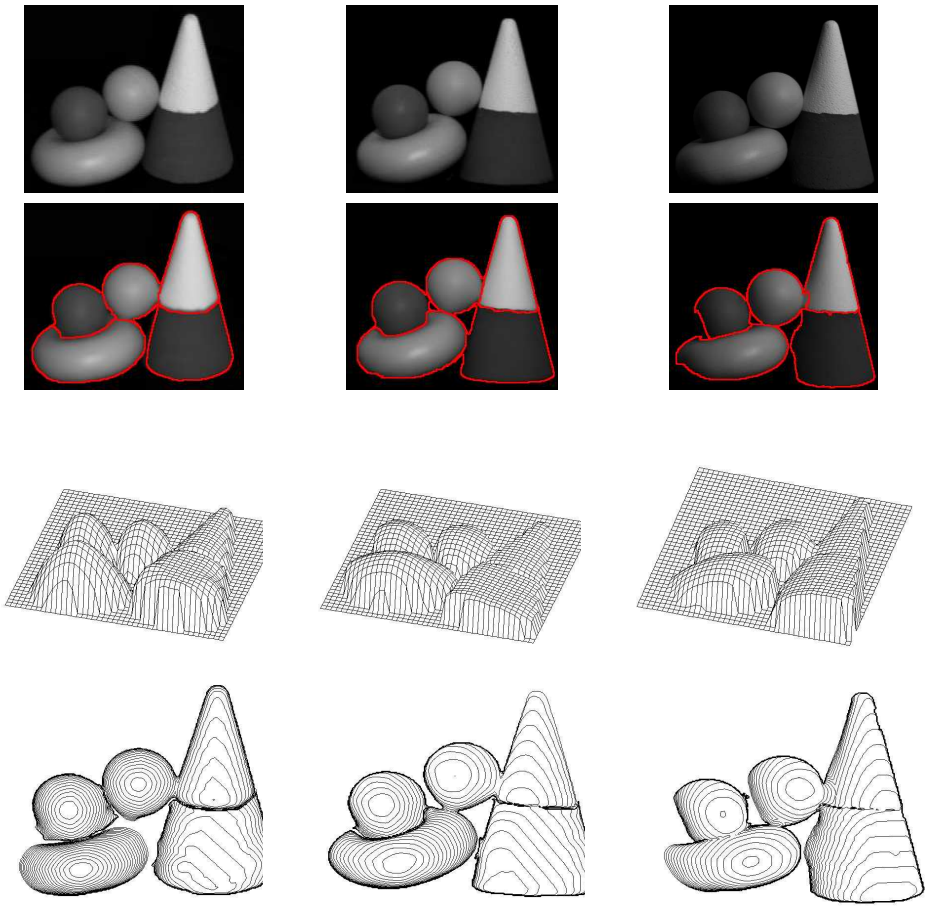


Fig. 7.15. Results of IS^{2}R for the third scene of the first set of images: [1st row] scenes (images have been enhanced to improve visualization and printing); [2nd row] segmentation results (region contours are superimposed over original images); [3rd row] height maps after rotation; [4th row] contour plots. The columns order corresponds to figure 7.12.

$\boldsymbol{s} = (0.16, 0.17, -0.97)$	$\boldsymbol{s} = (0.04, 0.54, -0.84)$	$\boldsymbol{s} = (0.56, 0.42, -0.71)$
$\boldsymbol{L_a \rho_{b,C}} = (63.01, 47.80, 40.15)$	$\boldsymbol{L_a \rho_{b,C}} = (37.41, 37.40, 23.15)$	$\boldsymbol{L_a \rho_{b,C}} = (19.30, 12.33, 17.31)$
$\boldsymbol{L_d \rho_{b,C}} = (177.84, 137.32, 102.76)$	$\boldsymbol{L_d \rho_{b,C}} = (116.72, 124.89, 92.40)$	$\boldsymbol{L_d \rho_{b,C}} = (177.69, 179.10, 115.80)$
(a)	(b)	(c)

Fig. 7.16. Lighting conditions for IS^{2}R corresponding to the colour images of the first set of images.

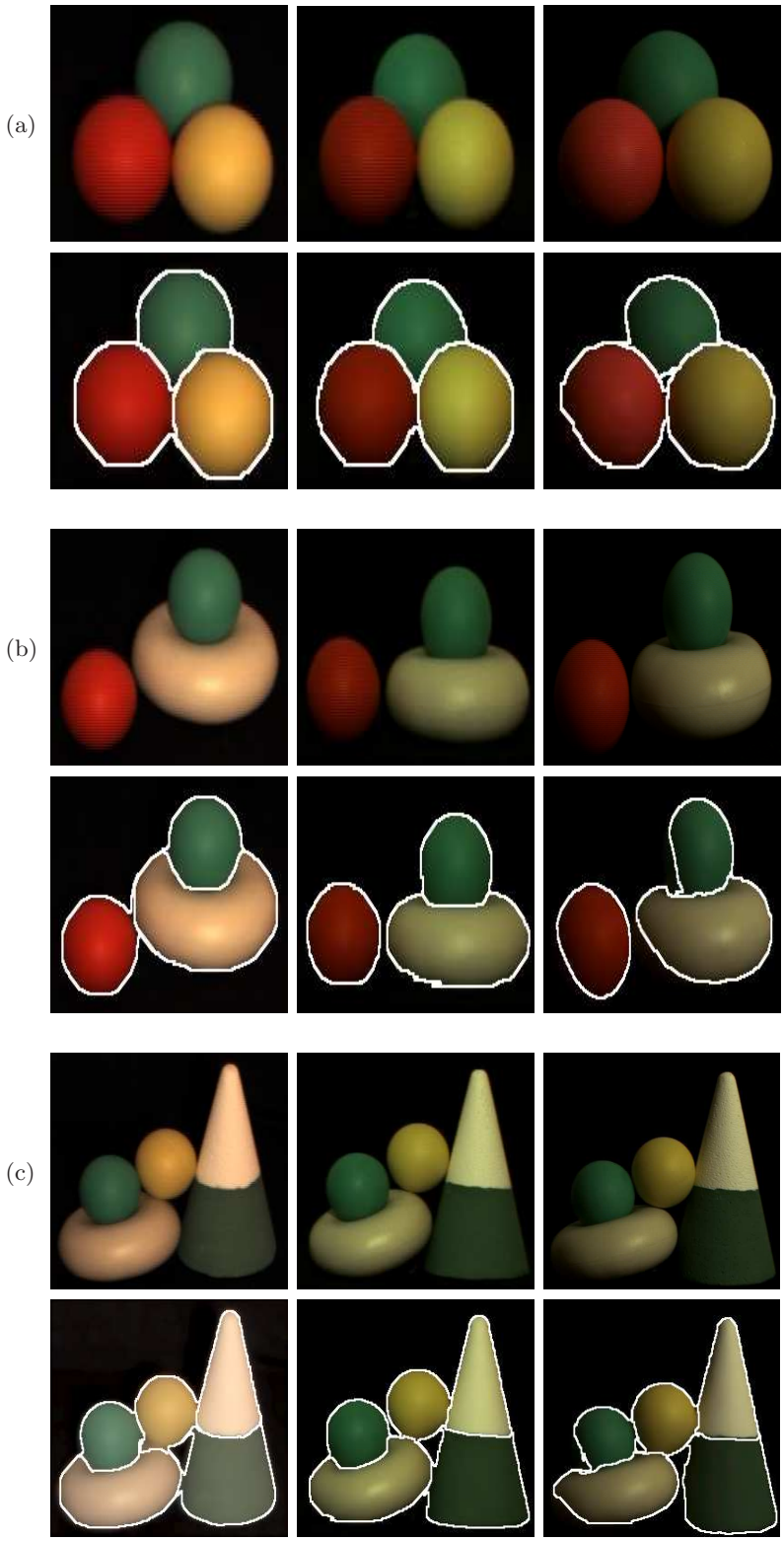
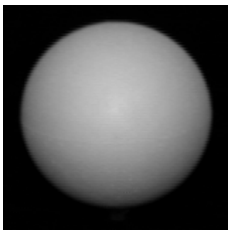


Fig. 7.17. Results of IS^{2}R for colour images of the three scenes of the first set of images. For every group of images, the first row is for the original images and the second row show the same image with region contours superimposed after segmentation.



$\boldsymbol{s} = (0.02, 0.36, -0.93)$
$L_a \rho_{b,C} = 51.11$
$L_d \rho_{b,C} = 141.59$

Fig. 7.18. Lighting conditions for IS^{2}R corresponding to the second set of experiments with real images.

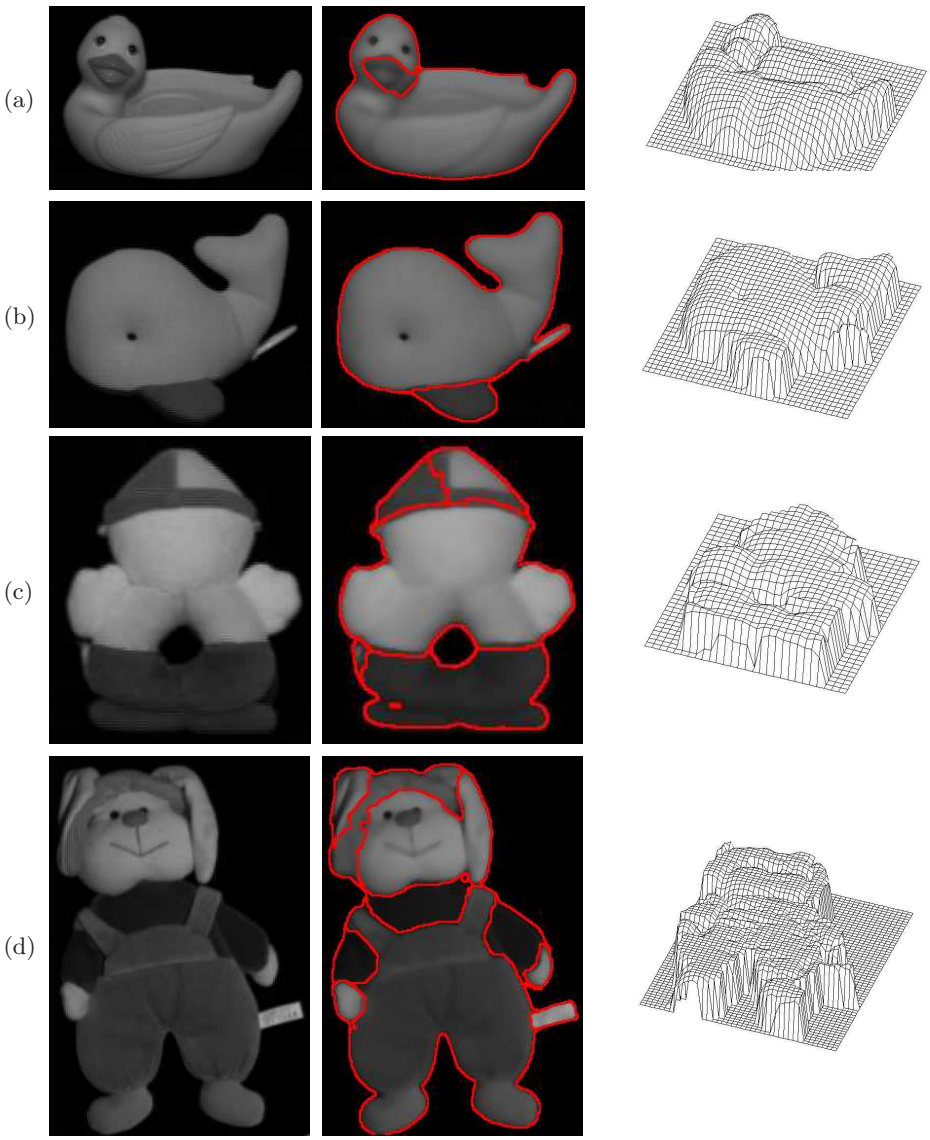


Fig. 7.19. Results of IS^{2}R for the gray-level images of the TOYS scenes of the second set of images: [1st column] scenes (images have been enhanced to improve visualization and printing); [2nd column] segmentation results (region contours are superimposed over original images); [3rd column] height maps.

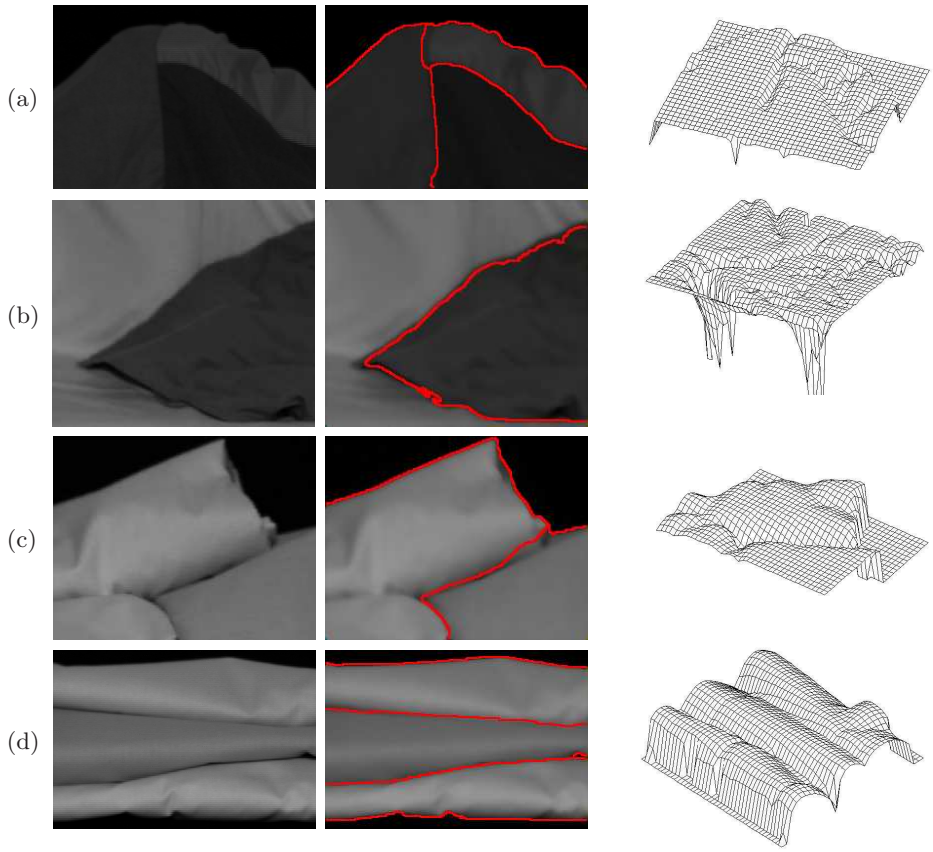


Fig. 7.20. Results of IS^{2}R for the gray-level images of the CLOTHING scenes of the second set of images: [1st column] scenes (images have been enhanced to improve visualization and printing); [2nd column] segmentation results (region contours are superimposed over original images); [3rd column] height maps.

$$\boldsymbol{s} = (0.01, 0.32, -0.95)$$
$$\boldsymbol{L_a \rho_{b,C}} = (67.16, 40.35, 26.15)$$
$$\boldsymbol{L_d \rho_{b,C}} = (171.73, 138.56, 118.98)$$

Fig. 7.21. Lighting conditions for IS^{2}R corresponding to the second set of experiments with real images.

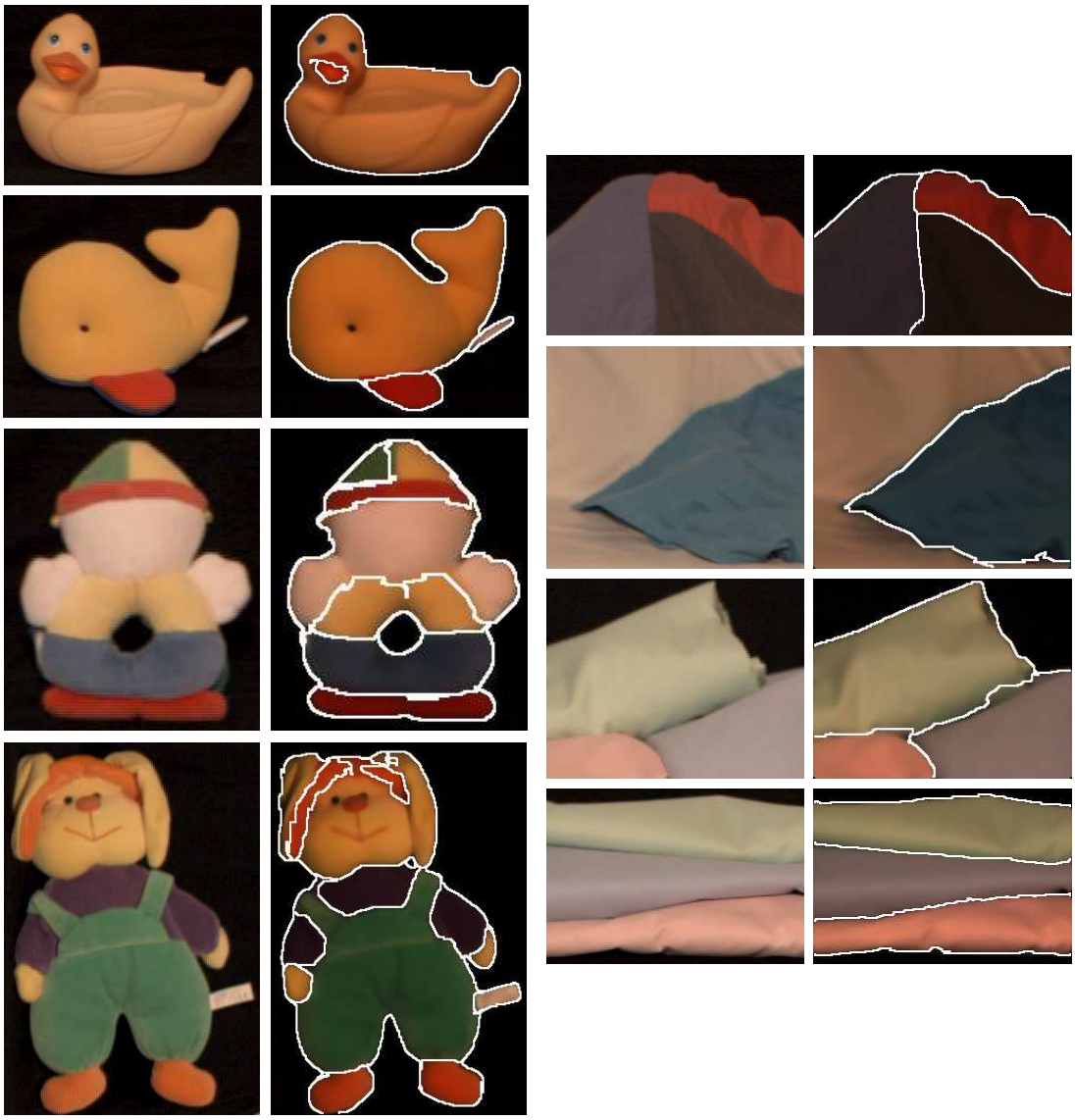


Fig. 7.22. Results of IS^{2}R for colour images of the TOYS and CLOTHING scenes of the second set of images: [1st column] TOYS scenes; [2nd column] segmentation results (region contours are superimposed over original images); [3rd column] CLOTHING scenes; [4th column] segmentation results (region contours are superimposed over original images).

To finish with the experiments with real images, figure 7.23 shows results for colour images not satisfying the image formation model required from a theoretical point of view by IS^{2}R, or just images more complex than the ones considered up until now. More precisely, the first row corresponds to images where glossy objects appear, what leads to IS^{2}R to consider specularities as singular points and, thus, to create false regions. The two images on the left of the second row correspond to natural scenes, where reflectance uniformity is typically violated. As a consequence, IS^{2}R tends to produce over-segmentation. Finally, the two images on the right of the second row are well-known images typically used to test segmentation and edge detection algorithms. Furthermore, those images are likely to have not been captured with gamma correction turned off. In the *peppers* image, again specularities are put in different regions. Besides, over- and under-segmentation also appears. A common feature of the

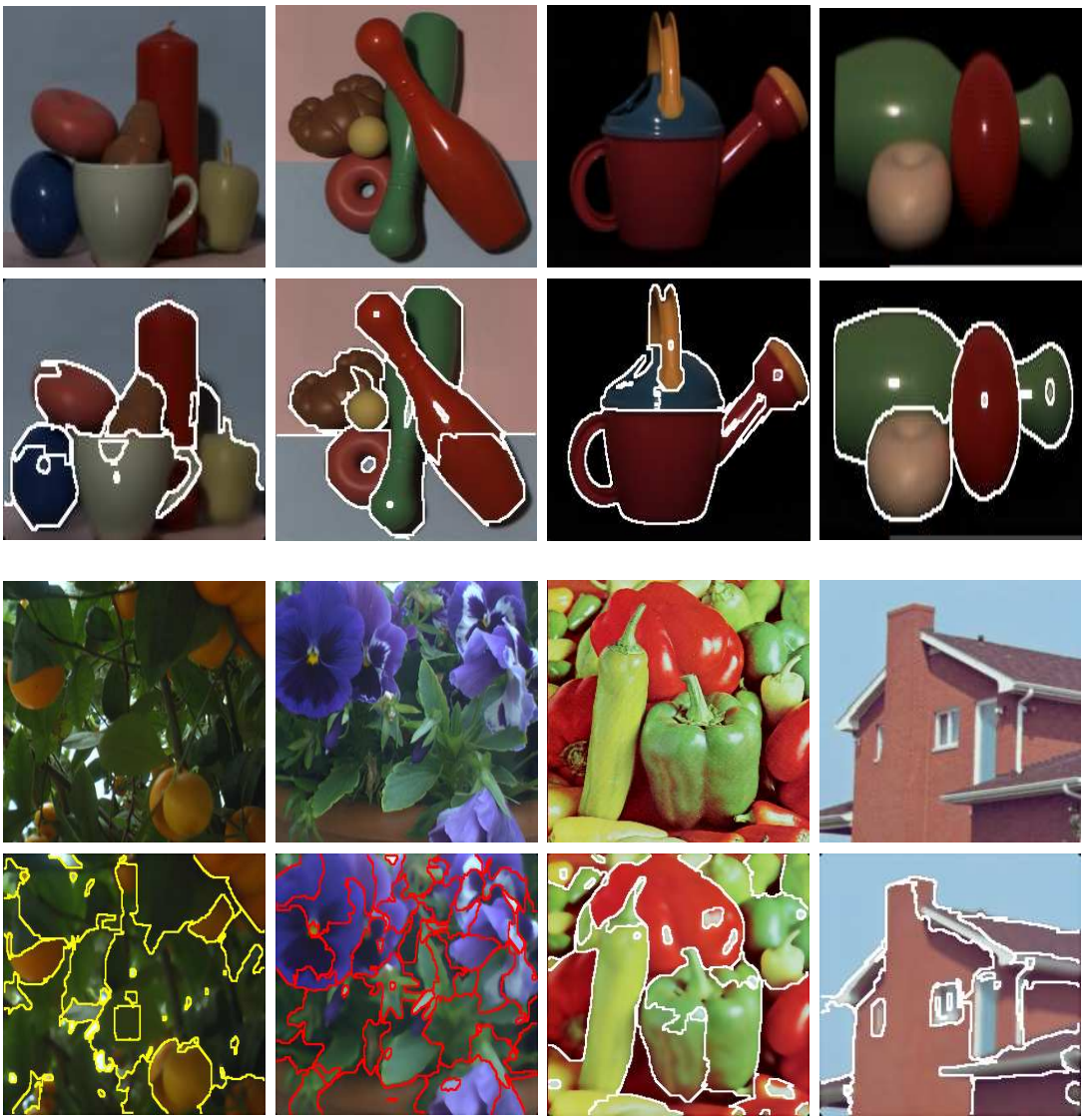


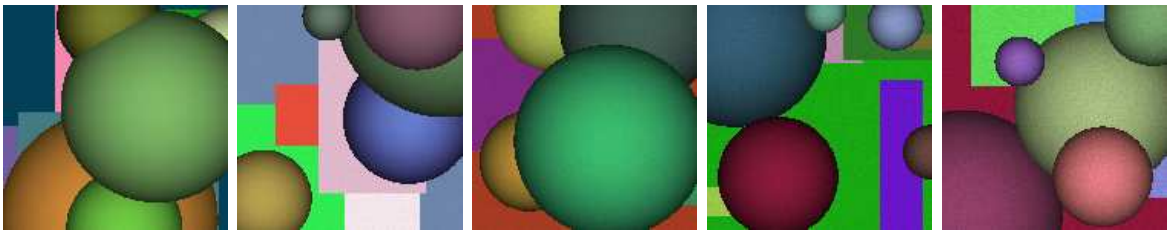
Fig. 7.23. Results of IS^{2}R for the third set of images. (Region contours are superimposed over original images.)

images in the lower row of figure 7.23 is that no knowledge was available about the lighting parameters. As well, uncertainties are artificial for the standard images.

7.9.4 Comparison with other segmentation algorithms

To finish, this section compares IS^{2}R with KLN, GEV and C&M with both gray-level and colour synthetic and real images.

First, IS^{2}R was faced against increasing noise conditions over synthetic images and compared with KLN and GEV. To this end, the synthetic images were generated following pseudocode 6.3 using new noise models corresponding to multiplying by a scalar k the standard deviations of the noise sources involved in the process of image formation (i.e. σ_K, σ_{dc} and σ_f^c, in equation 6.11b of section 6.2.2). $k = 1..5$ were the levels of noise considered, where $k = 1$ means normal noise conditions. Figure 7.24 shows examples of synthetic images for all levels. Note that $k = 5$ yields an uncertainty of 8.39 at maximum intensity, which means the corresponding confidence interval includes up to 16 levels of intensity for just $t = 1$; that is to say, the amount of noise corrupting the images for this level can



(a) (b) (c) (d) (e)

Fig. 7.24. Examples of noisy synthetic images used for comparing IS^{2}R, KLN and GEV: (a) noise×1; (b) noise×2; (c) noise×3; (d) noise×4; (e) noise×5.

Table 7.7. Parameters for IS^{2}R during the comparison under increasing noise conditions.

R_{sp}	$\Delta_{\mathcal{R}^0}$	$R_{\mathcal{R}^0}$	R_{lf}	ϵ_f	n_D	n_A	ϵ_1	ϵ_2	ϵ_3
7	10	15	5	100	6	1	2	3	3

Table 7.8. Parameters for KLN during the comparison with IS^{2}R under increasing noise conditions. (The description of the parameters corresponds to the one used in the original paper [129, table 6].)

parameter description	**value**				
	noise×1	noise×2	noise×3	noise×4	noise×5
window size (for initial grouping)			10 × 10		
minimal area size			100 pixels		
camera noise	2	3	6	9	15
cylinder width	6	8	16	24	30
minimal intensity			15		

Table 7.9. Parameters for GEV during the comparison with IS^{2}R under increasing noise conditions.

parameter description	**value**
standard deviation for gaussian derivative operators	1.0
maximum standard deviation for homogeneous regions (normalized between 0 and 255)	5.0
Tlow for hysteresis thresholding	0.9
Thigh for hysteresis thresholding	0.9

result very large. For every noise level, 100 images were generated and performance measures were computed for all the algorithms.

According to the varying noise conditions, the intensity uncertainties used by IS^{2}R were recalculated for each level. The parameter values for IS^{2}R during this experiment were chosen as the ones leading to the best results across the experiments of the previous sections. They can be found in table 7.7. The parameters of KLN were also modified from noise level to noise level in order to obtain the best performance under the different operating conditions (see table 7.8). As for GEV, curiously, it produced better results if the parameters were left constant and equal to the ones given in table 7.9.

The results for this experiment can be found in table 7.10 and in figure 7.25. As can be observed, IS^{2}R yields the best levels of correctly grouped pixels and under-segmentation, but tends to produce increasing values of over-segmentation as noise goes up. As for the localization of material boundaries, IS^{2}R is the best up to $k = 3$. For increasing levels of noise, this aspect of segmentation performance worsens up to reaching an average of more than 2 pixels of displacement with regard to true contours at the largest noise level.

Table 7.10. Comparison between IS^{2}R, KLN and GEV under increasing noise conditions by means of synthetic images: every entry of the table consists of average - standard deviation [minimum,maximum].

noise	measure	IS^{2}R	KLN	GEV
×1	CG	96.20 - 6.72 [63.14,100.00]	92.89 - 11.71 [27.75,100.00]	74.24 - 22.70 [6.95,96.03]
	OS	13.31 - 17.99 [0.00,97.02]	9.19 - 12.54 [0.00,55.43]	30.06 - 20.85 [1.28,100.00]
	US	3.17 - 6.65 [0.00,34.93]	6.42 - 11.70 [0.00,72.00]	23.70 - 23.34 [1.34,93.03]
	Baddeley	0.93 - 0.37 [0.01, 1.98]	1.15 - 0.30 [0.26, 1.88]	1.67 - 0.27 [1.08, 2.45]
×2	CG	96.83 - 5.39 [70.05,100.00]	95.55 - 9.93 [62.91,100.00]	77.32 - 17.38 [26.46,95.91]
	OS	15.43 - 16.81 [0.00,99.05]	6.13 - 11.03 [0.00,52.98]	27.91 - 21.54 [0.23,100.00]
	US	2.56 - 5.35 [0.00,29.95]	4.21 - 9.77 [0.00,37.00]	20.43 - 17.90 [1.35,72.62]
	Baddeley	0.96 - 0.33 [0.04, 1.73]	1.35 - 0.33 [0.04, 1.87]	1.61 - 0.27 [1.07, 2.42]
×3	CG	95.91 - 5.46 [69.50,100.00]	88.06 - 16.93 [37.95,100.00]	79.36 - 15.80 [30.21,95.81]
	OS	34.90 - 26.92 [0.00,100.00]	8.28 - 11.91 [0.00,45.72]	30.36 - 19.78 [1.36,100.00]
	US	3.33 - 5.44 [0.00,29.83]	11.21 - 16.95 [0.00,61.91]	18.32 - 16.41 [1.35,69.10]
	Baddeley	1.27 - 0.35 [0.33, 2.14]	1.28 - 0.33 [0.42, 2.24]	1.66 - 0.29 [0.87, 2.65]
×4	CG	94.42 - 6.84 [73.48,100.00]	71.39 - 30.60 [0.20,100.00]	80.75 - 18.29 [4.97,96.88]
	OS	53.15 - 24.74 [0.21,100.00]	12.49 - 19.15 [0.00,100.00]	34.16 - 22.55 [0.61,90.96]
	US	4.80 - 6.75 [0.00,24.82]	28.03 - 30.93 [0.00,99.80]	16.96 - 18.87 [0.00,94.89]
	Baddeley	1.64 - 0.35 [0.92, 2.57]	1.63 - 0.34 [0.84, 2.56]	1.77 - 0.35 [1.15, 2.87]
×5	CG	91.02 - 8.28 [59.98,100.00]	63.56 - 27.99 [7.94,100.00]	77.59 - 18.82 [19.67,94.49]
	OS	74.65 - 17.57 [10.32,100.00]	18.40 - 17.31 [0.00,64.59]	34.35 - 20.10 [4.63,84.88]
	US	8.28 - 8.28 [0.00,39.41]	35.52 - 28.42 [0.00,92.06]	20.15 - 19.37 [1.70,80.03]
	Baddeley	2.17 - 0.33 [1.39, 2.90]	1.55 - 0.43 [0.82, 2.79]	1.87 - 0.36 [1.10, 3.01]

In the second experiment of this section, IS^{2}R, KLN, GEV and C&M are compared by means of some of the noiseless synthetic and real gray-level and colour images used in other experiments. The ground truths for the real images were manually generated making edges coincide with whole material transitions so that the reference edge map can, and in fact do, contain thick edges (see figure 7.26). Since KLN and GEV are not able to handle gray-level images, only results for colour images are provided for them. The parameter values used for both algorithms appear in tables 7.11 and 7.12. Quantitative

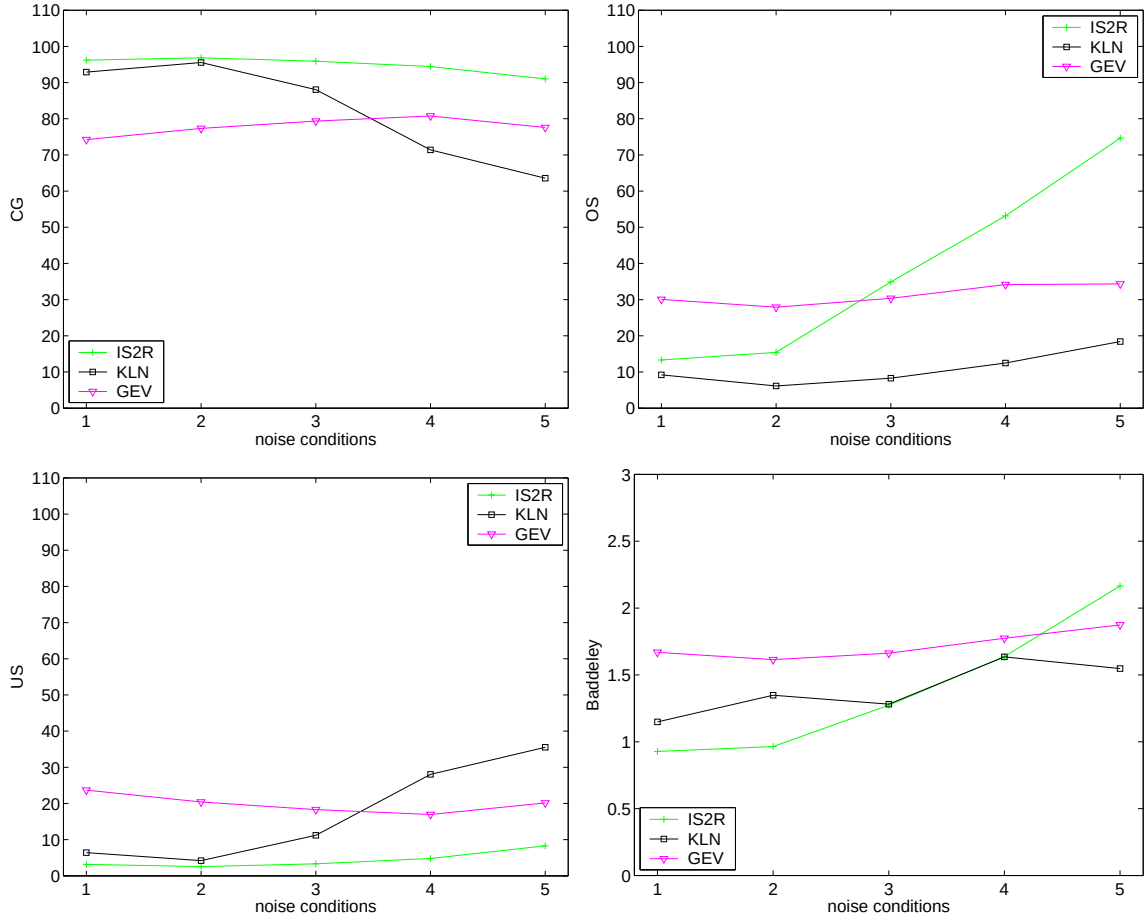


Fig. 7.25. Comparison between IS^{2}R, KLN and GEV under increasing noise conditions by means of synthetic images: plots of average values for the performance measures.

Table 7.11. Parameters for KLN during the comparison with IS^{2}R by means of gray-level and colour synthetic and real images. (The description of the parameters corresponds to the one used in the original paper [129, table 6].)

parameter description	**value for image** 7.11(a; left,middle,right)	7.17(c; left,middle,right)
window size (for initial grouping)	8×8	10×10
minimal area size	50 pixels	100 pixels
camera noise	6	9
cylinder width	2	4
minimal intensity	5	20

results are shown in tables 7.13 and 7.14 while the segmentation results appear in figures 7.27, 7.28, and 7.29. In those figures, results for IS^{2}R are replicated to make the qualitative comparison easier. As can be observed, IS^{2}R and KLN perform in a similar way, with a slight qualitative outperformance of IS^{2}R with regard to contour localization. GEV performs reasonably well with the synthetic images but poorly with the colour real images, specially in case 7.29(c). Finally, due to the high curvature of the objects in the scenes, C&M tends to produce considerable levels of both over- and under-segmentation, except for the cases of figure 7.28.

Table 7.12. Parameters for GEV during the comparison with IS^{2}R by means of gray-level and colour synthetic and real images.

parameter description	**value for image**			
	7.11(a; left,middle,right)	7.17(c)		
		(left)	(middle)	(right)
standard deviation for gaussian derivative operators	1.0	1.0	1.0	1.0
maximum standard deviation for homogeneous regions (normalized between 0 and 255)	2.0	5.0	10.0	10.0
Tlow for hysteresis thresholding	0.4	0.4	0.4	0.4
Thigh for hysteresis thresholding	0.4	0.8	0.8	0.8

Table 7.13. Comparison between IS^{2}R, KLN, GEV and C&M by means of gray-level synthetic and real images. (n.a. stands for *not available*.)

		IS^{2}R	KLN	GEV	C&M
7.8(left col)	CG	100.00	n.a.	n.a.	100.00
	OS	0.00	n.a.	n.a.	27.48
	US	0.00	n.a.	n.a.	0.00
	Baddeley	0.17	n.a.	n.a.	1.18
7.8(mid. col)	CG	100.00	n.a.	n.a.	27.45
	OS	0.00	n.a.	n.a.	15.00
	US	0.00	n.a.	n.a.	72.55
	Baddeley	0.17	n.a.	n.a.	1.07
7.8(rgt. col)	CG	100.00	n.a.	n.a.	52.20
	OS	0.00	n.a.	n.a.	13.82
	US	0.00	n.a.	n.a.	47.80
	Baddeley	0.18	n.a.	n.a.	1.05
7.15(left col)	CG	98.38	n.a.	n.a.	20.07
	OS	0.00	n.a.	n.a.	40.29
	US	0.00	n.a.	n.a.	79.93
	Baddeley	0.61	n.a.	n.a.	0.95
7.15(mid. col)	CG	98.76	n.a.	n.a.	21.63
	OS	19.18	n.a.	n.a.	52.52
	US	0.00	n.a.	n.a.	78.37
	Baddeley	0.48	n.a.	n.a.	0.89
7.15(rgt. col)	CG	98.52	n.a.	n.a.	63.76
	OS	58.89	n.a.	n.a.	100.00
	US	0.00	n.a.	n.a.	35.79
	Baddeley	0.67	n.a.	n.a.	1.29

7.10 Conclusions

This chapter has presented a segmentation algorithm called IS^{2}R (*Image Segmentation by Scene Reconstruction*) insensitive to scene curvature. To this end, reflectance values of the scene objects are estimated by means of singular points and the scene shape is progressively reconstructed in order to predict intensity values to be compared with real intensity values in a so-called consistency test, what

Table 7.14. Comparison between IS^{2}R, KLN, GEV and C&M by means of colour synthetic and real images.

		IS^{2}R	KLN	GEV	C&M
7.11(a,left)	CG	100.00	100.00	98.90	100.00
	OS	0.00	0.00	0.00	0.00
	US	0.00	0.00	1.09	0.00
	Baddeley	0.17	1.17	1.21	0.41
7.11(a,middle)	CG	100.00	100.00	98.86	100.00
	OS	0.00	0.00	0.00	0.00
	US	0.00	0.00	1.09	0.00
	Baddeley	0.17	0.84	1.21	0.69
7.11(a,right)	CG	100.00	100.00	98.86	100.00
	OS	0.00	0.00	0.00	0.00
	US	0.00	0.00	1.09	0.00
	Baddeley	0.17	0.84	1.21	0.71
7.17(c,left)	CG	98.69	97.62	95.73	99.18
	OS	0.03	1.49	9.99	40.27
	US	0.00	0.82	1.21	0.15
	Baddeley	0.54	1.09	1.01	0.92
7.17(c,middle)	CG	98.97	98.76	66.39	98.32
	OS	19.18	39.75	19.18	52.60
	US	0.00	0.54	32.99	0.26
	Baddeley	0.43	1.32	0.79	0.99
7.17(c,right)	CG	97.68	98.45	79.28	82.34
	OS	64.81	0.10	91.35	100.00
	US	0.00	0.31	19.72	17.12
	Baddeley	0.84	1.32	1.39	1.29

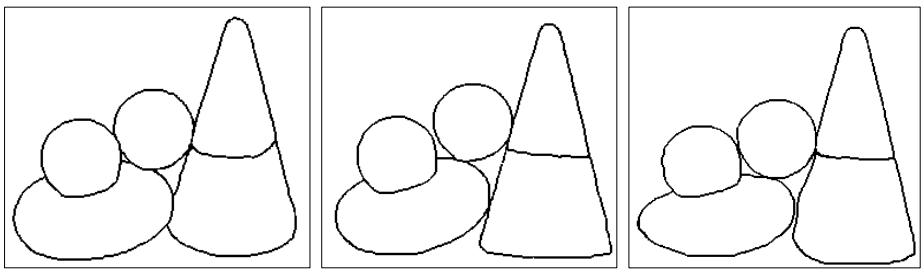


Fig. 7.26. Ground truths for the real images used in the comparison between IS^{2}R, KLN, GEV and C&M: figures 7.15 and 7.17(c).

allows making a decision about whether a pixel must join a certain region or not. Experimental results for both synthetic and real images have been provided.

During the experiments performed, both colour and gray-level images have been considered and no difference in performance has been noticed. On the other hand, algorithm parameters have resulted to be quite stable, except those related with spatial dimensions, specially for the determination of singular points. In this case, some fine tuning has been needed.

Although the experimental results prove IS^{2}R is able to deal with scenes with curved objects, its general applicability has been compromised by the limited sort of images which can correctly be

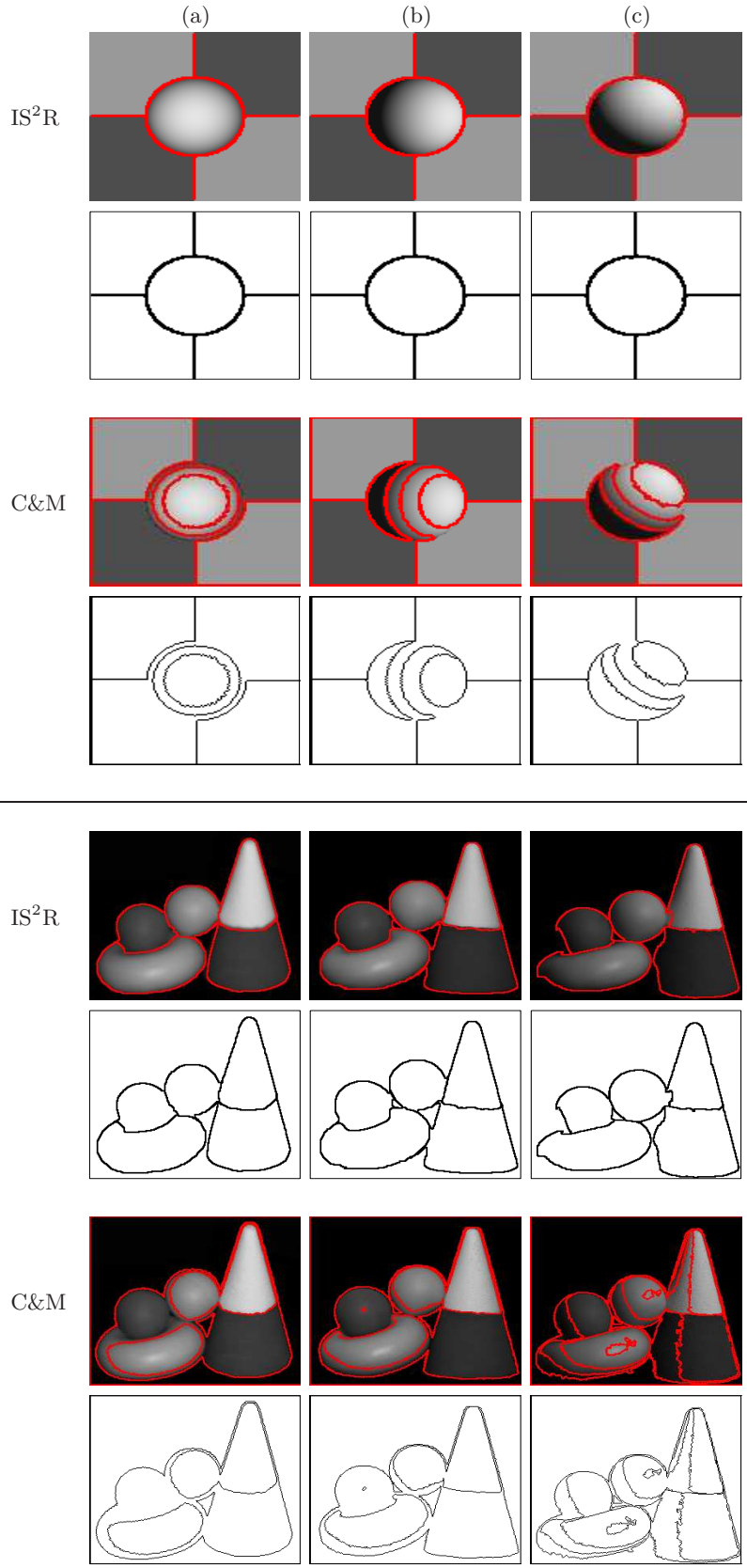


Fig. 7.27. Segmentation results for IS^{2}R and C&M for the images of figures 7.8 and 7.15.

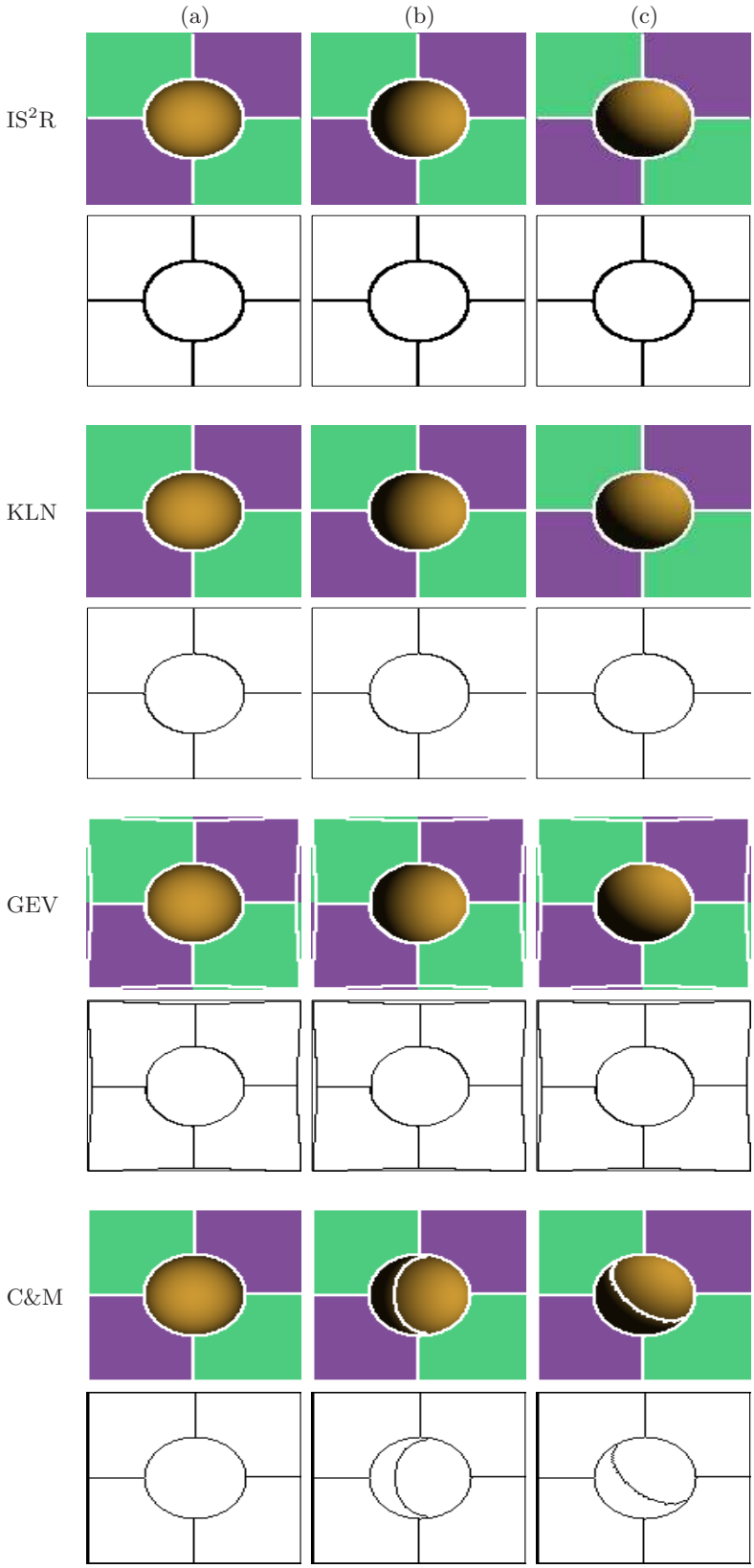


Fig. 7.28. Segmentation results for IS^{2}R, KLN, GEV and C&M for the images of figure 7.11(a).

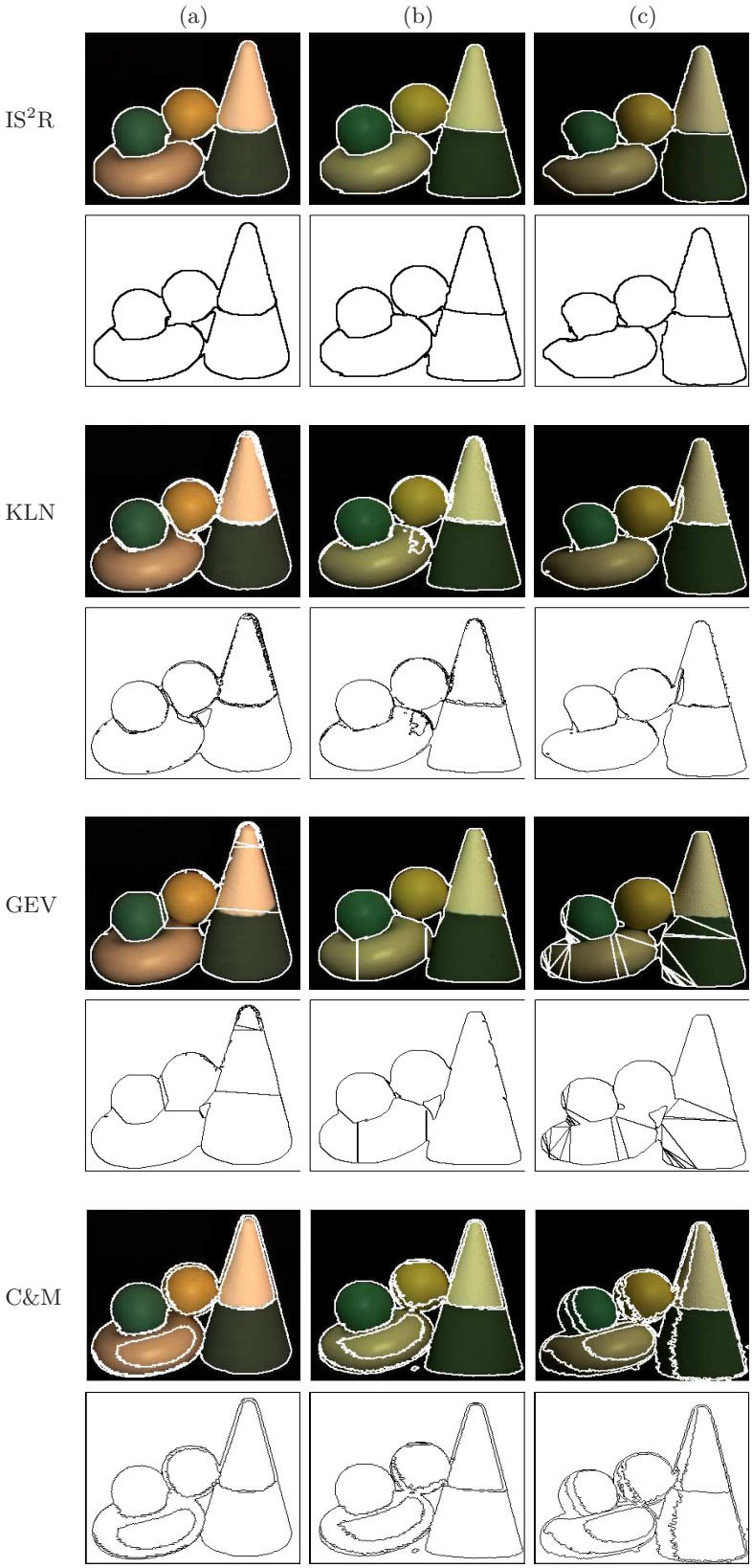


Fig. 7.29. Segmentation results for IS^{2}R, KLN, GEV and C&M for the images of figure 7.17(c).

segmented. In particular, IS^2R is not very tolerant to deviations of the theoretical image formation model, namely specularities, reflectance non-uniformities and small details. One of the clear points for this limitation is the SFS algorithm which is embedded in IS^2R for estimating shape information. Nevertheless, the very basis of IS^2R is also the cause for its low performance in certain cases, since the use of the prediction of the intensity values means a strong agreement must exist between images and the image formation model, what not always takes place. This fact suggests that other strategies based on contours instead of regions can succeed with complex images. Furthermore, IS^2R needs information from the scene which can limit its applicability in some scenarios.

It is also noteworthy the fact that the lighting estimation method CPAM-TEAV of chapter 5 has produced very consistent estimates when analyzing the same conditions of illumination over gray-level and colour images (figures 7.12 and 7.16, and figures 7.18 and 7.21).

Finally, regarding execution times, although they have not been provided in the experimental results section, some data can be given now. On the one hand, the different experiments performed have shown that IS^2R can require from 2-3 seconds up to 1-2 minutes to produce a segmentation. This coincides with its dependence on the image data involved in every case. In the cases of KLN and C&M, both have needed on the order of 1-2 minutes at most. Finally, the execution time of GEV has oscillated between less than a minute and up to 15 minutes depending on the image. These long times are because the splitting stage of GEV implies considering several triangulation scenarios for every edge point in order to take the triangulation best describing the scene, so that the larger the number of edges, the larger the execution time. All the executions were performed over a Pentium IV @ 3GHz machine.

8

Analysis of Colour Channels Coupling and Edge Detection

This chapter shows that, according to the physics-based model of image formation introduced in chapter 4, in uniform reflectance areas colour channels are coupled by the reflectance of the surface material, while, in reflectance transition zones, such coupling can be broken in a number of ways. Consequently, material changes can be found by looking for violations of this coupling. This fact is used to devise an edge detection algorithm based on colour channel coupling analysis called C^3E (*Colour Channel Coupling-based Edge detection*). This edge detection algorithm produces an edge map whose edges correspond to scene locations where a reflectance change is taking place. Moreover, chapter 9 proposes and discusses an image segmentation algorithm using this edge map as the starting point.

The rest of the chapter is organized as follows: sections 8.1, 8.2, 8.3 and 8.4 investigate on the properties of uniform reflectance areas according to the image formation model presented in chapter 4; section 8.5 enunciates a compatibility relationship among pixels based on the colour channel coupling analysis proposed in the previous sections; section 8.6 discusses about the coverage of reflectance transitions achieved by analysis of colour channels coupling; section 8.7 comments on the use of all the aforementioned to compute an edge map representing the reflectance changes of the imaged scene; next, section 8.8 comments on the thickness of the edges produced and proposes a thinning strategy; section 8.9 presents the experiments performed and the corresponding results, and, finally, conclusions appear in section 8.10.

8.1 General Properties of Uniformly Coloured Image Areas

In the following, a set of properties of uniformly coloured image areas will be discussed. For the time being, noiseless intensity values will be assumed. Noisy measurements will be considered in a later section, when discussing the implementation of an edge detection algorithm whose basis lies on the aforementioned properties.

Within an image area whose pixels correspond to scene points belonging to the same material, the ambient, body and interface colours are constant (i.e. $\rho_b(i,j;\lambda) = \rho_b(\lambda)$ and $\rho_i(i,j;\lambda) = \rho_i(\lambda)$ and, by extension, also $\rho_a(i,j;\lambda) = \rho_a(\lambda)$ because of its dependence on exclusively ρ_b and ρ_i). In a noiseless environment, changes in the pixel values between image locations are, thus, only due to changes in the geometrical factors m_b and m_i. As a consequence, in areas of uniform reflectance, colour channels keep coupled, in the sense that they are not free to take any intensity value, but they depend on the values taken by the other colour channels through the body and interface reflectances and the way they are combined in the image formation model.

This coupling can be observed from the shape of the cluster constituted by the set of pixels corresponding to a uniformly coloured object in colour space. As it is predicted by the Dichromatic Reflection Model by Shafer [252] (section 2.3.4), in general, these pixels are confined within the parallelogram defined by the ambient, body and interface colours; in case of a non-glossy object, the parallelogram reduces to the straight line defined by the ambient and the body colours. The coupling manifests, therefore, in the form of this confinement; that is to say, the corresponding pixels are not allowed to "jump" outside the cluster. This fact, although not explicitly mentioned, is exploited by

many physics-based segmentation algorithms, either by the analysis of colour clusters or by means of photometric invariants (see [68,76,129,194,203,204], just to cite some of them). Image noise, however, deforms clusters and make photometric invariants to be not so invariant, which makes the quality of the segmentation conditional on the accuracy of the estimations of the clusters parameters or the invariants.

To overcome this problem, part of this work has been devoted to formulating a set of necessary compatibility conditions which are satisfied by pixels corresponding to the same scene material. In this way, failure to meet any of these conditions implies the pixels involved cannot be grouped into the same region. As a result of this study, up to three necessary compatibility conditions have been identified. They are formulated in the form of the following properties:

Property 1 *For any pair of colour channels c_1 and c_2 and any two image locations (i_1, j_1) and (i_2, j_2) corresponding to the same scene material,*

$$\begin{aligned}
&(1)\ D^{c_1}(i_1,j_1) \geq D^{c_2}(i_1,j_1) \Leftrightarrow D^{c_1}(i_2,j_2) \geq D^{c_2}(i_2,j_2)\\
&(2)\ D^{c_1}(i_1,j_1) \leq D^{c_2}(i_1,j_1) \Leftrightarrow D^{c_1}(i_2,j_2) \leq D^{c_2}(i_2,j_2)
\end{aligned}$$

Property 2 *For any pair of colour channels c_1 and c_2, any image location (i, j) not corresponding to a material change, and any direction ξ over the image plane,*

$$\begin{aligned}
&(1)\ \left(\frac{dD^{c_1}(i,j)}{d\xi}\right) \geq 0 \Leftrightarrow \left(\frac{dD^{c_2}(i,j)}{d\xi}\right) \geq 0\\
&(2)\ \left(\frac{dD^{c_1}(i,j)}{d\xi}\right) \leq 0 \Leftrightarrow \left(\frac{dD^{c_2}(i,j)}{d\xi}\right) \leq 0
\end{aligned}$$

Property 3 *For any pair of colour channels c_1 and c_2 and any two image locations (i_1, j_1) and (i_2, j_2) corresponding to the same scene material,*

$$\begin{aligned}
(1)\quad & D^{c_1}(i_1,j_1) \geq D^{c_1}(i_2,j_2) \text{ and } D^{c_1}(i_1,j_1) \geq D^{c_2}(i_1,j_1), \text{ or}\\
& D^{c_1}(i_1,j_1) \leq D^{c_1}(i_2,j_2) \text{ and } D^{c_1}(i_1,j_1) \leq D^{c_2}(i_1,j_1)\\
& \Leftrightarrow D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1) \geq D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2)\\
(2)\quad & D^{c_1}(i_1,j_1) \geq D^{c_1}(i_2,j_2) \text{ and } D^{c_1}(i_1,j_1) \leq D^{c_2}(i_1,j_1), \text{ or}\\
& D^{c_1}(i_1,j_1) \leq D^{c_1}(i_2,j_2) \text{ and } D^{c_1}(i_1,j_1) \geq D^{c_2}(i_1,j_1)\\
& \Leftrightarrow D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1) \leq D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2)
\end{aligned}$$

The previous properties characterize in fact the three following facts:

- *Colour channels do not cross one another* (property 1). As a consequence, colour channels are ordered in regions of uniform reflectance in the sense that, if for pixel (i_1, j_1) the intensity for channel c_1 is "above" the intensity of channel c_2 (i.e. the intensity in channel c_1 is larger than in channel c_2), then for any pixel (i_2, j_2) in the same region, colour channel c_1 will also be "above" colour channel c_2.

- *Colour channels vary in a coordinated way: when one changes, so do the others, and in the same sense, all increase or all decrease* (property 2). Observe that another form of this property is as follows: "*Given two image locations (i_1, j_1) and (i_2, j_2) corresponding to the same scene material, $D^{c_1}(i_1, j_1) \geq D^{c_1}(i_2, j_2)$ if and only if $D^{c_2}(i_1, j_1) \geq D^{c_2}(i_2, j_2)$, and the same applies for $\leq$*".

- *As the intensity in one channel decreases, so does the difference between colour channel intensities; the opposite happens when the intensity in one channel increases* (property 3). Notice that the property is formulated so as to take into account the four possible cases:

(a) if $D^{c_1}(i_1, j_1) \geq D^{c_2}(i_1, j_1)$ (channel c_1 is "above" channel c_2) and

(i) $D^{c_1}(i_1, j_1) \geq D^{c_1}(i_2, j_2)$ (intensity decreases between (i_1, j_1) and (i_2, j_2)),

$D^{c_1}(i_1, j_1) - D^{c_2}(i_1, j_1) \geq 0$ is above $D^{c_1}(i_2, j_2) - D^{c_2}(i_2, j_2) \geq 0$ (case (1), 1st row)

(ii) $D^{c_1}(i_1, j_1) \leq D^{c_1}(i_2, j_2)$ (intensity increases between (i_1, j_1) and (i_2, j_2)),

$D^{c_1}(i_1, j_1) - D^{c_2}(i_1, j_1) \geq 0$ is below $D^{c_1}(i_2, j_2) - D^{c_2}(i_2, j_2) \geq 0$ (case (2), 2nd row)

(b) if $D^{c_1}(i_1, j_1) \leq D^{c_2}(i_1, j_1)$ (channel c_1 is "below" channel c_2) and

(i) $D^{c_1}(i_1, j_1) \geq D^{c_1}(i_2, j_2)$ (intensity decreases between (i_1, j_1) and (i_2, j_2)),

$D^{c_1}(i_1, j_1) - D^{c_2}(i_1, j_1) \leq 0$ is below $D^{c_1}(i_2, j_2) - D^{c_2}(i_2, j_2) \leq 0$ (case (2), 1st row)

(ii) $D^{c_1}(i_1, j_1) \leq D^{c_1}(i_2, j_2)$ (intensity increases between (i_1, j_1) and (i_2, j_2)),

$D^{c_1}(i_1, j_1) - D^{c_2}(i_1, j_1) \leq 0$ is above $D^{c_1}(i_2, j_2) - D^{c_2}(i_2, j_2) \leq 0$ (case (1), 2nd row)

8.2 Particularization of the General Properties

The fulfillment of properties 1-3 depends on the particular instantiation of the image formation model (equations 4.4 and 4.6). Four cases are distinguished:

- *Case 1: non-glossy pixels/without ambient lighting*
- *Case 2: non-glossy pixels/with ambient lighting*
- *Case 3: glossy pixels/without ambient lighting*
- *Case 4: glossy pixels/with ambient lighting*

They are all revised in the following sections.

8.2.1 Case 1: non-glossy pixels/without ambient lighting

If interface reflection and ambient lighting are not taken into account, the intensity of every colour channel is reduced to the following expression:

$$D^c(i, j) = m_b(i, j) C_b^c(i, j) \,. \tag{8.1}$$

In such a case, the three properties are met. In other words, assuming the model of equation 8.1, in uniform reflectance areas, colour channels do not cross one another, they vary coordinately, and the difference between colour channel intensities decreases as the intensity in anyone of them decreases. These facts are stated in the following three propositions, whose proofs can be found in appendix B.

Proposition 8.1. *Assuming ambient lighting is negligible and directional illumination is uniform throughout the scene, for any pair of colour channels c_1 and c_2, and any two image locations (i_1, j_1) and (i_2, j_2) coming from the same scene material, such that interface reflection is negligible, i.e. $D_i^c(i_1, j_1) = 0$ and $D_i^c(i_2, j_2) = 0$,*

$$(1)\ D^{c_1}(i_1, j_1) \geq D^{c_2}(i_1, j_1) \Leftrightarrow D^{c_1}(i_2, j_2) \geq D^{c_2}(i_2, j_2)$$
$$(2)\ D^{c_1}(i_1, j_1) \leq D^{c_2}(i_1, j_1) \Leftrightarrow D^{c_1}(i_2, j_2) \leq D^{c_2}(i_2, j_2)$$

Proposition 8.2. *Assuming ambient lighting is negligible and directional illumination is uniform throughout the scene, for any pair of colour channels c_1 and c_2, any image location (i, j) not corresponding to a material change, such that interface reflection is negligible, i.e. $D_i^c(i, j) = 0$, and any direction ξ over the image plane,*

$$(1)\ \left(\frac{dD^{c_1}(i,j)}{d\xi}\right) \geq 0 \Leftrightarrow \left(\frac{dD^{c_2}(i,j)}{d\xi}\right) \geq 0$$
$$(2)\ \left(\frac{dD^{c_1}(i,j)}{d\xi}\right) \leq 0 \Leftrightarrow \left(\frac{dD^{c_2}(i,j)}{d\xi}\right) \leq 0$$

Proposition 8.3. *Assuming ambient lighting is negligible and directional illumination is uniform throughout the scene, for any pair of colour channels c_1 and c_2, and any two image locations (i_1, j_1) and (i_2, j_2) coming from the same scene material, such that interface reflection is negligible, i.e. $D_i^c(i_1, j_1) = 0$ and $D_i^c(i_2, j_2) = 0$,*

$$\begin{aligned}
(1)\quad & D^{c_1}(i_1,j_1) \geq D^{c_1}(i_2,j_2) \text{ and } D^{c_1}(i_1,j_1) \geq D^{c_2}(i_1,j_1), \text{ or} \\
& D^{c_1}(i_1,j_1) \leq D^{c_1}(i_2,j_2) \text{ and } D^{c_1}(i_1,j_1) \leq D^{c_2}(i_1,j_1) \\
& \Leftrightarrow D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1) \geq D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2) \\
(2)\quad & D^{c_1}(i_1,j_1) \geq D^{c_1}(i_2,j_2) \text{ and } D^{c_1}(i_1,j_1) \leq D^{c_2}(i_1,j_1), \text{ or} \\
& D^{c_1}(i_1,j_1) \leq D^{c_1}(i_2,j_2) \text{ and } D^{c_1}(i_1,j_1) \geq D^{c_2}(i_1,j_1) \\
& \Leftrightarrow D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1) \leq D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2)
\end{aligned}$$

8.2.2 Case 2: non-glossy pixels/with ambient lighting

Now, the resultant image formation model is given by:

$$D^c(i,j) = C_a^c(i,j) + m_b(i,j)C_b^c(i,j)\,. \tag{8.2}$$

In this case, only property 2 is met in general. Properties 1 and 3 require an additional hypothesis to be applicable. All these facts are stated in the following lemma and propositions: (As before, formal proofs can be found in appendix B.)

Lemma 8.4. *If ambient illumination is proportional to directional illumination (i.e. $L_a(\lambda) = \alpha L_d(\lambda)$ $[\alpha \geq 0], \forall\lambda$) and interface reflection is not involved in the reflected light, then, for any material and any two colour channels c_1 and c_2:*

$$(1)\ C_a^{c_1} \geq C_a^{c_2} \Leftrightarrow C_b^{c_1} \geq C_b^{c_2}$$

$$(2)\ C_a^{c_1} \leq C_a^{c_2} \Leftrightarrow C_b^{c_1} \leq C_b^{c_2}$$

Proposition 8.5. *Assuming directional illumination is uniform throughout the scene and ambient illumination is proportional to directional illumination, for any pair of colour channels c_1 and c_2, and any two image locations (i_1, j_1) and (i_2, j_2) coming from the same scene material, such that interface reflection is negligible, i.e. $D_i^c(i_1, j_1) = 0$ and $D_i^c(i_2, j_2) = 0$,*

$$(1)\ D^{c_1}(i_1,j_1) \geq D^{c_2}(i_1,j_1) \Leftrightarrow D^{c_1}(i_2,j_2) \geq D^{c_2}(i_2,j_2)$$
$$(2)\ D^{c_1}(i_1,j_1) \leq D^{c_2}(i_1,j_1) \Leftrightarrow D^{c_1}(i_2,j_2) \leq D^{c_2}(i_2,j_2)$$

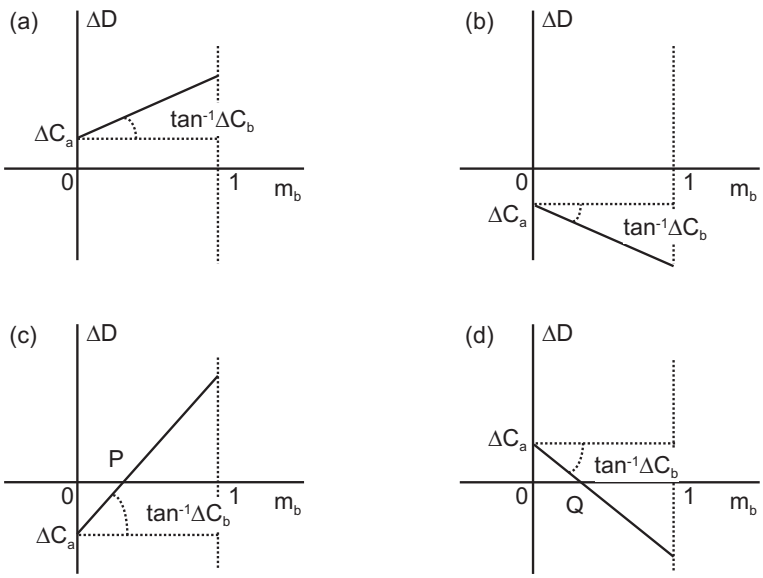


Fig. 8.1. Four possible configurations of the straight line $\Delta D(i,j) = \Delta C_a + m_b(i,j)\Delta C_b$.

Proposition 8.6. *Assuming directional illumination is uniform throughout the scene, for any pair of colour channels c_1 and c_2, any image location (i,j) not corresponding to a material change, such that such that interface reflection is negligible, i.e. $D_i^c(i,j) = 0$, and any direction ξ over the image plane,*

$$(1)\quad \left(\frac{dD^{c_1}(i,j)}{d\xi}\right) \geq 0 \Leftrightarrow \left(\frac{dD^{c_2}(i,j)}{d\xi}\right) \geq 0$$
$$(2)\quad \left(\frac{dD^{c_1}(i,j)}{d\xi}\right) \leq 0 \Leftrightarrow \left(\frac{dD^{c_2}(i,j)}{d\xi}\right) \leq 0$$

Proposition 8.7. *Assuming directional illumination is uniform throughout the scene and ambient illumination is proportional to directional illumination, for any pair of colour channels c_1 and c_2, and any two image locations (i_1,j_1) and (i_2,j_2) coming from the same scene material, such that such that interface reflection is negligible, i.e. $D_i^c(i_1,j_1) = 0$ and $D_i^c(i_2,j_2) = 0$,*

$$\begin{aligned}
(1)\quad & D^{c_1}(i_1,j_1) \geq D^{c_1}(i_2,j_2) \text{ and } D^{c_1}(i_1,j_1) \geq D^{c_2}(i_1,j_1), \text{ or} \\
& D^{c_1}(i_1,j_1) \leq D^{c_1}(i_2,j_2) \text{ and } D^{c_1}(i_1,j_1) \leq D^{c_2}(i_1,j_1) \\
& \Leftrightarrow D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1) \geq D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2) \\
(2)\quad & D^{c_1}(i_1,j_1) \geq D^{c_1}(i_2,j_2) \text{ and } D^{c_1}(i_1,j_1) \leq D^{c_2}(i_1,j_1), \text{ or} \\
& D^{c_1}(i_1,j_1) \leq D^{c_1}(i_2,j_2) \text{ and } D^{c_1}(i_1,j_1) \geq D^{c_2}(i_1,j_1) \\
& \Leftrightarrow D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1) \leq D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2)
\end{aligned}$$

As for properties 1 and 3, the introduction of the ambient reflection term has led to the fact that the sign of $D^{c_1}(i,j) - D^{c_2}(i,j)$ does not determine, in the general case, the sign of the difference for any other image point for the same scene material. In effect, for all the points (i,j) belonging to the same scene material, the differences between colour channels c_1 and c_2, $\Delta D(i,j) = D^{c_1}(i,j) - D^{c_2}(i,j) = (C_a^{c_1} - C_a^{c_2}) + m_b(i,j)(C_b^{c_1} - C_b^{c_2}) = \Delta C_a + m_b(i,j)\Delta C_b$, lie in a straight line with variables ΔD and m_b and parameters ΔC_a (intercept with ΔD axis) and ΔC_b (slope with m_b axis). Observe that this straight line can adopt any of the four configurations shown in figure 8.1. In cases (a) and (b), the signs

of ΔC_a and ΔC_b are the same and, thus, coincide with the sign of ΔD for any value of m_b. However, in cases (c) and (d), the signs of ΔC_a and ΔC_b are different and only if $m_b(i_1, j_1)$ and $m_b(i_2, j_2)$ are both above or both below $P = Q = -\frac{\Delta C_a}{\Delta C_b}$ (see figure 8.1) the signs of the differences $D^{c_1}(i_1, j_1) - D^{c_2}(i_1, j_1)$ and $D^{c_1}(i_2, j_2) - D^{c_2}(i_2, j_2)$ are the same. The constraint relating ambient lighting with directional illumination through the scale factor $\alpha \geq 0$ in propositions 8.5 and 8.7 force situations (a) and (b) everywhere in the scene.

8.2.3 Case 3: glossy pixels/without ambient lighting

In this case, the intensity of every colour channel is given by the following expression:

$$D^c(i, j) = m_b(i, j)C_b^c(i, j) + m_i(i, j)C_i^c(i, j)\,. \tag{8.3}$$

When interface reflection is introduced in the formation model, only property 1 is satisfied after a slight modification of the formulation, as it is indicated in the following proposition: (For the formal proof see again appendix B.)

Proposition 8.8. *Assuming ambient lighting is negligible and directional illumination is uniform throughout the scene, and assuming the NIR formation model, for any pair of colour channels c_1 and c_2, and any two image locations (i_1, j_1) and (i_2, j_2) coming from the same scene material,*

$$(1)\ D^{c_1}(i_1, j_1) \geq D^{c_2}(i_1, j_1)\frac{L_d^{c_1}}{L_d^{c_2}} \Leftrightarrow D^{c_1}(i_2, j_2) \geq D^{c_2}(i_2, j_2)\frac{L_d^{c_1}}{L_d^{c_2}}$$
$$(2)\ D^{c_1}(i_1, j_1) \leq D^{c_2}(i_1, j_1)\frac{L_d^{c_1}}{L_d^{c_2}} \Leftrightarrow D^{c_1}(i_2, j_2) \leq D^{c_2}(i_2, j_2)\frac{L_d^{c_1}}{L_d^{c_2}}$$

where $L_d^c = q_0^c \int_\Lambda L_d(\lambda)\tau^c(\lambda)s(\lambda)d\lambda$ is the (scaled) radiance of the directional light source for colour channel c.[1]

[1] The procedure of multiplying colour channels by the quotient $L_d^{c_1}/L_d^{c_2}$ has the effect of whitening the image. For instance, in RGB space: (In the following, co-ordinates (i, j) will be omitted to simplify the notation.)

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = m_b \begin{pmatrix} R_b \\ G_b \\ B_b \end{pmatrix} + m_i \begin{pmatrix} R_i \\ G_i \\ B_i \end{pmatrix}$$

Now, using the mean-value theorem over equation 4.4b (assume the continuity conditions required are satisfied), $R_b = \overline{\rho}_b^R \left[q_0^R \int_\Lambda L_d(\lambda)\tau^R(\lambda)s(\lambda)d\lambda\right]$, where $\overline{\rho}_b^R = \rho_b(\lambda_0)$ for some $\lambda_0 \in \Lambda$. In this way, R_b can be put as $\overline{\rho}_b^R L_d^R$, and, analogously, $G_b = \overline{\rho}_b^G L_d^G$ and $B_b = \overline{\rho}_b^B L_d^B$. If the same is applied to the interface components, then:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = m_b \begin{pmatrix} \overline{\rho}_b^R L_d^R \\ \overline{\rho}_b^G L_d^G \\ \overline{\rho}_b^B L_d^B \end{pmatrix} + m_i \begin{pmatrix} \overline{\rho}_i^R L_d^R \\ \overline{\rho}_i^G L_d^G \\ \overline{\rho}_i^B L_d^B \end{pmatrix}$$

Therefore:

$$\begin{pmatrix} R \\ G\frac{L_d^R}{L_d^G} \\ B\frac{L_d^R}{L_d^B} \end{pmatrix} = m_b \begin{pmatrix} \overline{\rho}_b^R L_d^R \\ \overline{\rho}_b^G L_d^R \\ \overline{\rho}_b^B L_d^R \end{pmatrix} + m_i \begin{pmatrix} \overline{\rho}_i^R L_d^R \\ \overline{\rho}_i^G L_d^R \\ \overline{\rho}_i^B L_d^R \end{pmatrix}$$

That is to say, the image intensities no longer depend on the colour of the illuminant, as if it was white-coloured (i.e. $L(\lambda) = L, \forall\lambda$).

This effect can be achieved in real cameras through the function known as *white balance*. Making use of this function, Stokman obtained invariance against specularities assuming the NIR model [262].

In order for the previous proposition to be applicable, L_d^c must thus be estimated for every colour channel c for the particular scene prior to its segmentation. Any of the methods presented in chapter 5 can be used for this purpose.

As for property 2, it is no longer true in general when interface reflection is incorporated into the reflection model. This is because the derivative of the intensity in every channel not only depends now on the sign of the derivative of m_b but also on the derivative of m_i:

$$\left(\frac{dD^{c_1}(i,j)}{d\xi}\right) = m_b'(i,j)C_b^{c_1} + m_i'(i,j)C_i^{c_1}$$
$$\left(\frac{dD^{c_2}(i,j)}{d\xi}\right) = m_b'(i,j)C_b^{c_2} + m_i'(i,j)C_i^{c_2},$$

where $m_b'(i,j)$ and $m_i'(i,j)$ are the first-order derivatives along direction ξ of $m_b(i,j)$ and $m_i(i,j)$, respectively. As a consequence, in general:

$$\begin{array}{ccc} m_b'(i,j)C_b^{c_1} + m_i'(i,j)C_i^{c_1} \lesseqqgtr 0 & & m_b'(i,j)C_b^{c_2} + m_i'(i,j)C_i^{c_2} \lesseqqgtr 0 \\ \Updownarrow & & \Updownarrow \\ m_b'(i,j)C_b^{c_1} \lesseqqgtr -m_i'(i,j)C_i^{c_1} & \not\Leftrightarrow & m_b'(i,j)C_b^{c_2} \lesseqqgtr -m_i'(i,j)C_i^{c_2} \end{array}$$

Observe that the derivatives of both channels c_1 and c_2 will have the same sign if m_b' and m_i' are both positive or both negative, or, in case their signs do not coincide, if the derivative of the same reflection component dominates the sum in both cases (i.e. $|m_b'(i,j)C_b^{c_1}| \lesseqqgtr |m_i'(i,j)C_i^{c_1}|$ and $|m_b'(i,j)C_b^{c_2}| \lesseqqgtr |m_i'(i,j)C_i^{c_2}|$). By way of example, in the intensity profiles of figure 8.2, the dotted lines enclose an area where m_b' and m_i' have different sign, but the derivatives of every colour channel have all the same sign.

Finally, as for property 3, figure 8.2 shows precisely a case where it is not met. In effect, the intensity levels of the green and blue channels within the area enclosed by the dotted lines are such that, when the intensity increases in both, the difference between both channels does not increase, but decreases.

8.2.4 Case 4: glossy pixels/with ambient lighting

In the general case:

$$D^c(i,j) = C_a^c(i,j) + m_b(i,j)C_b^c(i,j) + m_i(i,j)C_i^c(i,j)\,. \tag{8.4}$$

With the incorporation of the interface reflection term, only property 1 holds after the changes in the formulation already introduced in case 3. The NIR model is again assumed, together with an ambient illumination proportional to the directional lighting. The following lemma and proposition states the aforementioned in a formal way: (See appendix B for the formal proofs as before.)

Lemma 8.9. *If ambient illumination is proportional to directional illumination (i.e. $L_a(\lambda) = \alpha L_d(\lambda)$ $[\alpha \geq 0], \forall\lambda$) and if the NIR formation model is assumed, for any material and any two colour channels c_1 and c_2:*

$$(1)\ C_a^{c_1} - C_a^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}} \geq 0 \Leftrightarrow C_b^{c_1} - C_b^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}} \geq 0$$

$$(2)\ C_a^{c_1} - C_a^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}} \leq 0 \Leftrightarrow C_b^{c_1} - C_b^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}} \leq 0$$

where $L_d^c = q_0^c \int_\Lambda L_d(\lambda)\tau^c(\lambda)s(\lambda)d\lambda$ is the (scaled) radiance of directional illumination for colour channel c.

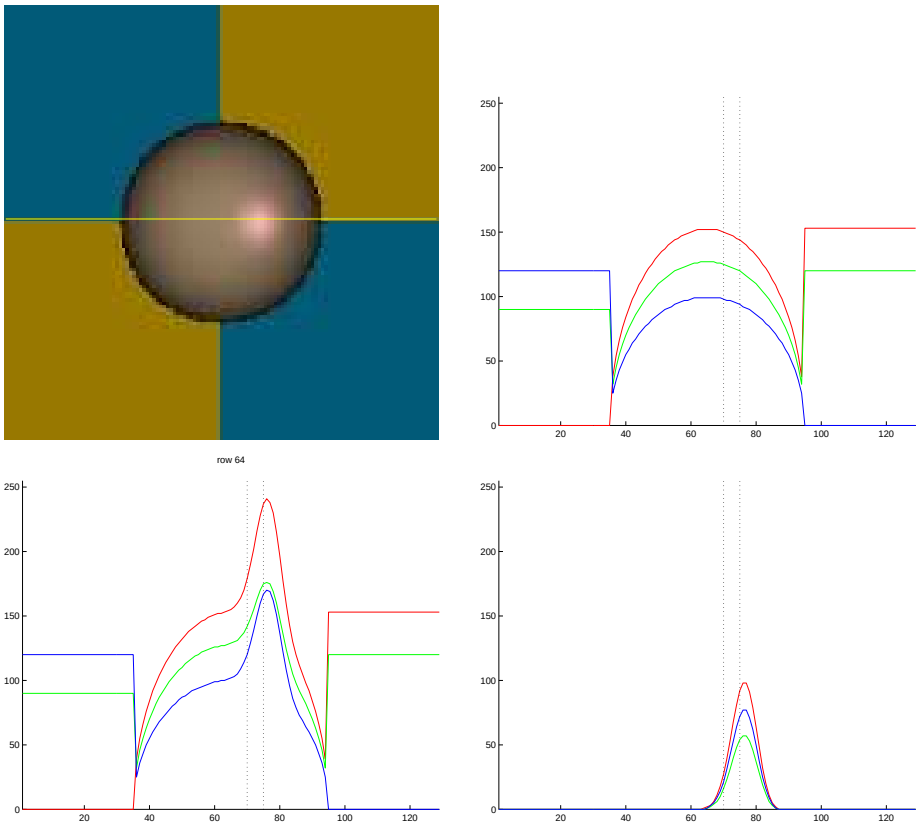


Fig. 8.2. [top,left] A synthetic image without ambient lighting. [bottom,left] Intensity profile for row 64 (in yellow) for every colour channel. [top,right] Body component profiles. [bottom,right] Interface component profiles.

Proposition 8.10. *Assuming directional illumination is uniform throughout the scene and that ambient illumination is proportional to directional illumination, and assuming the NIR formation model, for any pair of colour channels c_1 and c_2, and any two image locations (i_1, j_1) and (i_2, j_2) coming from the same scene material,*

$$(1)\ D^{c_1}(i_1,j_1) \geq D^{c_2}(i_1,j_1)\frac{L_d^{c_1}}{L_d^{c_2}} \Leftrightarrow D^{c_1}(i_2,j_2) \geq D^{c_2}(i_2,j_2)\frac{L_d^{c_1}}{L_d^{c_2}}$$

$$(2)\ D^{c_1}(i_1,j_1) \leq D^{c_2}(i_1,j_1)\frac{L_d^{c_1}}{L_d^{c_2}} \Leftrightarrow D^{c_1}(i_2,j_2) \leq D^{c_2}(i_2,j_2)\frac{L_d^{c_1}}{L_d^{c_2}}$$

where $L_d^c = q_0^c \int_\Lambda L_d(\lambda)\tau^c(\lambda)s(\lambda)d\lambda$ is the (scaled) radiance of directional illumination for colour channel c.

As for properties 2 and 3, the same remarks as for case 3 apply here.

8.3 Usefulness of the Properties

By way of recapitulation of section 8.2, table 8.1 summarizes the dependence of the fulfillment of properties 1-3 on the particular instantiation of the image formation model. Some conclusions about the applicability of those properties can be drawn now that the four cases have been revised:

- When the interface reflection term is introduced, only property 1 is satisfied in general within uniform reflectance areas. This means that false material changes can be detected sometimes inside

Table 8.1. Fulfillment of properties 1-3 depending on the particular instantiation of the image formation model.

case	properties fulfilled and conditions
(1) matte pixels, no ambient lighting	properties 1-3 always
(2) matte pixels, ambient lighting	property 2 always; properties 1 and 3 if $L_a(\lambda) = \alpha L_d(\lambda), \alpha > 0$
(3) glossy pixels, no ambient lighting	property 1 under the NIR model and white-balanced images
(4) glossy pixels, ambient lighting	property 1 under the NIR model, white-balanced images and $L_a(\lambda) = \alpha L_d(\lambda), \alpha > 0$

or near specularities. As it has been discussed before, it depends on whether the sign of m'_b or m'_i are the same or not, and on which term dominates the value of the derivative: the body/ambient reflection terms or the interface reflection term. Therefore, not always will specularities give rise to false material changes.

- Properties 1 and 3 require that L_a is proportional to L_d in order for them to be applicable when ambient lighting is not negligible. As was said in section 2.3, the ambient lighting comes, in an outdoor environment, from the sky through diffraction in the atmosphere, or, in a room, from inter-reflections on the walls and other objects. Given these facts, it is obvious that some relationship can be expected between ambient lighting and directional lighting. Unfortunately, no evidence about the proportionality relationship has been found in the related literature, apart from other authors also making use of it [227]. However, the experiments performed have shown that, in general, violations of properties 1 and 3 are closely related with reflectance transitions.

8.4 Cases which Fail to Satisfy the Properties

The following propositions characterize which are the circumstances under which properties 1-3 are not fulfilled: (As before, formal proofs are given in appendix B.)

Proposition 8.11. *Assuming directional illumination is uniform throughout the scene and that ambient illumination is proportional to directional illumination, and assuming the NIR formation model, for any pair of colour channels c_1 and c_2 and any two image locations (i_1, j_1) and (i_2, j_2):*

$$(1) \quad \left.\begin{array}{l} D^{c_1}(i_1,j_1) > D^{c_2}(i_1,j_1)\frac{L_d^{c_1}}{L_d^{c_2}} \\ D^{c_1}(i_2,j_2) < D^{c_2}(i_2,j_2)\frac{L_d^{c_1}}{L_d^{c_2}} \end{array}\right\} \Leftrightarrow \left\{\begin{array}{l} C_b^{c_1}(i_1,j_1) > C_b^{c_2}(i_1,j_1)\frac{L_d^{c_1}}{L_d^{c_2}} \\ C_b^{c_1}(i_2,j_2) < C_b^{c_2}(i_2,j_2)\frac{L_d^{c_1}}{L_d^{c_2}} \end{array}\right.$$

$$(2) \quad \left.\begin{array}{l} D^{c_1}(i_1,j_1) < D^{c_2}(i_1,j_1)\frac{L_d^{c_1}}{L_d^{c_2}} \\ D^{c_1}(i_2,j_2) > D^{c_2}(i_2,j_2)\frac{L_d^{c_1}}{L_d^{c_2}} \end{array}\right\} \Leftrightarrow \left\{\begin{array}{l} C_b^{c_1}(i_1,j_1) < C_b^{c_2}(i_1,j_1)\frac{L_d^{c_1}}{L_d^{c_2}} \\ C_b^{c_1}(i_2,j_2) > C_b^{c_2}(i_2,j_2)\frac{L_d^{c_1}}{L_d^{c_2}} \end{array}\right.$$

where $L_d^c = q_0^c \int_\Lambda L_d(\lambda)\tau^c(\lambda)s(\lambda)d\lambda$ is the (scaled) radiance coming from the directional light source for colour channel c.

Proposition 8.12. *Assuming directional illumination is uniform throughout the scene and that ambient illumination is proportional to directional illumination, for any pair of colour channels c_1 and c_2 and any two image locations (i_1, j_1) and (i_2, j_2) such that $D_i^c(i_1, j_1) = 0$ and $D_i^c(i_2, j_2) = 0$,*

$$(1)\quad \left.\begin{array}{l} D^{c_1}(i_1,j_1) - D^{c_1}(i_2,j_2) > 0 \\ D^{c_2}(i_1,j_1) - D^{c_2}(i_2,j_2) < 0 \end{array}\right\} \Rightarrow \frac{C_b^{c_2}(i_1,j_1)}{C_b^{c_1}(i_1,j_1)} < \frac{C_b^{c_2}(i_2,j_2)}{C_b^{c_1}(i_2,j_2)}$$

$$(2)\quad \left.\begin{array}{l} D^{c_1}(i_1,j_1) - D^{c_1}(i_2,j_2) < 0 \\ D^{c_2}(i_1,j_1) - D^{c_2}(i_2,j_2) > 0 \end{array}\right\} \Rightarrow \frac{C_b^{c_2}(i_1,j_1)}{C_b^{c_1}(i_1,j_1)} > \frac{C_b^{c_2}(i_2,j_2)}{C_b^{c_1}(i_2,j_2)}$$

Proposition 8.13. *Assuming directional illumination is uniform throughout the scene and that ambient illumination is proportional to directional illumination, for any pair of colour channels c_1 and c_2 and any two image locations (i_1, j_1) and (i_2, j_2) such that $D_i^c(i_1, j_1) = 0$ and $D_i^c(i_2, j_2) = 0$,*

$$(1)\quad \left.\begin{array}{l} D^{c_1}(i_1,j_1) \geq D^{c_1}(i_2,j_2) \text{ and } D^{c_2}(i_1,j_1) \geq D^{c_2}(i_2,j_2) \\ D^{c_1}(i_1,j_1) \geq D^{c_2}(i_1,j_1) \text{ and } D^{c_1}(i_2,j_2) \geq D^{c_2}(i_2,j_2) \\ D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1) < D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2) \end{array}\right\}$$

$$\Rightarrow \left\{\begin{array}{l} C_b^{c_1}(i_1,j_1) > C_b^{c_1}(i_2,j_2) \\ C_b^{c_2}(i_1,j_1) > C_b^{c_2}(i_2,j_2) \end{array}\right.$$

$$(2)\quad \left.\begin{array}{l} D^{c_1}(i_1,j_1) \leq D^{c_1}(i_2,j_2) \text{ and } D^{c_2}(i_1,j_1) \leq D^{c_2}(i_2,j_2) \\ D^{c_1}(i_1,j_1) \leq D^{c_2}(i_1,j_1) \text{ and } D^{c_1}(i_2,j_2) \leq D^{c_2}(i_2,j_2) \\ D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1) < D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2) \end{array}\right\}$$

$$\Rightarrow \left\{\begin{array}{l} C_b^{c_1}(i_1,j_1) < C_b^{c_1}(i_2,j_2) \\ C_b^{c_2}(i_1,j_1) < C_b^{c_2}(i_2,j_2) \end{array}\right.$$

$$(3)\quad \left.\begin{array}{l} D^{c_1}(i_1,j_1) \geq D^{c_1}(i_2,j_2) \text{ and } D^{c_2}(i_1,j_1) \geq D^{c_2}(i_2,j_2) \\ D^{c_1}(i_1,j_1) \leq D^{c_2}(i_1,j_1) \text{ and } D^{c_1}(i_2,j_2) \leq D^{c_2}(i_2,j_2) \\ D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1) > D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2) \end{array}\right\}$$

$$\Rightarrow \left\{\begin{array}{l} C_b^{c_1}(i_1,j_1) > C_b^{c_1}(i_2,j_2) \\ C_b^{c_2}(i_1,j_1) > C_b^{c_2}(i_2,j_2) \end{array}\right.$$

$$(4)\quad \left.\begin{array}{l} D^{c_1}(i_1,j_1) \leq D^{c_1}(i_2,j_2) \text{ and } D^{c_2}(i_1,j_1) \leq D^{c_2}(i_2,j_2) \\ D^{c_1}(i_1,j_1) \geq D^{c_2}(i_1,j_1) \text{ and } D^{c_1}(i_2,j_2) \geq D^{c_2}(i_2,j_2) \\ D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1) > D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2) \end{array}\right\}$$

$$\Rightarrow \left\{\begin{array}{l} C_b^{c_1}(i_1,j_1) < C_b^{c_1}(i_2,j_2) \\ C_b^{c_2}(i_1,j_1) < C_b^{c_2}(i_2,j_2) \end{array}\right.$$

These last propositions deserve several observations:

- First of all, notice that every proposition identifies situations for which the composite body reflectances corresponding to image locations (i_1, j_1) and (i_2, j_2) must necessarily be different due to failure to fulfill one of properties 1-3. Although the reflectance changes are depicted by means of the respective composite body reflectances $C_b^{c_1}$ and $C_b^{c_2}$, they can also be expressed in terms of ρ_b as follows. Through the mean value theorem, the composite body reflectance for a certain colour channel c can be put as a product of two terms:

$$\begin{aligned} C_b^c(i,j) &= q_0^c \int_\Lambda L_d(\lambda)\rho_b(i,j;\lambda)\tau^c(\lambda)s(\lambda)\ d\lambda \\ &= \rho_b(i,j;\lambda_0)\left[q_0^c \int_\Lambda L_d(\lambda)\tau^c(\lambda)s(\lambda)d\lambda\right], \end{aligned}$$

Table 8.2. Reflectance transitions expressed by propositions 8.11-8.13.

proposition	reflectance transition	
8.11	CHannel Crossing (CHC)	$\rho_b^{c_1}(i_1,j_1) > \rho_b^{c_2}(i_1,j_1)$ $\rho_b^{c_1}(i_2,j_2) < \rho_b^{c_2}(i_2,j_2)$
		$\rho_b^{c_1}(i_1,j_1) < \rho_b^{c_2}(i_1,j_1)$ $\rho_b^{c_1}(i_2,j_2) > \rho_b^{c_2}(i_2,j_2)$
8.12	Non-Coinciding Derivative (NCD)	$\frac{\rho_b^{c_2}(i_1,j_1)}{\rho_b^{c_1}(i_1,j_1)} < \frac{\rho_b^{c_2}(i_2,j_2)}{\rho_b^{c_1}(i_2,j_2)}$
		$\frac{\rho_b^{c_2}(i_1,j_1)}{\rho_b^{c_1}(i_1,j_1)} > \frac{\rho_b^{c_2}(i_2,j_2)}{\rho_b^{c_1}(i_2,j_2)}$
8.13	Non-Decreasing Difference (NDD)	$\rho_b^{c_1}(i_1,j_1) > \rho_b^{c_1}(i_2,j_2)$ $\rho_b^{c_2}(i_1,j_1) > \rho_b^{c_2}(i_2,j_2)$
		$\rho_b^{c_1}(i_1,j_1) < \rho_b^{c_1}(i_2,j_2)$ $\rho_b^{c_2}(i_1,j_1) < \rho_b^{c_2}(i_2,j_2)$

for some $\lambda_0 \in \Lambda$. Therefore, if $\rho_b^c(i,j) = \rho_b(i,j;\lambda_0)$ is defined for every colour channel and image location[2], then the three propositions correspond to the reflectance transitions enumerated in table 8.2. (The relationships between ρ_b^c's which are given are immediate from the proofs of the corresponding propositions.)

- Regarding proposition 8.11, cases (1) and (2) of its enunciate provide the only two ways to fail to fulfill property1: the order between two colour channels intensity is reversed from location (i_1,j_1) to location (i_2,j_2). If those location happen to be adjacent to one another, then a reflectance transition of the type indicated in table 8.2 necessarily exists between (i_1,j_1) and (i_2,j_2), which materializes as a colour channel crossing in-between both image locations.
  Besides, observe that the proposition is written as a double implication. In order to illustrate this fact, let us consider, by way of example, case (1) of the proposition. In this case, channel c_1 is "above" channel c_2 at (i_1,j_1) (i.e. the intensity in channel c_1 is larger than the intensity in channel c_2), while at (i_2,j_2) is just the opposite. The proposition says that, provided that the hypotheses hold, this order relationship takes place if and only if the same relationship holds between the corresponding composite reflectances. Both facts are thus equivalent.
  Finally, another version of proposition 8.11 which does not need estimates for L_d^c nor the NIR model, at the expense of not tolerating specularities, can be found below in the form of proposition 8.14: (As before, the proof is in appendix B.)

  Proposition 8.14. *Assuming directional illumination is uniform throughout the scene and that ambient illumination is proportional to directional illumination, for any pair of colour channels c_1 and c_2 and any two image locations (i_1,j_1) and (i_2,j_2) such that $D_i^c(i_1,j_1) = 0$ and $D_i^c(i_2,j_2) = 0$,*

[2] Observe that $\rho_b^c(i,j) = \rho_b(i,j;\lambda_0)$ is the same for every (i,j) corresponding to the same scene material, provided that the illumination is uniform throughout the scene. Therefore, it is in fact a property of the material, although just for that particular scene. In effect, contrary to what can seem at first sight, $\rho_b^c(i,j)$ is not really illumination-independent since, in fact:

$$\rho_b^c(i,j) = \frac{\int_\Lambda L_d(\lambda)\rho_b(i,j;\lambda)\tau^c(\lambda)s(\lambda)\ d\lambda}{\int_\Lambda L_d(\lambda)\tau^c(\lambda)s(\lambda)d\lambda} \tag{8.5}$$

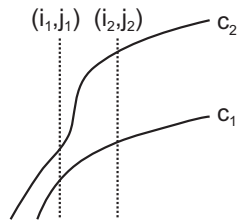


Fig. 8.3. A case where possibly $\frac{C_b^{c_2}(i_1,j_1)}{C_b^{c_1}(i_1,j_1)} < \frac{C_b^{c_2}(i_2,j_2)}{C_b^{c_1}(i_2,j_2)}$, but $D^{c_1}(i_1,j_1) - D^{c_1}(i_2,j_2) < 0$ and $D^{c_2}(i_1,j_1) - D^{c_2}(i_2,j_2) < 0$.

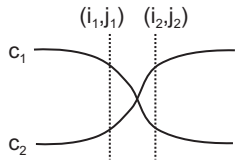


Fig. 8.4. A CHC and an NCD transition taking place simultaneously.

$$(1)\quad \left.\begin{array}{l} D^{c_1}(i_1,j_1) > D^{c_2}(i_1,j_1) \\ D^{c_1}(i_2,j_2) < D^{c_2}(i_2,j_2) \end{array}\right\} \Leftrightarrow \left\{\begin{array}{l} C_b^{c_1}(i_1,j_1) > C_b^{c_2}(i_1,j_1) \\ C_b^{c_1}(i_2,j_2) < C_b^{c_2}(i_2,j_2) \end{array}\right.$$

$$(2)\quad \left.\begin{array}{l} D^{c_1}(i_1,j_1) < D^{c_2}(i_1,j_1) \\ D^{c_1}(i_2,j_2) > D^{c_2}(i_2,j_2) \end{array}\right\} \Leftrightarrow \left\{\begin{array}{l} C_b^{c_1}(i_1,j_1) < C_b^{c_2}(i_1,j_1) \\ C_b^{c_1}(i_2,j_2) > C_b^{c_2}(i_2,j_2) \end{array}\right.$$

- Proposition 8.12 provides the conditions which characterize a reflectance transition for which at least two channels are forced to evolve without coordination; that is to say, when the intensity in one increases, in the other one decreases (transition of type *non-coinciding derivative* or NCD). Analogously to proposition 8.11, the cases enumerated in proposition 8.12 correspond to the two ways how property 2 can be unsatisfied.
  On the other hand, notice that the formulation of proposition 8.12 does not make use of a derivative-like notation, unlike the corresponding property 2, but utilizes a notation based on two points (i_1, j_1) and (i_2, j_2) like the other two propositions. This is to elude the problem of the non-existence of intensity derivatives at reflectance transition locations.
  Observe that the proposition is written as a single implication. This is because the relationships between composite body reflectances expressed through the consequent of the proposition apply also to other cases where the colour channels do evolve in a coordinated way (see figure 8.3 for an illustration of this point). Therefore, both facts are not equivalent.
  Finally, notice that there is a certain degree of overlapping between CHC and NCD transitions. By way of example, observe in figure 8.4 that every channel crossing implies two colour channels evolving without coordination during the reflectance transition. In case it was necessary to avoid the overlapping, an additional condition should be added to the two cases of proposition 8.12: $D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1) \geq 0$ and $D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2) \geq 0$, or $D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1) \leq 0$ and $D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2) \leq 0$.
- In the third place, proposition 8.13 characterizes a reflectance transition during which property 3 does not hold. Therefore, for at least two channels, the difference between their intensities does not decrease as the intensity in both decreases (transition of type *non-decreasing difference* or NDD). As it can be seen, the formulation of proposition 8.13 is a bit special in the sense that the conditions that constitute the antecedents of the different implications include more terms than the conditions appearing in the formulation of property 3. This is because they are needed in the proof, although they also avoid a certain degree of overlapping between propositions 8.11/8.12 and proposition 8.13. In effect, by way of example, in case (1) of proposition 8.13, $D^{c_1}(i_1,j_1) \geq D^{c_1}(i_2,j_2)$ and $D^{c_2}(i_1,j_1) \geq D^{c_2}(i_2,j_2)$ are both needed simultaneously because, if they were not

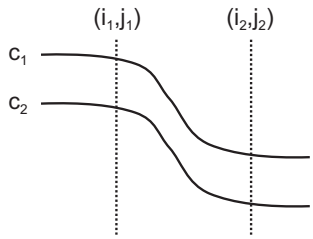


Fig. 8.5. A case where possibly $C_b^{c_1}(i_1, j_1) > C_b^{c_1}(i_2, j_2)$ and $C_b^{c_2}(i_1, j_1) > C_b^{c_2}(i_2, j_2)$, but $D^{c_1}(i_1, j_1) \geq D^{c_1}(i_2, j_2)$, $D^{c_2}(i_1, j_1) \geq D^{c_2}(i_2, j_2)$, $D^{c_1}(i_1, j_1) \geq D^{c_2}(i_1, j_1)$, $D^{c_1}(i_2, j_2) \geq D^{c_2}(i_2, j_2)$, and $D^{c_1}(i_1, j_1) - D^{c_2}(i_1, j_1) \geq D^{c_1}(i_2, j_2) - D^{c_2}(i_2, j_2)$.

fulfilled at the same time, colour channels c_1 and c_2 would not evolve coordinately, and a transition of type NCD would have been found. Analogously, $D^{c_1}(i_1, j_1) \geq D^{c_2}(i_1, j_1)$ and $D^{c_1}(i_2, j_2) \geq D^{c_2}(i_2, j_2)$ are also necessary in order to ensure a transition of type CHC is not taking place. These two extra conditions do not appear in the formulation of property 3 because, as for every property both pixels (i_1, j_1) and (i_2, j_2) are assumed to have a common reflectance, $D^{c_1}(i_1, j_1) \geq D^{c_1}(i_2, j_2)$ implies $D^{c_1}(i_1, j_1) \geq D^{c_1}(i_2, j_2)$ (property 2) and $D^{c_1}(i_1, j_1) \geq D^{c_2}(i_1, j_1)$ implies $D^{c_1}(i_2, j_2) \geq D^{c_2}(i_2, j_2)$ (property 1).

Finally, proposition 8.13 is not a double implication because, as in the case of proposition 8.12, the sort of reflectance transitions which are consequence of the conditions stated in the proposition can also arise due to other circumstances (see figure 8.5 for an illustration).

8.5 Formulation of a Compatibility Relationship between Pixels

Properties 1-3 can be said to lead to a set of necessary compatibility conditions between pixels, in the sense that if two pixels fail to satisfy any of the three properties for any of the two-by-two combinations between colour channels, then both pixels cannot correspond to the same scene material. In this section, all three compatibility conditions will be unified in a single compatibility relationship between image pixels by means of their geometric interpretation.

In effect, given a certain combination of colour channels, say c_1 and c_2, and two pairs, p and q, of intensity values, $(D_p^{c_1}, D_p^{c_2})$ and $(D_q^{c_1}, D_q^{c_2})$, all three properties can be stated geometrically considering the space of intensity values taken by both channels:

- $(D_p^{c_1}, D_p^{c_2})$, with $D_p^{c_1} \geq D_p^{c_2}$, and $(D_q^{c_1}, D_q^{c_2})$ satisfy property 1 if and only if $D_q^{c_1} \geq D_q^{c_2}$. This is expressed in figure 8.6(a) as the shaded area below the straight line $D^{c_1} = D^{c_2}$. In case $D_p^{c_1} \leq D_p^{c_2}$, the same order relationship must take place in the pair $(D_q^{c_1}, D_q^{c_2})$ for them to be compatible in the sense of property 1. This corresponds to the complement of the shaded area in figure 8.6(a). Notice that if $D_p^{c_1} = D_p^{c_2}$, any combination of intensity values for channels c_1 and c_2 is compatible with $(D_p^{c_1}, D_p^{c_2})$. This agrees with the aim of developing a set of necessary compatibility conditions, since, in this particular case, the pair $(D_p^{c_1}, D_p^{c_2})$ does not provide any clue about the order relationship between colour channels for the corresponding material, and therefore no reflectance transition should be suspected whatever $(D_q^{c_1}, D_q^{c_2})$ could be.
- $(D_p^{c_1}, D_p^{c_2})$ and $(D_q^{c_1}, D_q^{c_2})$ satisfy property 2 if intensity varies in the same sense for both channels from a to b: it increases in both, $D_p^{c_1} - D_q^{c_1} \geq 0$ and $D_p^{c_2} - D_q^{c_2} \geq 0$, or it decreases in both, $D_p^{c_1} - D_q^{c_1} \leq 0$ and $D_p^{c_2} - D_q^{c_2} \leq 0$. In figure 8.6(b), the shaded area include all the pairs $(D_q^{c_1}, D_q^{c_2})$ which verify these conditions for $(D_p^{c_1}, D_p^{c_2})$ fixed. Observe that, in this case, the compatibility zone has nothing to do with the order relationship between colour channels.
- Finally, $(D_p^{c_1}, D_p^{c_2})$, with $D_p^{c_1} \geq D_p^{c_2}$, and $(D_q^{c_1}, D_q^{c_2})$ are such that an increase in intensity in one channel corresponds to an increase in the difference between colour channels, or, inversely, a decrease in intensity corresponds to a decrease in the difference (property 3) if:

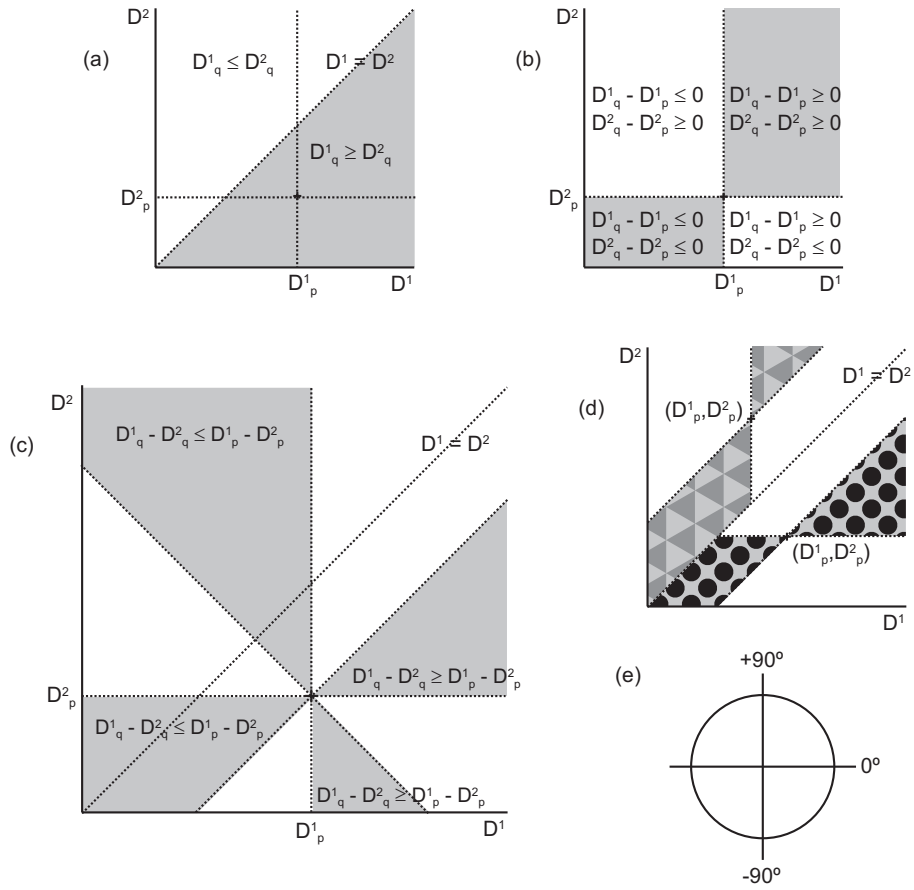


Fig. 8.6. Geometrical interpretation of properties 1-3. (In the figures, D_k^n is an abbreviation of $D_k^{c_n}$.)

(i) $D_q^{c_1} - D_q^{c_2} \geq D_p^{c_1} - D_p^{c_2}$ when $D_q^{c_1} \geq D_p^{c_1}$ and $D_q^{c_2} \geq D_p^{c_2}$ (upper right shaded area of figure 8.6(c)), or

(ii) $D_q^{c_1} - D_q^{c_2} \leq D_p^{c_1} - D_p^{c_2}$ when $D_q^{c_1} \leq D_p^{c_1}$ and $D_q^{c_2} \leq D_p^{c_2}$ (lower left shaded area of figure 8.6(c)), or

(iii) $D_q^{c_1} - D_q^{c_2} \leq D_p^{c_1} - D_p^{c_2}$ when $D_q^{c_1} \leq D_p^{c_1}$ and $D_q^{c_2} \geq D_p^{c_2}$ (upper left shaded area of figure 8.6(c)), or

(iv) $D_q^{c_1} - D_q^{c_2} \geq D_p^{c_1} - D_p^{c_2}$ when $D_q^{c_1} \geq D_p^{c_1}$ and $D_q^{c_2} \leq D_p^{c_2}$ (lower right shaded area of figure 8.6(c)).

As well as for property 1, in case $D_p^{c_1} \leq D_p^{c_2}$, the compatibility area is the complement of the shaded area of figure 8.6(c). For instance, if $D_q^{c_1} \geq D_p^{c_1}$ and $D_q^{c_2} \geq D_p^{c_2}$, property 3 holds if $D_q^{c_2} - D_q^{c_1} \geq D_p^{c_2} - D_p^{c_1}$ (upper right non-shaded area of figure 8.6(c)). When $D_p^{c_1} = D_p^{c_2}$, as before, the compatibility area is the whole space of values taken by channels c_1 and c_2.
To finish, observe that, contrary to the original formulation of property 3, this geometrical interpretation does not take into account properties 1 and 2. In fact, notice that cases (iii) and (iv) above violate property 2.

To finish, figure 8.6(d) shows the compatible area for all three properties simultaneously (triangles correspond to the case $D_p^{c_1} \leq D_p^{c_2}$, while circles are for case $D_p^{c_1} \geq D_p^{c_2}$). In view of these graphs, a compatibility relation $\mathcal{C}$ can be summarized as:

$(D_q^{c_1}, D_q^{c_2})$ *is* $\mathcal{C}$*-compatible with* $(D_p^{c_1}, D_p^{c_2})$ *if:*
(i) when $D_p^{c_1} \geq D_p^{c_2}$*, the orientation of the vector joining both lies within* $[0°, 45°]$*; and*
(ii) when $D_p^{c_1} \leq D_p^{c_2}$*, the orientation of the vector joining both lies within* $[45°, 90°]$*.*

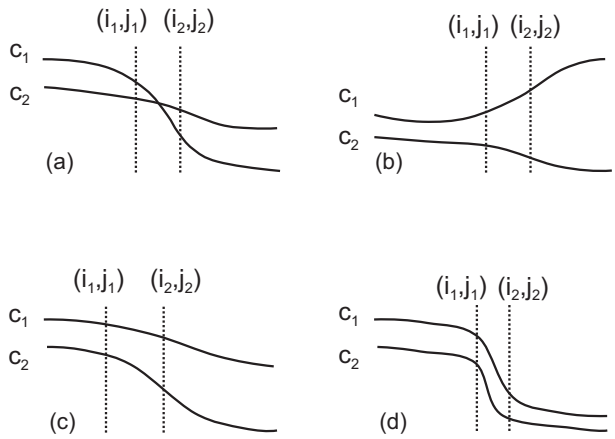


Fig. 8.7. Summary of how the body reflectance transitions dectected manifest and corresponding edges: (a) CHC edge; (b) NCD edge; (c) NDD edge; (d) GRD edge.

This formulation of $\mathcal{C}$ assumes that the origin of the vector joining $(D_p^{c_1}, D_p^{c_2})$ and $(D_q^{c_1}, D_q^{c_2})$ is chosen so that its orientation lies between $-90°$ and $90°$ (see figure 8.6(e) for the angle convention).

8.6 Coverage of Reflectance Transitions

The compatibility relationship $\mathcal{C}$ formulated above covers a broad spectrum of body reflectance changes. On the one hand, if property 1 is infringed, it is because at least one colour channel crosses one another. Otherwise, at that image pixel, colour channels do not cross one another, but, still, property 2 can be violated. In such a case, at least two colour channels diverge in the sense that, while in one channel the intensity increases, at the other one decreases. In case the intensity at all colour channels vary in a coordinated way —all increase or all decrease—, it can so happen that property 3 is not fulfilled; that is to say, when intensity decreases in both channels, the difference between them increases instead of decreasing, or vice versa. Figures 8.7(a)-(c) summarize graphically these three types of reflectance transitions.

By way of example, CHC reflectance transitions (see table 8.2) comprise, from a theoretical point of view, the 83.33 % of all possible transitions (see appendix C for the derivation of this percentage).What is most important from the previous fact is that $\mathcal{C}$ can be used to detect such an amount of reflectance transitions without resorting to explicitly computing the reflectance of the underlying surface.

However, unfortunately, there exist some reflectance transitions such that none of the cases of propositions 8.11-8.14 is fulfilled despite a reflectance change is arising there. These reflectance changes make manifest such as if the pixels involved corresponded all to the same scene material in the sense of not violating any of properties 1-3. Figures 8.3 and 8.5 are examples of this sort of reflectance transitions.

To investigate further this point, let us consider the situation of figure 8.3. In this case, the intensity increases in both channels, channel c_2 is "above" channel c_1 and the difference between c_2 and c_1 increases:

$$
\begin{aligned}
&D^{c_1}(i_1, j_1) < D^{c_1}(i_2, j_2)\\
&D^{c_2}(i_1, j_1) < D^{c_2}(i_2, j_2)\\
&D^{c_2}(i_1, j_1) > D^{c_1}(i_1, j_1)\\
&\qquad \Rightarrow D^{c_2}(i_1, j_1) - D^{c_1}(i_1, j_1) = (\alpha + m_b(i_1, j_1))\,(C_b^{c_2}(i_1, j_1) - C_b^{c_1}(i_1, j_1)) > 0
\end{aligned}
$$

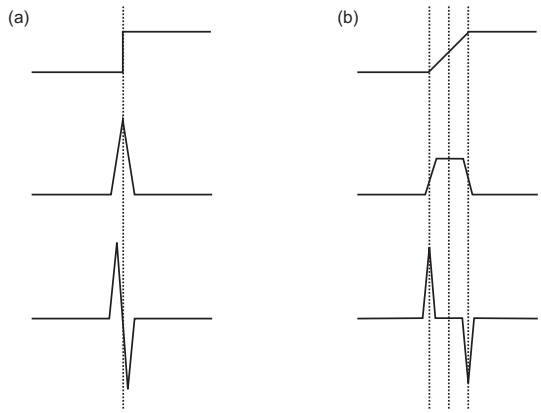


Fig. 8.8. Second-order derivative output for a step edge (a) and a ramp edge (b).

$$
\begin{aligned}
&D^{c_2}(i_2,j_2) > D^{c_1}(i_2,j_2)\\
&\qquad \Rightarrow D^{c_2}(i_2,j_2) - D^{c_1}(i_2,j_2) = (\alpha + m_b(i_2,j_2))\,(C_b^{c_2}(i_2,j_2) - C_b^{c_1}(i_2,j_2)) > 0\\
&D^{c_2}(i_1,j_1) - D^{c_1}(i_1,j_1) < D^{c_2}(i_2,j_2) - D^{c_1}(i_2,j_2)\\
&\qquad \Rightarrow (\alpha + m_b(i_1,j_1))\,\underbrace{(C_b^{c_2}(i_1,j_1) - C_b^{c_1}(i_1,j_1))}_{>0} < (\alpha + m_b(i_2,j_2))\,\underbrace{(C_b^{c_2}(i_2,j_2) - C_b^{c_1}(i_2,j_2))}_{>0}
\end{aligned}
$$

Observe that this last inequality can be satisfied under, among others, the two following sets of circumstances:

- $m_b(i_1,j_1) = m_b(i_2,j_2)$ and $C_b^{c_2}(i_1,j_1) - C_b^{c_1}(i_1,j_1) < C_b^{c_2}(i_2,j_2) - C_b^{c_1}(i_2,j_2)$, which means the changes in intensity $D^{c_1}(i_1,j_1) - D^{c_1}(i_2,j_2)$ and $D^{c_2}(i_1,j_1) - D^{c_2}(i_2,j_2)$ are due to a change in reflectance; or
- $m_b(i_1,j_1) < m_b(i_2,j_2)$, $C_b^{c_2}(i_1,j_1) = C_b^{c_2}(i_2,j_2)$ and $C_b^{c_1}(i_1,j_1) = C_b^{c_1}(i_2,j_2)$, which means the changes in intensity $D^{c_1}(i_1,j_1) - D^{c_1}(i_2,j_2)$ and $D^{c_2}(i_1,j_1) - D^{c_2}(i_2,j_2)$ are due to a change in the surface normal vectors between (i_1,j_1) and (i_2,j_2) because of the curvature of the surface.

Consequently, in this case, contrary to the situations depicted through propositions 8.11-8.14, there is no way to decide whether there is a change in reflectance or merely a change in m_b, and, thus, an ambiguity results. In other words, the interrelation between colour channels is not powerful enough in this case so as to disambiguate the source of the intensity variation. Notice that the analysis of images made so far is located at the signal level and, thus, the ambiguity lies at this level. Accordingly, information coming from the application domain level or from other sensor, if available, should be used to make the decision. For instance, if objects with smooth surfaces are expected in the scene, gradient information can result practical to locate the remaining reflectance transitions, which will be called GRD reflectance transitions from now on (see figure 8.7(d)).

In particular, second-order derivative operators are of common use for edge detection in piecewise constant images. On those cases, the idealized step/ramp edges give rise to a zero crossing at the edge because of the sudden change in the first-order derivative, which in turn makes the second-order derivative change its sign in order to return to a situation of nearly zero gradient (see figure 8.8). In general images (i.e. in images not necessarily consisting of a collection of tableaus), a more general analysis must be performed since other edges apart from the step/ramp edge appear.

To illustrate this point, consider the middle plot of figure 8.9, which shows the intensity profiles of row 60 of the uppermost synthetic image. As can be appreciated, the profiles are no longer piecewise constant but rather piecewise curved. Accordingly, for this sort of images, the output of a second-order derivative operator will be negative at places where the intensity is convex, while, at concave places, the output will be positive, as can be observed in the lower plot of figure 8.9. Inflection points take place at locations where the profiles change from being convex-shaped to be concave-shaped or vice versa. Although the inflection can happen in uniform reflectance areas as a consequence of the properties of

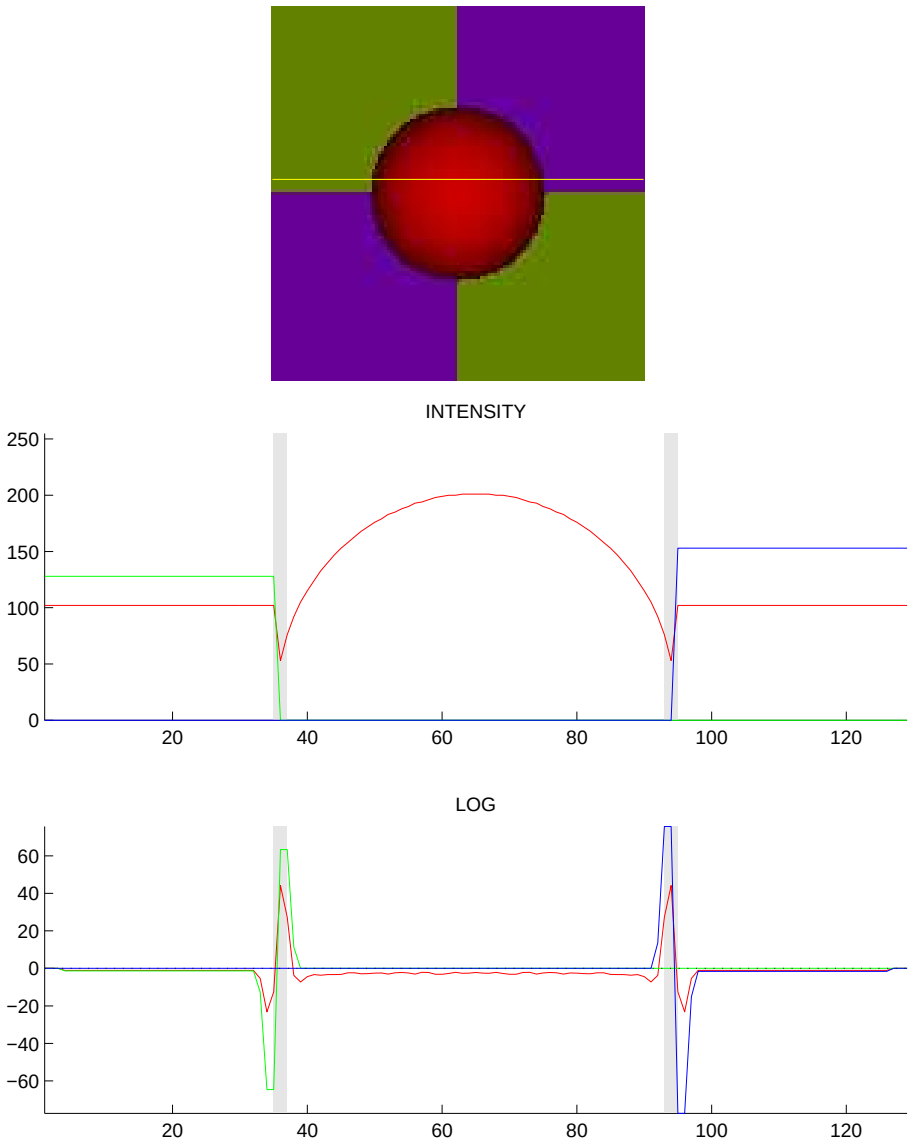


Fig. 8.9. Relationship between LOG zero crossings and reflectance transitions: [upper plot] intensity profiles for RGB colour channels, for row 60 (in yellow) of uppermost image (shadowed areas correspond to reflectance transition locations); [lower plot] result of convolving the upper profiles with LOG ($\sigma = 1$).

the corresponding scene surface, it particularly occurs at reflectance transitions, as it can be observed in figure 8.9. For instance, consider the left darkened area of the lower plot of figure 8.9 and the red profile. From left to right, the profile changes from being constant to be concave-shaped in the darkened area, and next convex-shaped past the darkened area. As a result, two zero crossings result, although only one of them is supported by a physical cause. A double response is thus produced, although of a different magnitude (i.e. the difference between the negative and the positive peaks in the LOG output) in each case. Another illustration between the relationship between reflectance transitions and LOG zero crossings can be found in the lower plot of figure 8.10.

8.7 Colour Channel Coupling-based Edge Detection (C^3E)

As has been shown in the preceding sections, the analysis of the coupling between colour channels allows locating reflectance transitions in the image. Therefore, an edge map can be built as this analysis is performed over the image, where every edge corresponds to a pixel involved in a reflectance change. As has been shown before, a small amount of reflectance transitions cannot be covered by means of colour channel coupling analysis, so that, in order to obtain a complete edge map, it is necessary to complement the previous analysis with additional means. In particular, a strategy based on locating LOG zero-crossings has been suggested, given the properties of second-order derivative operators when they are applied in general images. The rest of this section discusses the details of C^3E, the resulting final edge detection method.

8.7.1 Edge map computation

The edge map is built by checking, at every pixel, all three properties, as well as whether the pixel is involved in a LOG zero-crossing. To this end, a pixel (i_1, j_1) is considered an edge if:

1. for any of its 8 neighbours (i_2, j_2) and at least one combination of colour channels, both pixels are not compatible with one another in the sense of the relation $\mathcal{C}$ formulated in section 8.5; or
2. for any of its 8 neighbours (i_2, j_2) and at least one colour channel, the sign of LOG changes between (i_1, j_1) and (i_2, j_2), and (i_1, j_1) shows the LOG value closest to 0.

As a result of this process, four types of edges can be found in the edge map, each one related with one type of reflectance transition: *CHC* edge, *NCD* edge, *NDD* edge and *GRD* edge. They are all summarized in figure 8.7.

As mentioned before, when a reflectance change takes place, let us say along direction ξ, colour channels evolve in that direction so as to adopt the configuration corresponding to the new body reflectance. In a synthetic image, this evolution takes place between the two pixels which are adjacent over direction ξ, coming each one from a different scene material, so that the reflectance transition does not involve, in fact, any pixel. In a real image, however, a transition can comprise several pixels due to the round-off caused by aliasing in real cameras. By way of example, the upper plot of figure 8.10 illustrates this point. The plot corresponds to the RGB intensity profiles of row 50 of the real image shown at the top of the figure. The darkened areas approximately enclose the pixels involved in the different reflectance transitions taking place (in this case, all happen to make manifest in the form of crossings between colour channels). Observe that, during the transitions, it can be said, in a certain sense, that the reflectance of those pixels turns out to be undefined. Therefore, the most appropriate thing to do on these situations is to consider all of them as edges, so that thick edges must be expected in the edge map.

To finish, notice that, although not always will specularities cause CHC/NCD/NDD edges, as it has been discussed before, they will do induce edges of type GRD, except for "smooth" specularities.

8.7.2 Implementation issues

In order to counteract image noise in an adaptive way when computing the edge map, the intensity uncertainties computed in chapter 6 are incorporated into the edge detection strategy. Therefore, for every colour channel k, every digital noisy level D^k produced by the camera is associated to an interval $[D^k - \delta(D^k), D^k + \delta(D^k)]$. With the introduction of these uncertainties, either $\mathcal{C}$ and the strategy for locating LOG zero-crossings are redefined:

- With regard to $\mathcal{C}$, given (D_p^1, D_p^2), a rectangular uncertainty area around (D_p^1, D_p^2) covering t_1 uncertainties is defined, as it is depicted in figure 8.11. $\mathcal{C}$ is reformulated then as: *(D_q^1, D_q^2) is $\mathcal{C}$-compatible with (D_p^1, D_p^2) if any of the points belonging to the uncertainty area of (D_q^1, D_q^2) falls within the union of the compatibility zones of the points belonging to the uncertainty area of (D_p^1, D_p^2) (shaded area of figure 8.11).* The incompatibility between (D_p^1, D_p^2) and (D_q^1, D_q^2) stems, thus, from the fact that a very low probability exists that any pair of intensity levels which could lead, through noise,

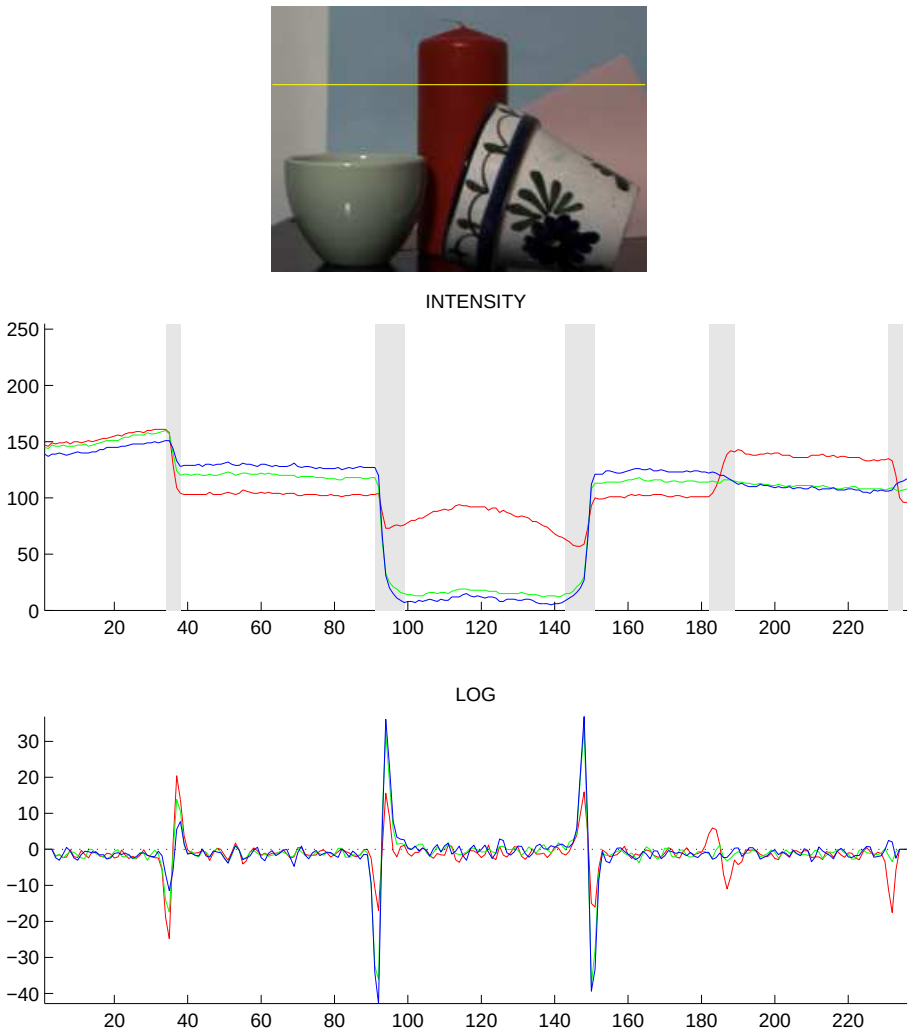


Fig. 8.10. Examples of reflectance transitions in a real image: [upper plot] intensity profiles for RGB colour channels, for row 50 (in yellow) of uppermost image (shadowed areas correspond to reflectance transition locations); [lower plot] result of convolving the upper profiles with LOG ($\sigma = 1$).

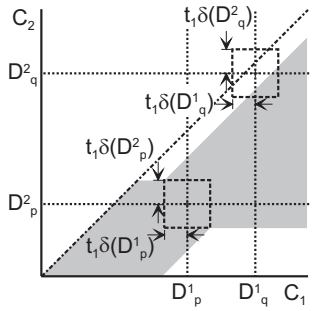


Fig. 8.11. Redefinition of the $\mathcal{C}$ compatibility relation. (The uncertainty areas have been magnified for illustration purposes. In the figure, D_k^n is an abbreviation of $D_k^{c_n}$.)

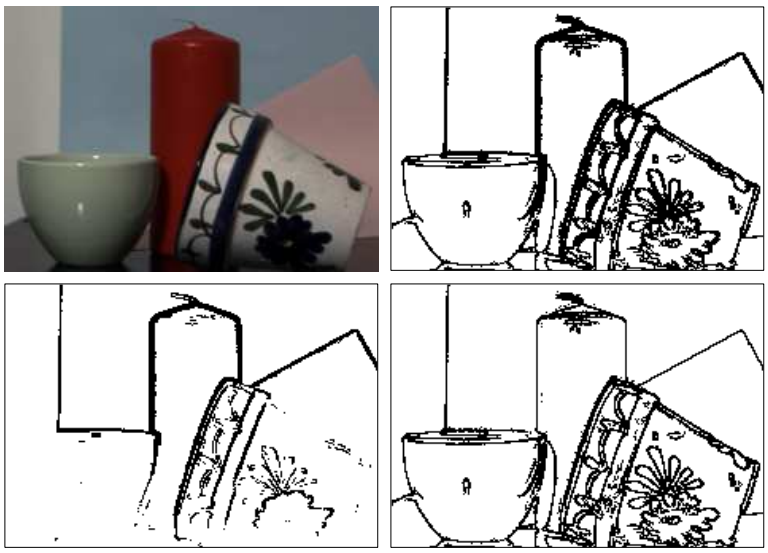


Fig. 8.12. Example of edge map (I): [top,left] original image; [top,right] final edge map; [bottom,left] CHC/NCD/NDD edges ($t_1 = 3$); [bottom,right] GRD edges ($t_2 = 3$).

to (D_p^1, D_p^2) was compatible, in the sense of properties 1-3, with any pair of intensity levels leading, through noise, to (D_q^1, D_q^2).

- As for LOG zero-crossings, the UTLOG strategy described in section 6.4.4 is adopted here. That is to say, whenever a zero-crossing is detected, it is deemed relevant if the positive and negative peak LOG values along the direction of detection are larger than t_2 times the respective uncertainties (see figure 6.9(a)). Remember that those uncertainties are computed using standard uncertainty propagation rules, by which, if the output of the LOG operator is calculated as $f = \sum_x D^c m(x)$, where $m(x)$ are the LOG mask constants, then [272]:

$$\delta(f) = \sqrt{\sum_x \delta(D^c(x))^2 m^2(x)} \tag{8.6}$$

8.7.3 Examples of edge maps and discussion

Figure 8.12 shows the partial and final edge maps for the image shown in the upper left corner. As can be observed, GRD edges tend to be found near specularities and also associated with the edges of scene objects or with shadows (see the candle and the cup, for instance), while CHC/NCD/NDD edges can have a considerable thickness (e.g. the candle). On the other hand, some spurious edges of type CHC/NCD/NDD appear when the corresponding reflectance transitions are similar to the ones shown in figures 8.3 and 8.5 (e.g. the borders between dark green and white zones inside the flowerpot). Otherwise, complete contours result between the regions involved (see the contours between the blue paper sheet and its neighbouring regions). Observe that both t_1 and t_2 have been set to 3. Finally, the combined edge map shown in figure 8.12(top,right) already reveals a certain degree of overdetection. For instance, observe the rim of the cup, a uniform reflectance area which results split in several pieces due to some false positive edges which sometimes proliferate until constituting a closed contour.

From the previous example and the plots shown in figures 8.9 and 8.10, it can be easily deduced that there can be an important degree of overlapping between GRD edges and CHC/NCD/NDD edges, what raises the question whether a classical feature such as LOG zero-crossings would be enough to locate reflectance transitions, and, consequently, build the edge map without detecting the colour edges. Certainly, the main reason for this coincidence is that reflectance transitions tend to produce abrupt changes in intensity and therefore LOG zero crossings. However, it is important not to forget that

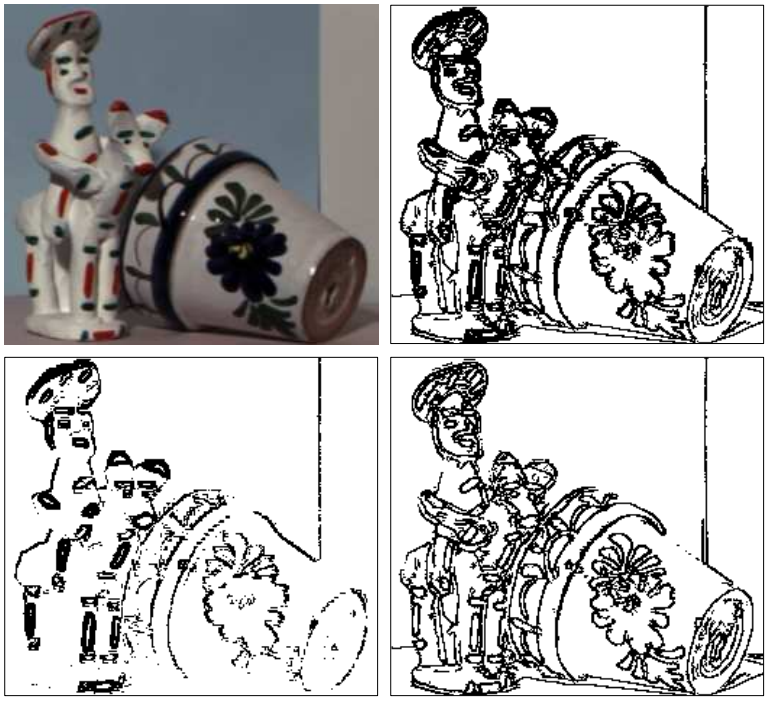


Fig. 8.13. Example of edge map (II): [top,left] original image; [top,right] final edge map; [bottom,left] CHC/NCD/NDD edges ($t_1 = 3$); [bottom,right] GRD edges ($t_2 = 3$).

CHC/NCD/NDD edges correspond to situations in the signal by which a reflectance transition (i.e. the physical cause) manifest, while LOG responds to variations in intensity related with changes in the concavity/convexity of the image surface. Consequently, it can so happen that, existing a clear evidence of a CHC/NCD/NDD edge, the sensitivity of the LOG operator, controlled by the σ parameter, is not sufficient to catch the corresponding change in intensity. Figure 8.13 illustrates this point. As can be observed, neither the map of colour edges alone or the map of GRD edges alone contain all the edges. The combination of both, however, yields a map where very few edges are missed.

8.8 Edge Thinning Strategy

As has been discussed and shown in previous sections, the edge map produced by the algorithm of section 8.7 tends to contain thick edges mainly because reflectance transitions in real images span several pixels along the edge direction. This section proposes an edge thinning strategy just in case this edge detection algorithm is used within a vision system such that the next procedure along the vision pipeline requires thin edges.

Non-maximum suppression is a simple and widely used method for edge thinning [25]. It deletes all edge responses that are not maximal in each cross section of the edge direction in their local neighbourhood. However, this procedure can remove real edges and, sometimes, alter the topological characteristics of the original edge map. A method directed by gradient magnitude which preserves edge connectivity is proposed next for the edge detector presented in section 8.7. Unfortunately, experimental results have shown unit thickness is not always achieved. In case this property is essential, it is suggested to compute the skeleton of the pre-thinned edge map. Good results have been obtained by means of the morphological method described in [136].

First of all, connected components of the edge map not including edges are found. Next, edges adjacent to these connected components are inserted in a heap so that, when pulled out, edges are visited in increasing order of gradient. Whenever an edge is visited, all its non-visited neighbouring

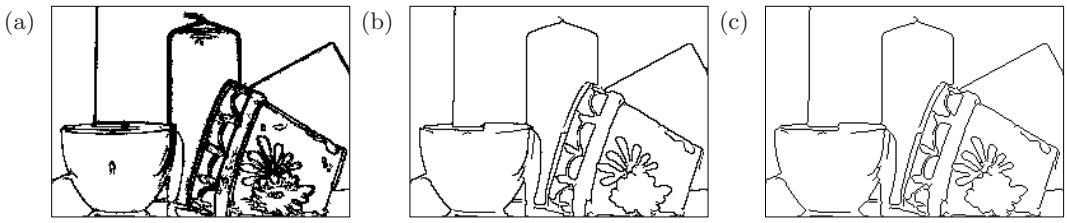


Fig. 8.14. Example of thinning: (a) edge map ($t = 3$ for both colour and GRD edges); (b) thinned edge map; (c) final edge map after the morphological thinning of (b).

edges are in turn introduced in the heap according to their own gradient. Besides, if an edge happens to be in the middle of two already visited edges along the gradient direction, the edge is labeled as a final edge; otherwise it is added to the connected component to which it was adjacent. Therefore, it is the gradient what is used to choose the representative of a set of edges found along a given direction because of a reflectance transition arising there.

In a second stage, once all the edge pixels have been assigned to a connected component and some of them have already been labeled as final edges, those in the borders between different components are also considered to be final edges. For every pair of adjacent candidates, if none of them were selected as final edge in the first stage, the one having largest gradient is selected now.

During this process, if a connected component happen to comprise a reduced amount of pixels, the connected component is removed and its pixels are added to the heap and then pulled out in increasing order of gradient. In this way, some spurious edges which appear due to, above all, the thickness of CHC/NCD/NDD edges are avoided.

Figure 8.14 shows the output of the thinning stage for the image appearing in figure 8.12.

8.9 Experimental Results

To prove experimentally the usefulness of C^3E, several results for synthetic and real images are provided. Furthermore, C^3E is extensively compared with the physics-based edge detection algorithm by Stokman and Gevers [79, 262] (S&G from now on). Results for the non-physics-based algorithm by Meer and Georgescu published in [168][3] (M&G from now on) are also shown.

As for parameters:

- In the case of C^3E, the only relevant parameters are the number of uncertainties to be used when detecting CHC/NCD/NDD edges, t_1, and the number of uncertainties associated to GRD edges, t_2. However, since the same value can be given to both, an only parameter results, which will be referenced as t. The value given to t in every experiment is indicated in each case.
  On the other hand, the value of the standard deviation for the LOG operator, σ_{LOG} used for detecting GRD edges was all the time set to 1.0. The size of the LOG mask was then determined as $2 \times \lceil 3\sigma_{LOG} \rceil + 1$.
  Additionally, in accordance with the post-processing stage of the segmentation/edge detection framework proposed in section 4.4 (figures 4.2 and 4.7):
  - connected sets of edges with a size below 10 pixels are removed, and
  - the connected components of non-edge pixels employed during the thinning step (section 8.8) are required to have a size above 50 pixels; otherwise, the corresponding pixels are inserted into the heap to join another connected component.
- Referring to S&G, it is a parameter-free algorithm which also uses intensity uncertainties to find local thresholds, so that no parameters needs to be set up. Besides, non-maxima suppression is used to thin the corresponding edge maps (as proposed by the authors) in order to have all the edge maps thinned throughout the algorithms comparison.

[3] The code is publicly available at http://www.caip.rutgers.edu/riul/research/code.html

- Finally, regarding M&G, the parameters set by default by the application they offer in their web page are used.

8.9.1 Experiments with synthetic images

In the experiments with synthetic images, C^3E was faced against two sets of 100 synthetic noisy images. Unless explicitly stated below when describing the particular experiment, both sets corresponded to normal noise conditions; that is to say, the noise was added, as on other occasions, according to the camera noise parameters estimated in section 6.4.1 following pseudocode 6.3. On the other hand, while one set consisted of scenes involving spheres and planes of different reflectances, the scenes of the other set only contained planes, in order to see if there were differences in performance depending on whether the scenes included or not curved objects. As for parameters, in all cases, the same number of uncertainties (t) was used for both detecting CHC/NCD/NDD edges and for GRD edges.

As a first experiment, and given the comments on edge detection performance evaluation expressed in section 4.4.2, Baddeley's performance measure (see table 4.8) was evaluated for C^3E, varying the number of uncertainties t used to compute the edge map. Figure 8.15 shows the results. As can be observed and could be easily guessed, the performance of C^3E varies with the number of uncertainties used; moreover, there is a value of t for which maximum performance is attained for both sets of images. More precisely, such performance is obtained for $t = 5$, although very similar values were obtained for values of t around 5. Clearly, this behaviour agrees with the statistical meaning of t: for low values of t a lot of false positives are generated, while, for large values, an increasing number of true edges are not detected (false negatives). As a result of this experiment and given the statistical meaning of t (see section 6.2.2), it seems that $t = 3..6$ is an adequate selection of values which should work on most occasions. In particular, $t = 6$ allows detecting edges corresponding to intensity variations whose probability of being due to noise is below $1/36 = 2.78$ %, while for $t = 3$ this value rises up to $1/9 = 11.11$ %.

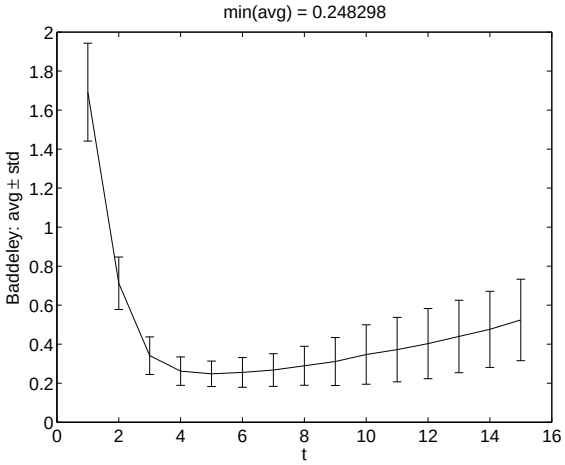
In the second experiment, C^3E and S&G were compared under normal noise conditions. According to the previous considerations about the values to be given to t, three values for t were considered: 3, 4 and 5. Furthermore, the four evaluation measures appearing in table 4.8 were computed in order not to make the comparison dependent on a particular evaluation measure. The numerical results appear in the tables of figure 8.16. As can be observed, Baddeley's measure indicates C^3E edge maps are better than the ones by S&G, while FOM and D measures are quite similar for both algorithms, irrespective of the value given to t. The source of this discrepancy is likely to lay in what was pointed out in section 4.4.2 regarding the lack of sensitiveness to false edge positives in FOM's case; as for the discrepancy percentage D, it should be taken into account that the location of misclassified pixels is not considered by this measure. On the other hand, the interval of Baddeley's measures obtained for S&G contains that of C^3E, which means that for some images S&G produced better edge maps than C^3E. However, at the same time, the larger interval length indicates a bigger instability in the results of S&G, which is corroborated by a larger standard deviation. All the aforementioned can also be observed graphically in the plots of figure 8.16 for $t = 4$.

As a final experiment, increasingly amplitude noise was added to both sets of synthetic images in order to observe the capability of adaptation to different levels of noise of C^3E and S&G. To this end, new noise models were generated multiplying by a scalar k the standard deviations corresponding to the noise sources (i.e. σ_K, σ_{dc} and σ_f^c, in equation 6.11b of section 6.2.2), as was already done in the experimental results section of chapter 7. Accordingly, the intensity uncertainties used by both algorithms were recalculated for each level. $k = 1..5$ were the levels of noise considered, where $k = 1$ means normal noise conditions. In order to illustrate the amount of noise added, figure 8.17 reproduces again some synthetic images for all levels.

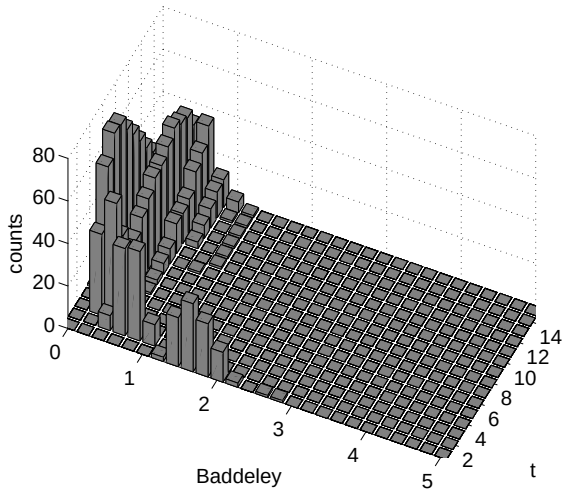
Figure 8.18 shows Baddeley's results for this last experiment and $t = 3..5$. Once more, C^3E gets better scorings than S&G and, besides, C^3E shows a stronger tolerance to noise, performing similarly as under normal noise conditions up to $k = 3$. From there on (i.e. $k = 4$ and 5), a slight trend towards worse edge maps can be observed in C^3E, which is quite reasonable given the amount of noise introduced. Finally, it is specially notorious the fact that, as noise increases, C^3E behaves better for $t = 3$ against $t = 4..5$.

(a)

$t \rightarrow$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
avg	1.69	0.71	0.34	0.26	0.25	0.26	0.27	0.29	0.31	0.35	0.37	0.40	0.44	0.48	0.52
std	0.25	0.13	0.10	0.07	0.07	0.08	0.08	0.10	0.12	0.15	0.17	0.18	0.19	0.20	0.21
min	1.25	0.37	0.14	0.11	0.09	0.08	0.09	0.09	0.09	0.09	0.09	0.09	0.10	0.13	0.15
max	2.72	1.00	0.63	0.41	0.42	0.53	0.55	0.61	0.94	1.18	1.19	1.19	1.19	1.20	1.21



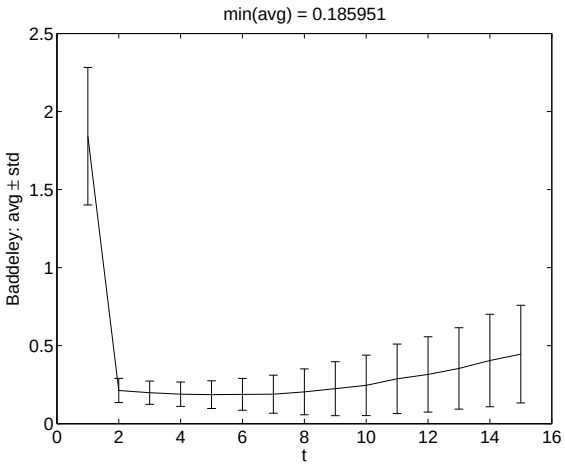
(b)



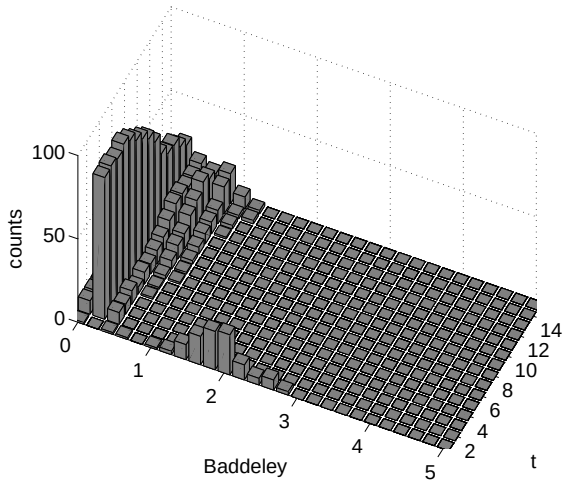
(c)

(d)

$t \rightarrow$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
avg	1.84	0.21	0.20	0.19	0.19	0.19	0.19	0.20	0.22	0.25	0.29	0.32	0.35	0.40	0.45
std	0.44	0.08	0.07	0.08	0.09	0.10	0.12	0.15	0.17	0.19	0.22	0.24	0.26	0.30	0.31
min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
max	2.78	0.36	0.34	0.48	0.56	0.56	0.76	0.76	0.77	0.77	0.84	0.97	1.05	1.16	1.16



(e)



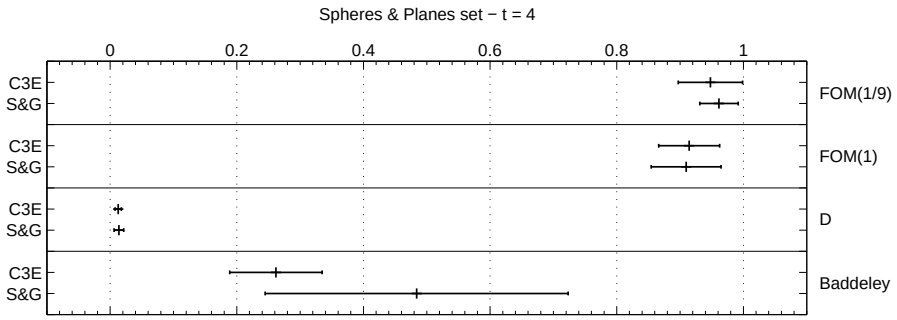
(f)

Fig. 8.15. Experimental results (Baddeley's measure) for C^3E and noisy synthetic images: (a), (b) and (c) are for the spheres and planes set, while (d), (e) and (f) are for the only-planes set; (a,d) average (avg), standard deviation (std), minimum (min) and maximum (max) for $t = 1..15$; (b,e) plot of average values and standard deviations for $t = 1..15$; (c,f) histograms of Baddeley's measures for $t = 1..15$.

(a)

	C^{3}E			S&G
	$t = 3$	$t = 4$	$t = 5$	
FOM(1/9)	0.96 - 0.03 [0.76,0.99]	0.95 - 0.05 [0.73,0.99]	0.93 - 0.06 [0.72,0.99]	0.96 - 0.03 [0.85,1.00]
FOM(1)	0.90 - 0.04 [0.73,0.97]	0.91 - 0.05 [0.71,0.98]	0.90 - 0.06 [0.70,0.98]	0.91 - 0.06 [0.76,0.99]
D×100	1.64 - 0.67 [0.35,3.45]	1.26 - 0.49 [0.28,2.60]	1.21 - 0.45 [0.29,2.31]	1.37 - 0.76 [0.15,3.46]
Baddeley	0.34 - 0.10 [0.14,0.63]	0.26 - 0.07 [0.11,0.41]	0.25 - 0.07 [0.09,0.42]	0.48 - 0.24 [0.07,1.23]

(b)



(c)

	C^{3}E			S&G
	$t = 3$	$t = 4$	$t = 5$	
FOM(1/9)	0.98 - 0.01 [0.94,1.00]	0.98 - 0.02 [0.90,1.00]	0.98 - 0.02 [0.89,1.00]	0.96 - 0.05 [0.75,1.00]
FOM(1)	0.96 - 0.02 [0.88,1.00]	0.96 - 0.02 [0.89,1.00]	0.97 - 0.02 [0.87,1.00]	0.96 - 0.05 [0.75,1.00]
D×100	0.62 - 0.40 [0.00,2.26]	0.57 - 0.39 [0.00,2.09]	0.55 - 0.37 [0.00,1.93]	0.32 - 0.42 [0.00,2.25]
Baddeley	0.20 - 0.07 [0.00,0.34]	0.19 - 0.08 [0.00,0.48]	0.19 - 0.09 [0.00,0.56]	0.29 - 0.24 [0.00,1.18]

(d)

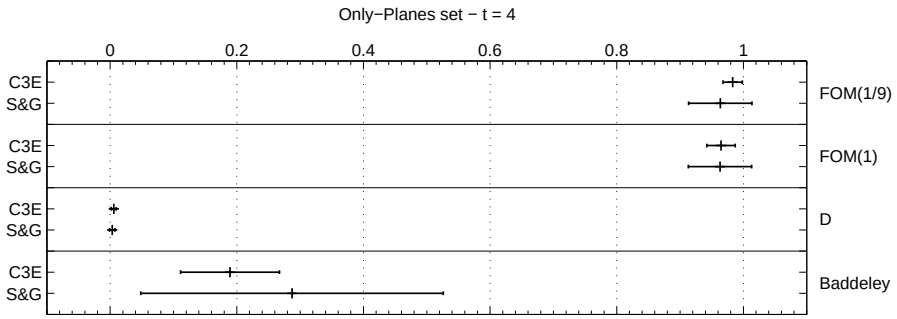
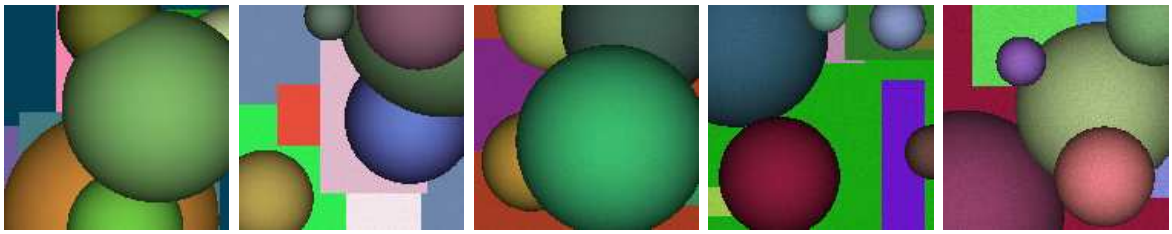


Fig. 8.16. Comparison between C^{3}E and S&G under normal noise conditions by means of synthetic images: (a) and (b) are for the spheres and planes set, while (c) and (d) are for the only-planes set; (a,c) every entry of the tables consists of average - standard deviation [minimum,maximum]; (b,d) plots of average and standard deviation values for $t = 4$.

8.9.2 Experiments with real images

This section presents experimental results for C^{3}E over a varied set of real images including images typically used for testing edge detection algorithms, i.e. standard images, and also "proprietary" images taken under the conditions outlined in section 4.4.1. As in the previous section, comparison results are also provided. On this occasion, performance measurements for both S&G and M&G are given.

On the one hand, quantitative results have been obtained on the basis of the six images appearing in figure 8.19, which are representative of the following types of scenes: two correspond to scenes with curved objects, (a) and (b); two contain mainly natural objects, (c) and (d); and, finally, two are



(a) (b) (c) (d) (e)

Fig. 8.17. Examples of noisy synthetic images used for comparing C^3E and S&G: (a) noise×1; (b) noise×2; (c) noise×3; (d) noise×4; (e) noise×5.

standard images, (e) and (f). Performance was evaluated using the manually-generated reference edge maps also shown in figure 8.19, which were drawn according to the following criteria:

- Edges cover whole reflectance transitions, so that the reference edge map can, and in fact do, contain thick edges.
- In accordance to the discussion about the object attribute that should steer edge detection and image segmentation tasks (section 4.3), edges were made to correspond to exclusively material transitions. This is specially relevant for images containing natural objects, such as the images of figure 8.19(c) and (d). As can be observed in the respective reference edge maps, the different leafs were not put apart since all of them correspond to, approximately, the same material; only when there is a noticeable pigmentation change an edge was drawn. An analogous reasoning can be applied to the *peppers* standard image of figure 8.19(f).
- Transitions between non-shadow and shadow areas were labeled as edges since none of the algorithms involved in the experimental results is able to tolerate them in practice.

It is important to notice that some of the images of figure 8.19 can more or less slightly deviate from the image formation model assumed by the physics-based algorithms involved in this comparison, specially regarding the standard images. This is because not all the surface materials appearing on those regions have been proved to satisfy the Dichromatic Reflection Model (it was initially conceived for inhomogeneous dielectrics, see section 2.3.4). However, this sort of deviations belong to the term *scene noise* introduced in section 4.4.3, and would, therefore, be useful to observe the degree of tolerance of C^3E and S&G to this sort of noise.

Figure 8.20 presents the results for the first experiment with real images. It shows Baddeley's C^3E values for the image appearing in figure 8.19(a), varying t between 1 and 15. As can be noticed, optimum results are obtained for $t = 3$ and $t = 4$, although the interval involving from $t = 2$ until $t = 5$ produces very similar values. The lower rows of figure 8.20 presents the edge maps (superimposed on the original image and alone) for $t = 2..6$, the values giving a performance measure below 1. They show how, as t increases, the number of edges decreases. Nevertheless, other edges not appearing for lower values of t appear with larger values (i.e. observe the top of the candle). This is the combined effect of the thinning and the post-processing steps. Finally, observe that for reasonable values of t the edges found tend to constitute closed contours, which will be essential for the image segmentation algorithm which will be proposed in chapter 9.

On the other hand, figure 8.21 and table 8.3 shows comparison results for all the images appearing in figure 8.19. As can be noticed, C^3E outperforms S&G numerically in all cases, sometimes with a larger difference (figure 8.19(a)) and other times very slightly (figure 8.19(c)). As for M&G, C^3E tends also to produce better edge maps, although, on some occasions, the numerical values are almost the same. The larger differences in performance between physics and non-physics-based algorithms appear, as could be easily guessed, for images containing curved objects (figures 8.19(a) and (b)). Finally, notice that, for some images, M&G produces better edge maps than S&G. However, this happens more notoriously with the standard images, for which, surely, gamma correction was not turned off, what can affect S&G; it seems, thus, that the non-linearity of camera output affects more to S&G than to C^3E.

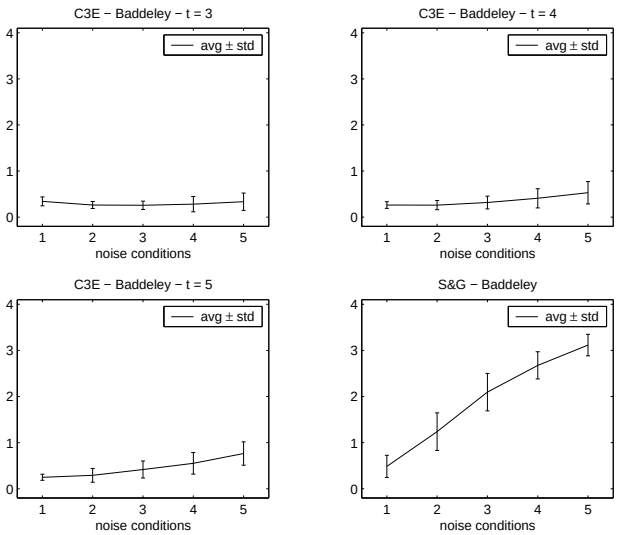
(a) Spheres & Planes set (Baddeley's measure)

	noise×1		noise×2		noise×3		noise×4		noise×5	
	C^3E	S&G	C^3E	S&G	C^3E	S&G	C^3E	S&G	C^3E	S&G
$t = 3$	0.34 - 0.10 [0.14,0.63]		0.26 - 0.08 [0.12,0.65]		0.26 - 0.09 [0.11,0.66]		0.28 - 0.17 [0.10,0.88]		0.33 - 0.19 [0.12,0.88]	
$t = 4$	0.26 - 0.07 [0.11,0.41]	0.48 - 0.24 [0.07,1.23]	0.26 - 0.10 [0.09,0.69]	1.24 - 0.41 [0.40,2.12]	0.32 - 0.14 [0.08,0.69]	2.09 - 0.41 [1.09,3.00]	0.41 - 0.21 [0.11,1.04]	2.68 - 0.30 [1.94,3.37]	0.53 - 0.24 [0.15,1.21]	3.11 - 0.23 [2.34,3.58]
$t = 5$	0.25 - 0.07 [0.09,0.42]		0.29 - 0.15 [0.06,1.04]		0.42 - 0.18 [0.12,0.94]		0.55 - 0.23 [0.12,1.17]		0.76 - 0.25 [0.21,1.40]	

(b) Only-Planes set (Baddeley's measure)

	noise×1		noise×2		noise×3		noise×4		noise×5	
	C^3E	S&G	C^3E	S&G	C^3E	S&G	C^3E	S&G	C^3E	S&G
$t = 3$	0.20 - 0.07 [0.00,0.34]		0.19 - 0.11 [0.00,0.76]		0.19 - 0.14 [0.00,0.91]		0.22 - 0.19 [0.00,0.86]		0.31 - 0.25 [0.00,0.93]	
$t = 4$	0.19 - 0.08 [0.00,0.48]	0.29 - 0.24 [0.00,1.18]	0.20 - 0.15 [0.00,0.91]	1.00 - 0.48 [0.00,2.58]	0.22 - 0.20 [0.00,1.07]	1.79 - 0.60 [0.00,3.47]	0.32 - 0.26 [0.00,1.17]	2.48 - 0.42 [1.37,3.62]	0.47 - 0.30 [0.00,1.08]	2.87 - 0.34 [2.14,4.18]
$t = 5$	0.19 - 0.09 [0.00,0.56]		0.21 - 0.17 [0.00,0.95]		0.32 - 0.25 [0.00,1.06]		0.48 - 0.30 [0.00,1.34]		0.65 - 0.35 [0.00,1.52]	

(c) Spheres & Planes set



(d) Only-Planes set

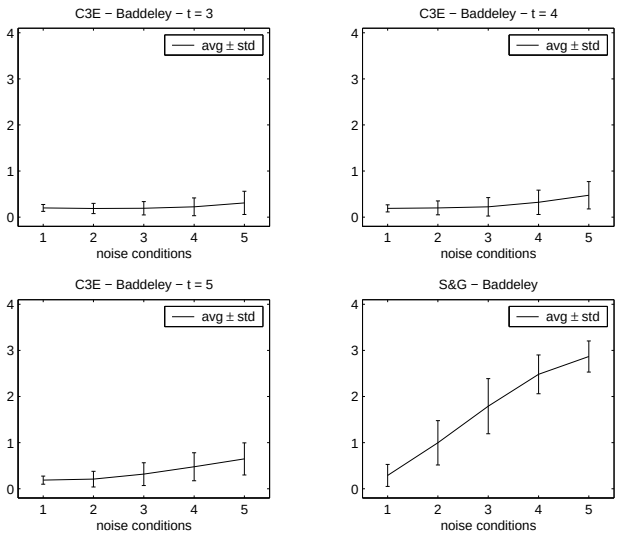


Fig. 8.18. Comparison between C^3E and S&G under increasing noise conditions by means of synthetic images and Baddeley's measure: (a,c) spheres and planes set; (b,d) only-planes set; (a,b) every entry of the table consists of average - standard deviation [minimum,maximum] for Baddeley's measure; (c,d) plots of average and standard deviation of Baddeley's measure for $t = 3..5$.

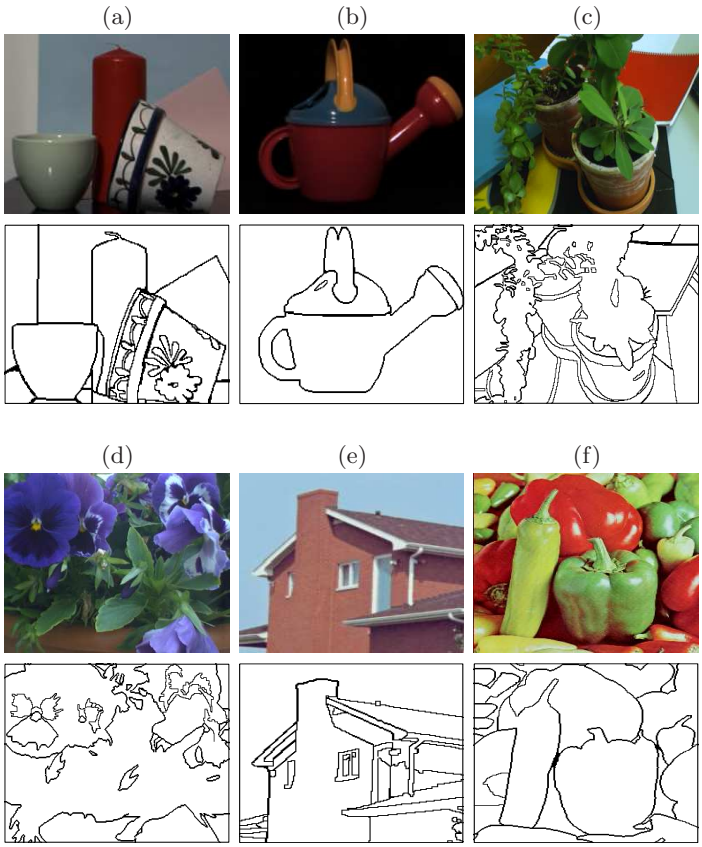


Fig. 8.19. Real test images and their manually-generated ground truths.

$t \rightarrow$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Baddeley	1.36	0.78	0.75	0.75	0.78	0.92	1.12	1.27	1.32	1.40	1.42	1.47	1.53	1.57	1.61

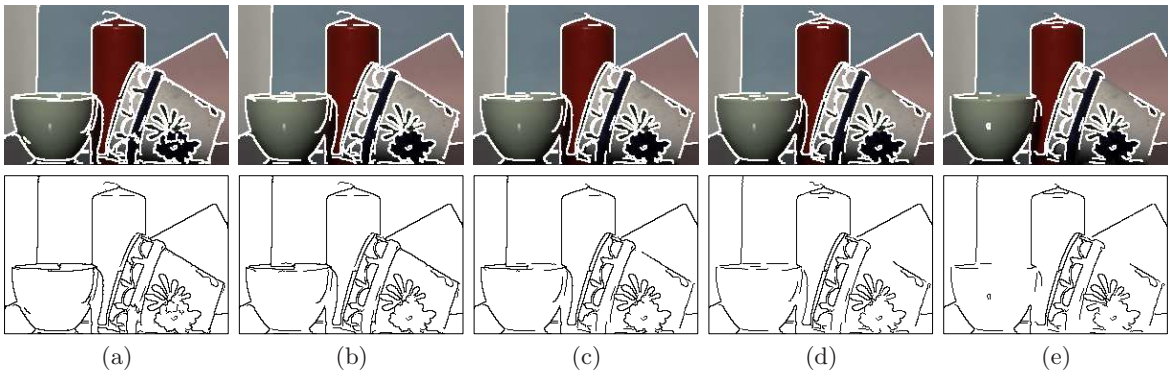


Fig. 8.20. Experimental results for C^3E and real image 8.19(a): [top] Baddeley's measure for $t = 1..15$; [bottom] resulting edge maps for (a) $t = 2$, (b) $t = 3$, (c) $t = 4$, (d) $t = 5$ and (e) $t = 6$.

		C^3E			S&G	M&G
		$t = 3$	$t = 4$	$t = 5$		
8.19(a)	Pratt(1/9)	0.68	0.65	0.62	0.56	0.58
	Pratt(1)	0.62	0.59	0.57	0.51	0.51
	D×100	6.21	6.15	6.20	6.78	7.79
	Baddeley	0.75	0.75	0.78	0.82	1.13
8.19(b)	Pratt(1/9)	0.73	0.71	0.69	0.59	0.64
	Pratt(1)	0.55	0.54	0.53	0.42	0.46
	D×100	5.17	5.03	4.87	5.47	5.83
	Baddeley	0.97	0.90	0.85	1.08	1.22
8.19(c)	Pratt(1/9)	0.69	0.70	0.72	0.67	0.64
	Pratt(1)	0.52	0.53	0.56	0.47	0.48
	D×100	12.67	12.23	11.50	14.04	12.28
	Baddeley	1.69	1.64	1.56	1.73	1.78
8.19(d)	Pratt(1/9)	0.52	0.56	0.60	0.44	0.47
	Pratt(1)	0.38	0.42	0.46	0.28	0.33
	D×100	13.88	12.20	10.99	18.57	13.82
	Baddeley	2.22	2.03	1.89	2.73	2.32
8.19(e)	Pratt(1/9)	0.88	0.88	0.86	0.80	0.70
	Pratt(1)	0.70	0.73	0.71	0.58	0.58
	D×100	6.59	6.05	6.05	8.80	6.02
	Baddeley	0.80	0.66	0.71	1.15	0.78
8.19(f)	Pratt(1/9)	0.47	0.53	0.56	0.46	0.54
	Pratt(1)	0.35	0.41	0.44	0.30	0.43
	D×100	12.67	10.56	9.42	13.71	9.04
	Baddeley	2.30	2.01	1.86	2.48	1.87

Table 8.3. Comparison between C^3E, S&G and M&G by means of real images.

At the qualitative level, it deserves special attention the behaviour of M&G with regard to, in particular, C^3E. Observing the results of figure 8.21(a) and (b), i.e. the two images containing curved objects, it can be noticed that M&G produces edges related with scene curvature while C^3E does not because of its basis on the physics of image formation. For instance, see the green cup of the image in figure 8.21(a). However, in the same image, it can be noticed that C^3E does not deal with small details as well as M&G, as can be observed in the ornamentation of the upper part of the flowerpot. Clearly, the contours are better delineated by M&G. This behaviour has to do with the fact that C^3E labels as edges all the pixels involved in a reflectance transition and, thus, thick edges result, what makes difficult the treatment of details of reduced size. The obvious solution for these cases is, of course, to increase image resolution. On the other hand, referring to S&G, this algorithm tends to produce an important quantity of spurious edges, probably due to the instabilities of photometric invariants it is based on, which is more notorious in the standard images and the images containing natural objects.

To finish with this section, figure 8.22 shows edge maps for another set of real images, for which no ground truth has been generated, so that only a qualitative analysis can be performed. As before, some of the images strictly conform to the assumed model of image formation, while others do not. As can be observed, M&G tends to miss some useful edges which are indeed identified by C^3E. See, among other cases, the big red and the small green balls in the second image, in the upper group of images, or the different treatment given to leaves in the images of the lower group. A similar behaviour exhibits S&G in the third image of the upper row. As before, S&G introduces a lot of spurious edges

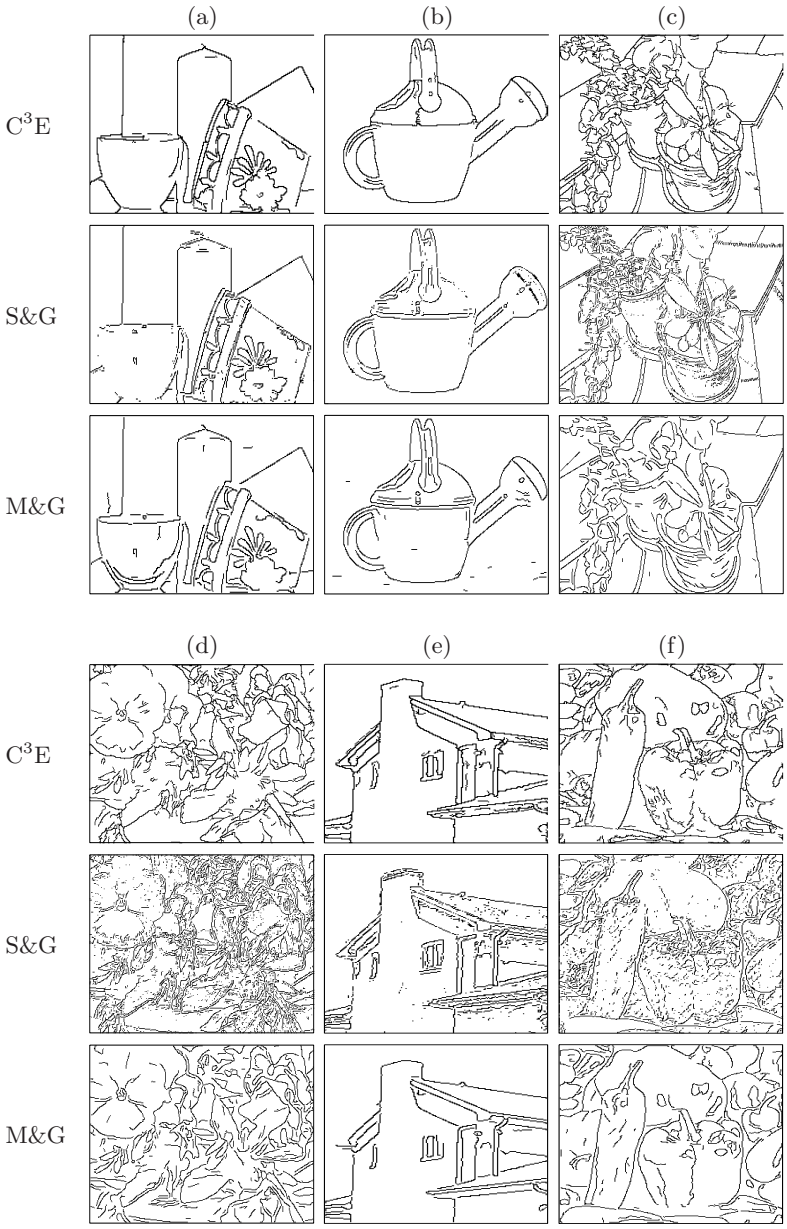


Fig. 8.21. Edge maps resulting for C^3E ($t = 4$), S&G and M&G for images in figure 8.19.

in the edge maps. Nevertheless, S&G seems to produce the best edge map for the *lena* image (third image of the second row) if the spurious edges are removed. Finally, observe that, once again, small details tend to put C^3E in trouble, such as in the case of the chain of the second image of the second row.

8.10 Conclusions

A study about what happens at the signal level when reflectance changes take place has been presented in this chapter. As has been shown, colour channels keep coupled in uniform reflectance areas, while the coupling is broken in a number of ways at pixels involved in reflectance transitions. This coupling

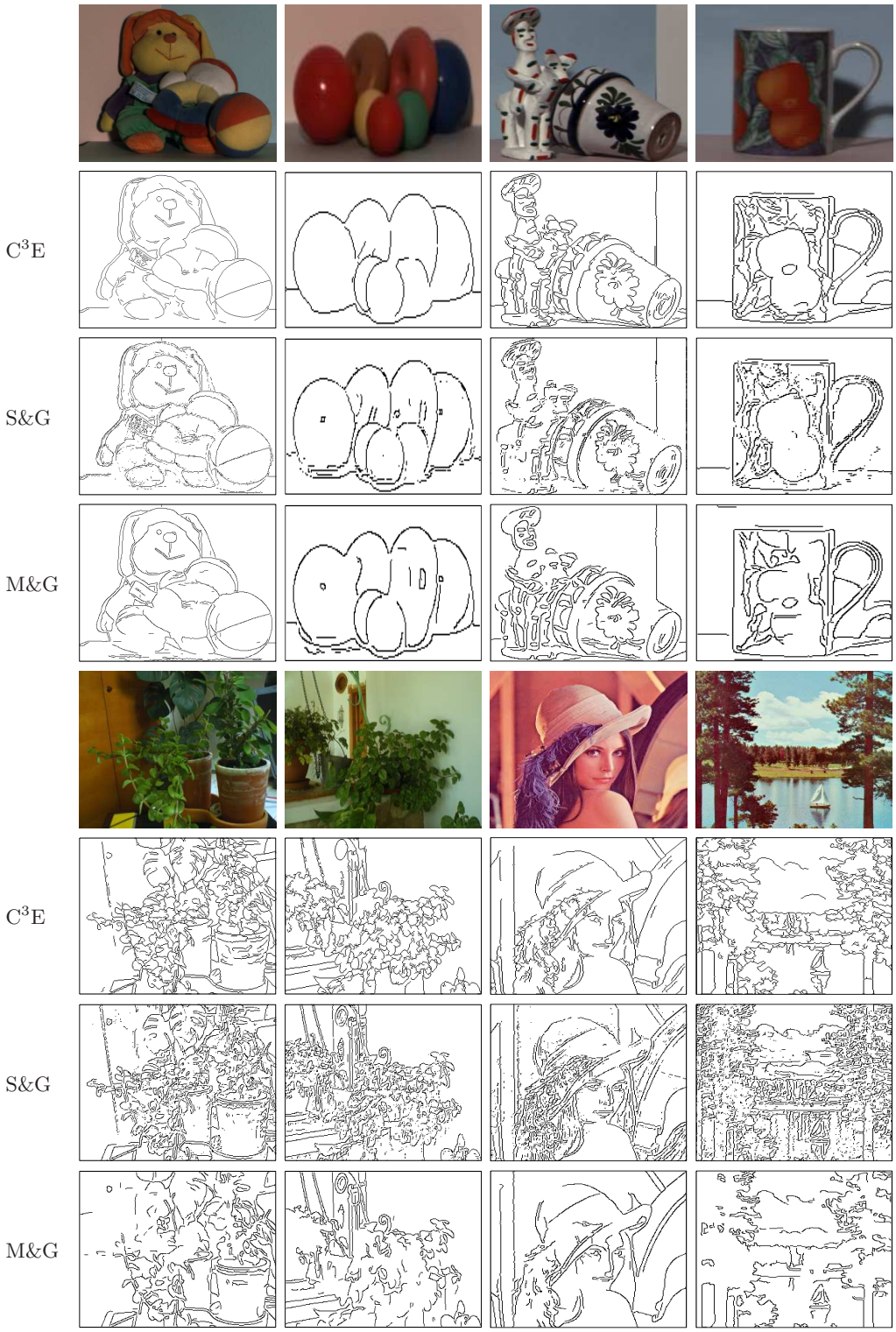


Fig. 8.22. Edge maps resulting for C^3E ($t = 4$), S&G and M&G for more real images.

has been expressed in the form of a set of properties which are fulfilled by all the pixels corresponding to the same surface material (i.e. same reflectance), so that, when any of them is not satisfied, the only possible cause of the unfulfillment reduces to a change in reflectance. The satisfaction of those properties has been studied for different instantiations of the image formation model, which have allowed determining the constraints to be met by the scene so that these properties can be effectively used to locate reflectance transitions. An important aspect of those properties is that they do not depend on the particular expression selected for the geometrical terms m_b and m_i (equation 4.6). The coupling properties have in turn been used to derive a number of types of edges which can be easily detected by analyzing the way how the colour channels are related among them.

An edge detector named C^3E (*Colour Channel Coupling-based Edge detection*) rooted on the facts resulting from the previous study has also been proposed. Experiments with synthetic and real images have been presented, showing the power of the approach for dealing with scenes with curved objects and different surface materials. Besides, C^3E has been compared with the physics-based edge detection algorithm by Stokman and Gevers [79, 262] (S&G) and with the recognized non-physics-based edge detection algorithm by Meer and Georgescu [168] (M&G). The following conclusions can be drawn from the different tests performed:

- The scene constraints concerning the fulfillment of the coupling properties have not been a serious problem for obtaining sound edge maps. Only in some cases, specularities have produced false CHC/NCD/NDD edges.
- C^3E does not show important differences between scenes where objects shading is noticeable with regard to scenes where it is not. On most occasions, four uncertainties have been enough to produce adequate edge maps.
- C^3E has exhibited a good tolerance to noise, although it has been observed that as the level of noise increased a lower number of uncertainties produced better edge maps. Intensity uncertainties have also resulted to be useful in this computer vision task.
- C^3E has outperformed S&G in all the experiments, although the differences were more notorious with synthetic scenes than with real images.
- C^3E has obtained better scorings than M&G particularly with images containing objects showing an important level of shading. In the rest of cases, a similar performance has been observed.

As for the computation time of the edge maps, it fluctuated around 1 second in all images in a Pentium IV @ 3GHz machine, for image sizes ranging from 256×256 to 512×512.

9

Image Segmentation by Analysis of Colour Channels Coupling

Looking for image regions satisfying an homogeneity criterion and looking for the region borders are, at least theoretically, complementary problems aiming at producing a segmentation of an image. This fact, together with the tendency of C^3E to produce closed contours for adequate values of t, has led to the development of a new segmentation strategy called C^3S (*Colour Channels Coupling-based Segmentation*). C^3S comprises C^3E and two more stages: one which makes regions grow according to the information provided by the (non-thinned) edge maps produced by C^3E, and another one which merges regions corresponding to the same scene material.

This chapter presents C^3S in the following way: section 9.1 outlines C^3S, while the details about the two additional stages of C^3S, region growing and region merging, appear in, respectively, sections 9.2 and 9.3; section 9.4 presents the experiments performed and the corresponding results, and, finally, conclusions appear in section 9.5.

9.1 Outline of the Segmentation Algorithm

The experiments performed have shown that C^3E non-thinned edge maps tend to contain, for adequate values of t (typically $t = 3..4$), closed contours enclosing regions of uniform reflectance. Therefore, a first partition of the image can be obtained if connected components not including edge pixels are found and edges are then added to the nearest most similar connected component. As the edge map is presumed to separate pixels corresponding to "steep" specularities from matte pixels, the regions resulting from the previous process are expected to be describable by, according to the Dichromatic Reflection Model, linear or point clusters in colour space.

Nevertheless, some times, contours are not finished due to the effects of noise and a path of non-edge pixels between two regions of different reflectance is created, leading to a region of non-uniform reflectance. In order to avoid this undesirable result, once the growth of a connected component has finished, the associated region is accepted only in case of a linear or point-like shape of the corresponding cluster in colour space; otherwise, the connected component is discarded. In such a case, the associated pixels are grouped again, although, this time, the cluster in colour space is not let to grow out of control as pixels are added: as soon as the cluster dimensionality exceeds that of a line, the growing stops and a new connected component is created and grown. The process continues until no more pixels of the original connected component are unlabeled.

A region merging stage follows next, in order to remove the probably low, but not generally zero, degree of over-segmentation which can result. This over-segmentation can be caused by: (1) the thickness of the edges produced by C^3E together with the effects of general noise over the image, which can lead to having a true region split in the edge map; (2) the separation of specular pixels from matte pixels, most probably due to the presence of GRD edges along the surroundings of a specularity. In accordance to this two sources of over-segmentation, the second stage of C^3S consists of two steps: one for merging matte regions and another one to join specular regions to their corresponding matte regions.

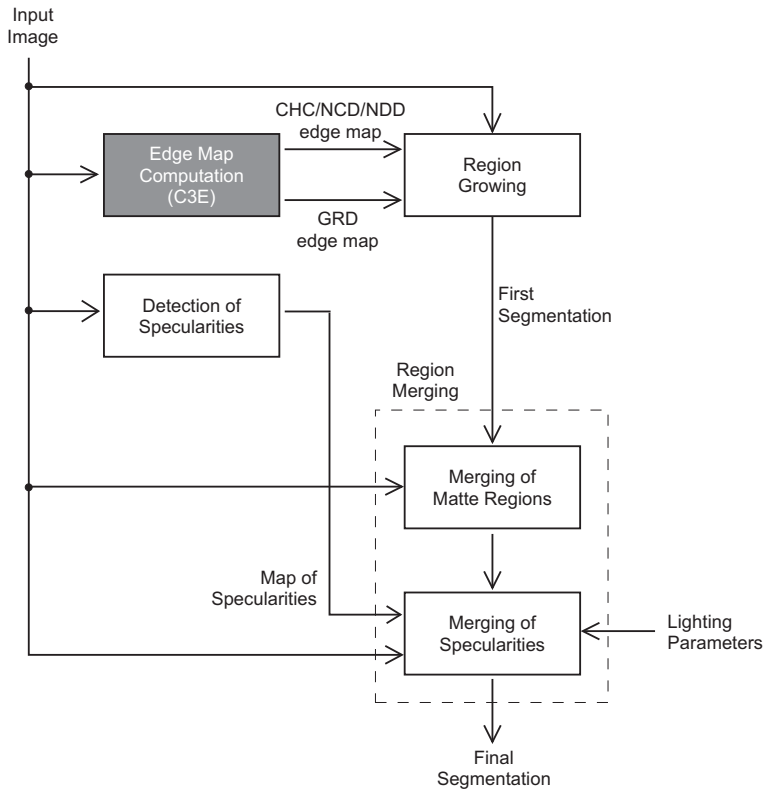


Fig. 9.1. Outline of C^3S.

Figure 9.1 shows graphically the sequence of processes involved in C^3S, together with the flux of information. The details for every one of those processes are given in the following sections.

9.2 Region Growing

As was already discussed in section 2.3.4, the Dichromatic Reflection Model predicts that pixels corresponding to uniformly coloured objects lie in a hyperplane in colour space, spanned by the common body and interface reflectances. Within the hyperplane, matte and glossy pixels fall into two generally different and more or less linear clusters, whose directions coincide with the corresponding body and interface reflectances. Therefore, since the edge map is likely to separate glossy pixels from matte pixels, the aforementioned connected components are expected to be describable by linear clusters in colour space. Consequently, at most one of the eigenvalues resultant from a *Principal Component Analysis* (PCA) over the connected component pixels should be large (i.e. the one related with the direction of the cluster).

Although the experiments performed have shown that the edge maps produced by C^3E tend to contain regions of homogenous reflectance and therefore describable by linear clusters in colour space, sometimes can happen that because of noise a contour is not finished so as to lead to a complete separation. In those cases, clusters of dimensionality above a straight line are obtained because pixels from different scene materials have been put together in the corresponding connected component. In order to distinguish linear clusters from clusters with a higher dimensionality, at least two eigenvalues are required to be below a threshold λ_{low}^2, defined as follows:

$$\lambda_{low} = t_3 \max_c \{\delta(255^c)\}\,, \tag{9.1}$$

Fig. 9.2. Example of output of the region growing stage of C^3S.

where $\delta(255^c)$ represents the uncertainty of the largest intensity level for colour channel c. In this way, the orthogonal distance from any point in the cluster to the straight line chosen as a descriptor can be said to be, roughly speaking, bounded by λ_{low}, which is, in turn, related to the maximum achievable uncertainty.

When a connected component does not pass the test described above is accordingly discarded. The respective pixels are grouped again avoiding pixels whose distance to the present cluster representative is above a threshold τ_{split}, where this representative is updated at every iteration of the procedure. This threshold determines the degree of splitting which is accepted in order to avoid planar and volumetric clusters, and, therefore, determines the effort which will have to be exerted later during the merging stage.

Finally, once all the non-edge pixels have been grouped, edge pixels join the adjacent region closest in colour space in two steps, taking into account the orthogonal distances between edge pixels and the straight lines describing the different connected components found. In the first step, colour edges not coinciding with GRD edges are added to existing regions. Next, connected components consisting of exclusively colour edges are determined and new regions are created. Finally, GRD edges, which have not been included in any region so far, join the nearest region in colour space. This somewhat complex part of the growing stage helps to overcome the thickness of colour edges, taking advantage of the generally narrowness of LOG-based edges. As a result of the growing stage, an initial partition of the image is obtained.

Observe that this process of incorporating edges to image regions can, and in fact does, make use of regions information. The edge thinning procedure of C^3E, which is similar to the growing stage of C^3S, can only make use, however, of local information, in that case in the form of gradient data. Therefore, C^3S can benefit from global information which is not available for C^3E.

By way of example, figure 9.2(right) shows the output of this stage for the image on figure 9.2(left). As can be noticed, over-segmentation appears around the rim of the cup, in the candle and also in several places of the flowerpot.

9.3 Region Merging

9.3.1 Merging of non-glossy regions

Because of false edge positives and due to the potentially large thickness of some edges, it is expected that sometimes regions of uniform reflectance appear split into several parts after the growing stage. In order to distinguish borders which separate regions which could reasonably be considered as perceptually different from those which do not, every region border must be analyzed to determine whether it makes sense from a perceptual point of view or not. This leads, therefore, to a merging strategy based on local analysis, since only pixels from both regions are involved in the merging decision.

Nevertheless, a global strategy could also be of application here. Such a strategy could be based on identifying separate clusters in colour space for the whole image and merging regions whose pixels belong to the same cluster (see [26, 153] for a survey on some segmentation techniques following

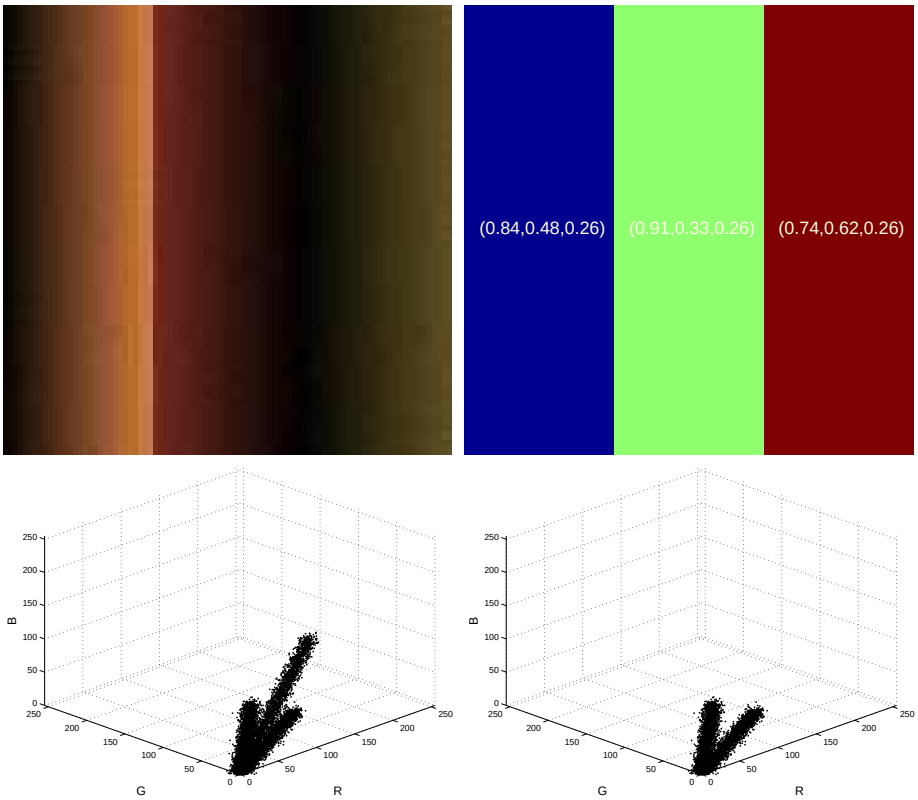


Fig. 9.3. Local analysis *versus* global analysis for region merging: [top,left] a synthetic image of a scene containing three materials; [top,right] segmentation ground truth and the unit vectors of the straight lines where the respective pixels lie on; [bottom,left] pixels in colour space, whole image; [bottom,right] pixels in colour space, only middle and right regions.

this research line). The quality of the final segmentation would strongly depend, however, on the ability to produce a cluster for every different scene material. Figure 9.3 illustrates this point with a synthetic image generated from three materials which should, therefore, lead to a segmentation in three regions. The ground truth segmentation together with the unit vectors for the straight lines where the pixels should lie on in colour space are given in figure 9.3(top,right). As can be observed in figure 9.3(bottom,left), all three regions form a rather compact cluster in colour space when all the image pixels are considered, what makes difficult producing a segmentation in the true three regions. However, if only the pixels for the adjacent regions in the middle and on the right of figure 9.3(top,left) are taken into account, the cloud of points of figure 9.3(bottom,left) clarifies in a structure of two clusters clearly distinguishable, as it is shown in figure 9.3(bottom,right). In other words, a local strategy allows putting the different regions into the context of their surroundings what can greatly contribute to produce a segmentation better than a strategy based on global analysis.

Following the local strategy and given two adjacent regions, they must be merged if there is evidence that they share a common reflectance. A first condition for this to occur is that their respective pixels lie on the same straight line in colour space. Nevertheless, this condition can only check that the corresponding reflectances have the same orientation. Two regions can still have their pixels aligned in colour space but the corresponding reflectances be different, as it is shown in figure 9.4. In such a case, colour space turns out to be useless unless both regions define separate clusters, which is just the opposite of what happens in the case of figure 9.4. A more general approach is thus needed.

In figure 9.4, both regions can be distinguished because of the high colour variation along the border separating them. In fact, it is an even more local cause what justifies putting them apart in the final segmentation, since that variation in intensity would not be meaningful if in the neighbourhood of the

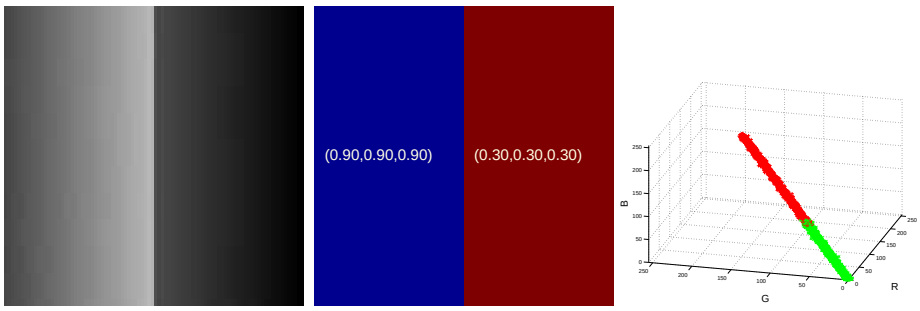


Fig. 9.4. [left] A synthetic image of a scene containing two materials whose reflectances are parallel to one another. [middle] Segmentation ground truth and the reflectances of both regions. [right] Pixels in colour space, whole image.

border, and within the regions, such variation was not infrequently high. However, from section 8.6, we know that, on situations where reflectances are parallel to each other, an ambiguity results as for the determination of the source of an intensity variation, so that, from a theoretical point of view, it could be either due to an abrupt change in the surface normal or a change in reflectance. As already discussed in section 8.6, adding the smoothness constraint for scene surfaces allows reducing the source of a large intensity change to a reflectance transition, although at the expense of classifying sides of the same object and same material in different regions. Nevertheless, the amount of colour variation along the border is the only clue available to make the decision without other knowledge than the sheer colour signal.

Summing up, two conditions must be met for the merging to occur: (1) the pixels of both regions must be "aligned" in colour space; and (2) a low colour variation, relative to the change in colour within the regions involved, must be found along the border separating both regions. The following explains how (1) and (2) are implemented in C^3S:

- To check condition (1), pixels corresponding to every pair of adjacent regions are clustered over the RGB colour space to try to determine whether both regions constitute in fact an only linear cluster. A *possibilistic C-means* algorithm (PCM) has been used for this purpose (see appendix D). In its general formulation, this is an iterative method which assigns a possibility $u_{ij} \in [0, 1]$ to every point x_i of the data set depending on its distance to everyone of the considered clusters θ_j, and then recomputes the representative of every cluster using all the points of the data set weighted by their possibilities of pertaining to the cluster. This process is repeated until convergence. Besides, because of the concept of possibility embedded into the clustering scheme, PCM possess the so-called mode-seeking property. As a consequence, if the structure of the data set consists of less clusters than initially hypothesized, after convergence, some cluster representatives will be more or less similar to one another, giving, thus, a correct interpretation of the data set structure.
  In this particular application of PCM, for every pair of regions, two clusters are initially guessed. The lines best describing each cluster, produced by PCA from the respective covariance matrices, are used as the initial cluster representatives. After every iteration, the cluster descriptors are updated, once again by means of PCA, but now using a covariance matrix computed using the pixels of both regions weighted by the corresponding possibilities u_{ij}. Furthermore, the orthogonal distance between a point and a straight line is used to determine the distance between pixels and cluster representatives. Finally, after convergence, function ζ_{pcm} is evaluated as follows:

$$\zeta_{pcm} = \frac{1}{n_1 + n_2} \sum_{i=1}^{n_1+n_2} |u_{i1} - u_{i2}| \tag{9.2}$$

  where θ_1 and θ_2 represent both clusters and n_1 and n_2 are the respective number of pixels. If $\zeta_{pcm} \leq \tau_{pcm}$ (i.e. all the pixels have a similar possibility of pertaining to the two clusters), both regions are considered aligned in colour space.

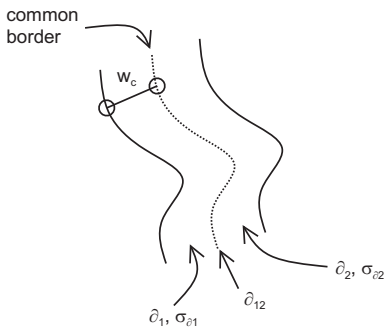


Fig. 9.5. Strips used to compute the statistical measures involved in ζ_{con}.

PCM formulation includes two parameters which, as indicated in appendix D, determine the quality of the clusters produced. On the one hand, the *scale* parameter η_j for every cluster should be set to the expected thickness of the wanted linear clusters, as suggested in [134] for clustering applications where linear structures are sought. Since such thickness is the consequence of image noise, η_j should be, thus, related with intensity uncertainties. As λ_{low} has been set as a thickness bound for a region to be considered valid during the growing stage, $\eta_j = \lambda_{low}, \forall j$, seems to be a good election. As for the *fuzzifier* q, a value of 1.1 has been chosen to make possibility functions decay rapidly when the distance between a pixel and the cluster representative surpasses η_j.

- To check condition (2), the change in colour along the common border must be compared with the colour variation due to changes in curvature in the nearness of the common border within each region. In this work, the colour variation is determined taking the pixel-by-pixel Euclidean distances in colour space. In other words, for a pixel not in the border but close to it, the eight distances with its eight neighbours are used; in case it is in the border, the distances involving this pixel and any other neighbour belonging to the opposite region are considered. Next, the mean pixel-by-pixel distance across the regions' common border $\overline{\partial}_{12}$ is compared with the mean pixel-by-pixel distance at every side of the border within each region, $\overline{\partial}_1$ and $\overline{\partial}_2$, as follows:

$$\zeta_{con} = \max\left\{\frac{\overline{\partial}_{12} - \overline{\partial}_1}{\sigma_{\partial 1}}, \frac{\overline{\partial}_{12} - \overline{\partial}_2}{\sigma_{\partial 2}}\right\}, \tag{9.3}$$

where $\sigma_{\partial 1}$ and $\sigma_{\partial 2}$ are the standard deviations of pixel-by-pixel distances for each region and are used to normalize the differences $\overline{\partial}_{12} - \overline{\partial}_1$ and $\overline{\partial}_{12} - \overline{\partial}_2$. In order to use only local information, the statistical measures involved in the computation of ζ_{con} are taken along a strip of width w_c at both sides of the border (see figure 9.5). If $\zeta_{con} \leq \tau_{con}$, the intensity variation across the border separating both regions is not considered relevant.

Once all the pairs of adjacent regions have been considered, those that fulfill conditions (1) and (2) are inserted in a heap ordered in ascending order, according to the value of ζ_{pcm} and ζ_{con}, where ζ_{con} is the main ordering key. In this way, the adjacency in the top of the heap corresponds to the pair of regions which are more likely to share a common reflectance. After their merging, the heap is reorganized to remove the old adjacencies and insert the new ones. That is to say, adjacencies are reconsidered after every merging and the most similar regions are always merged first. A more formal description of the merging stage of C^3S can be found in pseudocode 9.1.

To finish with this section, observe that there is a certain parallelism between conditions (1) and (2) and, respectively, the detection of edges CHC/NCD/NDD and GRD. In a certain sense, condition (1) and edges CHC/NCD/NDD are related with the detection of changes in the orientation of reflectance, while condition (2) and edges GRD are linked with changes between parallel reflectances. Given these similarities, it can seem redundant to use different techniques for detecting the same. However: (i) condition (1) provides an analysis finer than edges CHC/NCD/NDD; and (ii) conditions (1) and (2) make use of regional information which, when computing the edge map, is not available. Therefore,

Pseudocode 9.1 Algorithmic formulation of the merging stage of C^3S.

```
(1) initialization
        HeapInitialize(H);
        for every region R_i
        for every region R_j such that j > i
            if AdjacentRegions(R_i,R_j) then
                if ζ_con(R_i, R_j) ≤ τ_con and ζ_pcm(R_i, R_j) ≤ τ_pcm then
                    HeapInsert(H, R_i, R_j, ζ_con(R_i, R_j), ζ_pcm(R_i, R_j));
                endif
            endif
        endfor
        endfor
(2) while not EmptyHeap(H) do
        (R_i, R_j) = HeapRoot(H); HeapRemoveRoot(H);
        for every region R_k such that k ≠ i and k ≠ j
            if AdjacentRegions(R_k,R_i) then HeapRemove(H,R_k,R_i);
            endif
            if AdjacentRegions(R_k,R_j) then HeapRemove(H,R_k,R_j);
            endif
        endfor
        R_n = MergeRegions(R_i,R_j); DeleteRegion(R_i); DeleteRegion(R_j);
        for every region R_k such that k ≠ n
            if AdjacentRegions(R_k,R_n) then
                if ζ_con(R_k, R_n) ≤ τ_con and ζ_pcm(R_k, R_n) ≤ τ_pcm then
                    HeapInsert(H, R_k, R_n, ζ_con(R_k, R_n), ζ_pcm(R_k, R_n));
                endif
            endif
        endfor
    endwhile
```

conditions (1) and (2) and the whole merging stage are expected to be more powerful, with the aid of the edge map, to locate accurately reflectance changes.

9.3.2 Merging of specularities

General description

As has been mentioned before, it is possible that, in certain cases, a specularitity is surrounded by edges in the edge map and, consequently, its pixels result separated from the rest of pixels of the object (i.e. in a different region) after the region growing. In fact, because of the high intensity gradients which can be expected at the "base" of any specularity, it is likely that they still appear isolated in separate regions after the first step of the merging stage, and, thus, special measures must be taken to make specularities join their respective main regions.

The problem of detecting and handling specularities has been considered in several studies within the physics-based vision research area. The search of a certain configuration of clusters in colour space, the use of photometric invariants or transformed colour spaces, or the use of light polarizers are among the main approaches available (see [10, 68, 129, 179, 186, 249, 250, 268, 287] among many others). On the one hand, some of them do not tolerate, from a theoretical point of view, ambient illumination. On the other hand, some of them need several images of the same scene to accomplish the separation of reflection components. Finally, others assume a segmented image. The approach described below results quite simple to implement and manages to merge those specularities which, due to the way the edge map is built, are typically put in a separate region. In particular, the approach performs well because does not depend on the size of the specularities (contrary to methods based on particular configurations of clusters) or the amount of image noise or camera clipping (contrary to methods

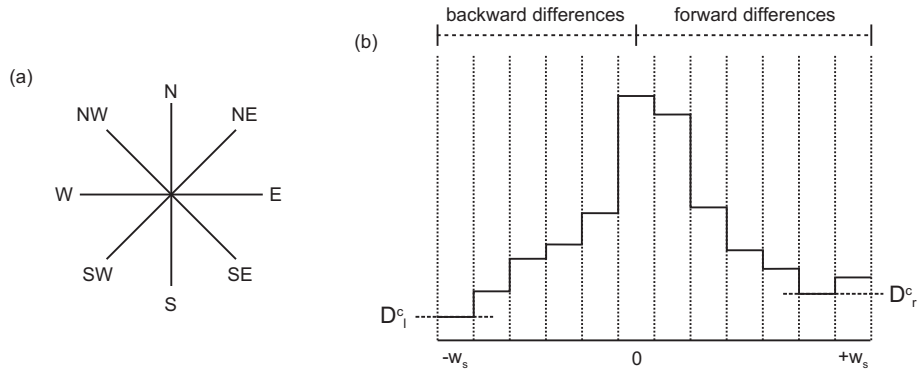


Fig. 9.6. Detection of specular pixels: (a) the four orientations examined; (b) example of neighbourhood and symbols involved.

relying on photometric invariants or transformed colour spaces). Furthermore, no special equipment is needed (contrary to methods based on light polarization).

The method proposed here makes use of an approach similar to [22,294], where regions in the image looking too bright to be Lambertian are found. To this end, any pixel in the image with the largest intensity within a neighbourhood of high dynamic range together with a large gradient, both in the same colour channel, is considered a candidate to be the "peak" of a possible specularity. Next, all the surrounding pixels whose intensity is above a certain percentage of the peak intensity are incorporated into the group. The spatial dimension of the set is finally computed to discard candidates not having point-wise or linear-wise shapes.

Once specular pixels have been located, those regions which contain a high percentage of them are considered to be one of those regions which get isolated because of the proliferation of edges around specularities. Accordingly, C^3S merges them with one of their respective adjacent regions. Since specularities are typically found in the middle of matte regions, "specular regions" usually have an only adjacent region, which surrounds them. In that case, there is an only possibility of merging and no decision has to be made. When there are two or more adjacent regions, C^3S selects the adjacent region with the largest common border. It is true that this decision seems rather arbitrary. However, it has resulted to work quite well during the experiments performed, as will be seen later. On the other hand, observe that the incidence of the particular criterion chosen is quite reduced because the likely situation corresponds to one adjacency per specular region.

Implementation details

From the implementation viewpoint, specular pixels are detected by scanning the whole image, looking for local maxima in bright areas of the image; therefore, no information from the previous segmentation stages is used and just the input image is involved in the process. The search for local maxima is performed along four one-dimensional neighbourhoods centered at every image pixel and oriented in directions N-S, E-W, NW-SE and NE-SW (see figure 9.6).

Once a local maximum is found, the dynamic range along a neighbourhood of $2w_s + 1$ pixels in the direction of detection is considered. First, minimum intensities at every side of the neighbourhood are obtained, $D^c_l = \min\{D^c(k), k = -w_s, \ldots, 0\}$ and $D^c_r = \min\{D^c(k), k = 0, \ldots, w_s\}$. Next, quotients $\mathrm{DR}^c(k) = D^c(k)/D^c_l, k = -w_s, \ldots, 0$, and $\mathrm{DR}^c(k) = D^c(k)/D^c_r, k = 0, \ldots, w_s$, are computed. For $k = 0$, the pixel under consideration, $\mathrm{DR}^c(k) = \max\{D^c(0)/D^c_l, D^c(0)/D^c_r\}$.

At the same time, normalized gradients are calculated for every pixel in the neighbourhood. More precisely, for every pixel, quotients $(D^c(k) - D^c(k-1))/L^c_d, k = -w_s, \ldots, 0$, and $(D^c(k+1) - D^c(k))/L^c_d, k = 0, \ldots, w_s$, are determined, where $L^c_d = q^c_0 \int_\Lambda L_d(\lambda)\tau^c(\lambda)s(\lambda)d\lambda$ is the (scaled) radiance coming from the directional light source for colour channel c (already introduced in section 8.2, proposition 8.8). If pixels $k-1$, k and $k+1$ are all assumed to correspond to the same scene material

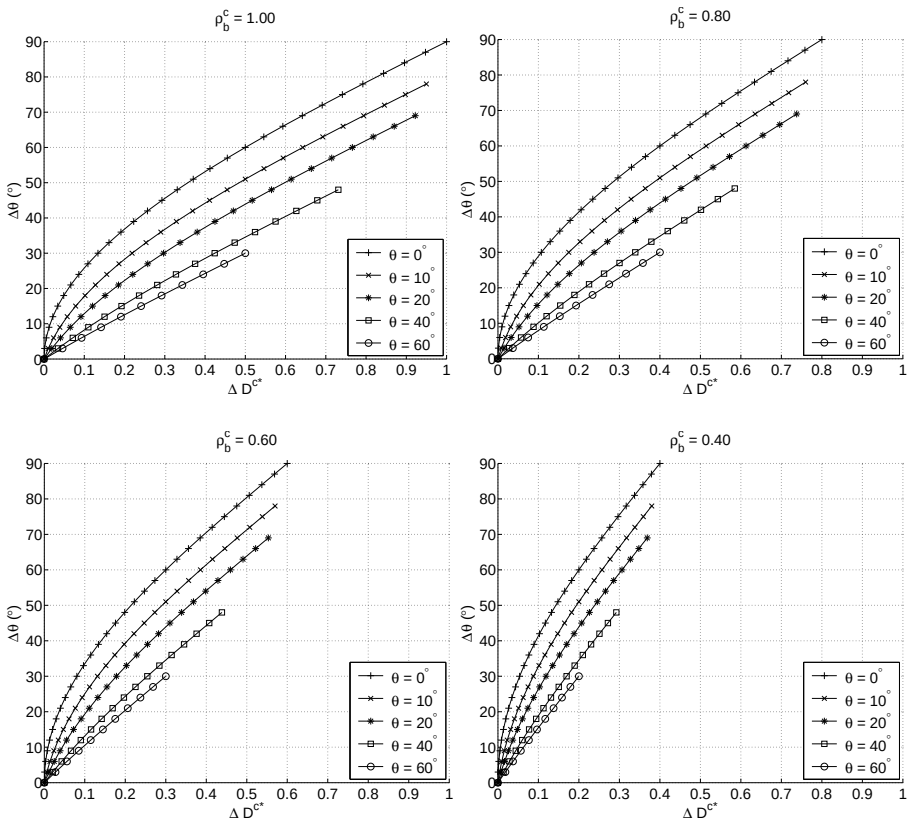


Fig. 9.7. Plots of $\Delta\theta$ for different values of ΔD^{c*}, θ and ρ_b^c.

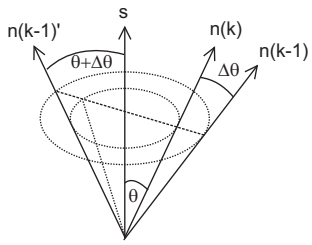


Fig. 9.8. Relationship between $\Delta\theta$ and surface normals of pixels involved.

and they do not show interface reflection, those quotients have the meaning expressed in equations 9.4 and 9.5:

$$\Delta D_{fw}^{c*}(k) = \frac{D^c(k+1) - D^c(k)}{L_d^c} = (m_b(k+1) - m_b(k))\rho_b^c\,, \tag{9.4}$$

$$\Delta D_{bk}^{c*}(k) = \frac{D^c(k) - D^c(k-1)}{L_d^c} = (m_b(k) - m_b(k-1))\rho_b^c\,, \tag{9.5}$$

where ρ_b^c was already introduced in equation 8.5. In this way, the gradient measures of equations 9.4 and 9.5 no longer depend on the illumination strength and both lie within the interval $[-1, +1]$.

Now, to investigate further into the meaning of $\Delta D_{bk}^{c*}(k)$ and $\Delta D_{fw}^{c*}(k)$, let us expand $m_b(k)$ and $m_b(k-1)$ as, respectively, $\cos(\theta)$ and $\cos(\theta + \Delta\theta)$. Figure 9.7 plots $\Delta\theta$ as a function of $\Delta D^{c*}(k)$ (i.e. any of $\Delta D_{bk}^{c*}(k)$ or $\Delta D_{fw}^{c*}(k)$), θ and ρ_b^c on the basis of equations 9.4 and 9.5. From these plots it can be seen that, for instance, $\Delta D^{c*}(k) = 0.3$ implies $\Delta\theta$ is, at least, almost 20°, irrespectively of ρ_b^c

Pseudocode 9.2 Test used by C^{3}S to detect a specularity peak. (**break** exits the loop.)

```
specularityPeak := false;
if DR^c(0) ≥ τ_sdr then
    ok1 := false; ok2 := false;
    for k := 0 to w_s
        if DR^c(−k) ≥ τ_sdr and ΔD^c*_bk(−k) ≥ τ_sgr then
            ok1 := true;
        endif
        if DR^c(k) ≥ τ_sdr and −ΔD^c*_fw(k) ≥ τ_sgr then
            ok2 := true;
        endif
        if ok1 and ok2 then break; endif
        if DR^c(−k) < τ_sdr and DR^c(k) < τ_sdr then break; endif
    endfor
    if ok1 and ok2 then specularityPeak := true; endif
endif
```

Table 9.1. Parameter values used for the images of figure 9.9.

parameter	left image	center image	right image
w_s	5	5	5
τ_{sdr}	1.5	1.5	1.5
τ_{sgr}	0.3	0.3	0.3
τ_{sppc}	0.66	0.66	0.66
τ_{sdim}	1.0	1.0	2.0

and θ. Now, observe figure 9.8, where $\Delta\theta$ is related with surface normal vectors at the corresponding image locations. As can be seen, given $\Delta\theta$, the angle between surface normal vectors lies in the interval $[\Delta\theta, 2\theta + \Delta\theta]$. Therefore, following the example, if $\Delta D^{c*}(k) = 0.3$, the angle between surface normal vectors is, at least, almost 20°. Summing up, ΔD^{c*}, a function of the intensity change between adjacent pixels, has been related with the minimum change in orientation which supports it, provided that the pixels involved do not show interface reflection. Therefore, if smooth surfaces are expected in the scene, so that changes in orientation between adjacent pixels should not be above, say 10°, and a change of orientation of, at least, 20° results, then we can deduce the intensity change cannot only be imputed to the scene objects curvature but to a wrong hypothesis: at least one of the pixels considered shows interface reflection, or there is a reflectance transition.

Once $\mathrm{DR}^c(k)$, $\Delta D^{c*}_{bk}(k)$ and $\Delta D^{c*}_{fw}(k)$ have all been calculated, given thresholds τ_{sdr} and τ_{sgr}, the test shown in figure 9.2 is applied for colour channel c. As can be observed, this code looks for a peak in intensity like the one shown in figure 9.6(b) in order to discard the aforementioned possibility of being a change in reflectance, instead of a specularity, the cause of a high dynamic range or a large gradient. In case the result of the test is positive for any colour channel and any of the above-mentioned directions, the pixel under consideration is labelled as a specularity peak.

Once the image has been scanned and all specularity peaks have been detected, the image area covered by the specularity associated to every specularity peak is located by determining the connected component including the peak and consisting of the surrounding pixels having an intensity level above a certain percentage τ_{sppc} of the peak corresponding intensity. Its spatial dimension is computed next. In case it cannot be considered a point or a line, the area detected is considered a false specularity and therefore discarded. The spatial dimension is found applying PCA to the set of image locations covered by the connected component. To be considered describable by a point or a line, at least one of the two corresponding eigenvalues is required to be below threshold τ_{sdim}.

By way of example, figure 9.9 shows the specularities detected for three images where specular reflection is noticeable. As can be observed, all the specularities have been correctly located, although, in the image on the left, there is a false specularity detected in the flowerpot because a whitish

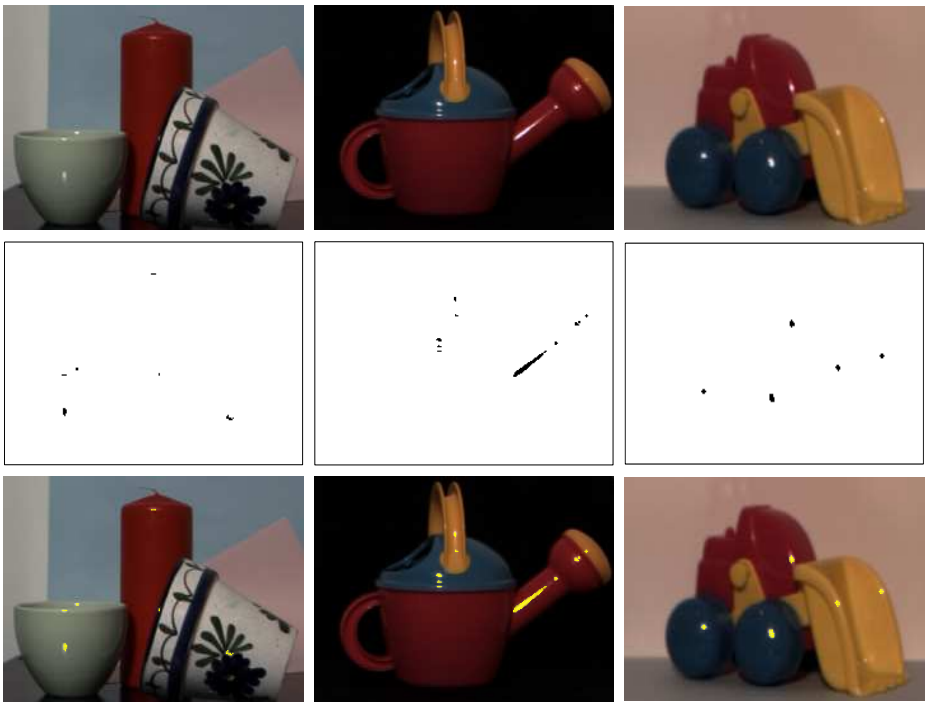


Fig. 9.9. Examples of detection of specularities: [1st row] original images; [2nd row] specularities detected in black; [3rd row] specularities superimposed in yellow over original images.

image area is between two blackish image areas, which gives rise to a high dynamic range and a large gradient, together with an adequate spatial dimension, which misleads the algorithm. Table 9.1 details the parameter values for this experiments. As can be seen, only τ_{sdim} had to be tuned for every image. This is logical since it depends on the size in pixels of the specularity, and this changes with image resolution, distance from the camera to the scene and the geometry of the glossy object; w_s can also show this dependence although the experiments have not shown a dependence so critical as in the case of τ_{sdim}. The remaining parameters, however, are quite stable. Finally, τ_{sppc} was set to $0.66 = \frac{2}{3}$ as suggested in [22].

To finish, once the map of specularities has been obtained, those regions which include specular pixels in a number above a certain percentage τ_{spct} join the adjacent region with which the number of pixels in the common border is the largest. Most times, however, regions containing specularities are totally surrounded by an only region (i.e. the number of adjacent regions is just 1). During the experiments performed, 50% of pixels have been required to be specular for a region to be qualified as a specularity put apart during the region growing stage.

9.3.3 An example of merging

By way of example of the output of the region merging stage, figure 9.10(right) shows the final segmentation for the image on figure 9.10(left), while the output of the growing stage is also repeated in figure 9.10(middle) for comparison purposes. As can be observed, most of the oversegmentation is removed after the merging stage. Only the edge corresponding to the rim of the cup and the shadow in the candle remain.

Fig. 9.10. Example of output of the region merging stage: [left] original image; [middle] output of the growing stage; [right] final segmentation.

9.4 Experimental Results

This section presents a set of results for C^3S for both synthetic and real images. Some of these experiments aim at showing the performance of C^3S when the different parameters involved are varied, while others involve other segmentation algorithms in a performance comparison. More specifically, three segmentation methods have been considered: the physics-based algorithms by Klinker et al. [129] (KLN from now on) and Gevers [68] (GEV from now on), together with the non-physics-based algorithm by Comaniciu and Meer published in [29][1] (C&M from now on).

As for parameters, KLN, GEV and C^3S have both a number of parameters which have to be set up according to the requirements of the particular experiment, in order to obtain the best performance; accordingly, they are explicitly indicated in every case. The $C_1C_2C_3$ space has been selected among the three colour spaces suggested by Gevers in [68] as the working space for the GEV algorithm, since it provides invariance to objects curvature; the RGB space will also be used in an experiment with real images, as will be seen later. The third colour space, $L_1L_2L_3$, which requires white-balanced cameras, will however not be used in any moment of the comparison. On the other hand, C&M selects automatically the algorithm parameters according to the type of result required: under-segmentation, over-segmentation and quantization. Given that non-physics-based segmentation algorithms tend to over-segment the image, the under-segmentation option is chosen in all the experiments where C&M is involved.

As for C^3S post-processing stage, this time no special actions were taken over the segmentation output.

Finally, as suggested in section 4.4.2, segmentation performance will be measured regarding both the classification and the contour localization aspects. Unless otherwise stated, the CG(90), OS(90) and US(90) measures defined in section 4.4.2 will be used for the former, while the second aspect will be measured mostly by means of Baddeley's measure. (In the two following sections, the (90) will be removed from CG, OS and US for the sake of simplicity.)

9.4.1 Experiments with synthetic images

The experimental results section of chapter 8 showed that the performance of C^3E did not depend on whether the scene objects shading was noticeable in the processed images or not. Because of this, this section for evaluating the performance of C^3S will directly present results for scenes with an important presence of objects shading. As on other occasions, synthetic images are contaminated with noise following pseudocode 6.3 and the camera noise parameters estimated in section 6.4.1, what will be referenced as normal noise conditions.

On the other hand, the values given to the different parameters determining C^3S behaviour are indicated in tables 9.2(a), (b), (c) and (d). As can be observed, many of those parameters were left constant, and the remaining ones were put as a function of the same parameter t of C^3E. As the experimental results will show below for synthetic images and in the next section for real images, this configuration of parameters for C^3S allows it to successfully segment many types of images, what

[1] The code is publicly available at http://www.caip.rutgers.edu/~comanici/segm_images.html

Table 9.2. Parameter values used by C^{3}S in the experiments with synthetic images.

(a)

edge map computation	
t_1	same value as t
σ_{LOG}	1.0
t_2	same value as t

(b)

region growing	
t_3	same value as t
λ_{low}	$t_3 \max_c\{\delta(255^c)\}$
τ_{split}	same value as λ_{low}

(c)

merging of matte regions	
q	1.1
η	same value as λ_{low}
τ_{pcm}	0.1
w_c	5
τ_{con}	same value as t

(d)

specularities detection and merging	
w_s	5
τ_{sdr}	1.5
τ_{sgr}	0.3
τ_{sppc}	0.66
τ_{sdim}	1.0
τ_{spct}	50%
L_d^c	according to the synthetic scene

minimizes the burden of having to set up the *a priori* rather large number of parameters, reducing it to configuring just one parameter most times.

Figure 9.11 shows results for C^{3}S and the spheres and planes set, varying t between 1 and 6. The table of figure 9.11(a) provides average values and standard deviations for CG, OS, US and Baddeley's measures; besides, the complete range of measurements attained is also given under the term *100 % range*, as well as the interval covering 90% of the values. On the other hand, figures 9.11(b), (c), (d) and (e) plot average and standard deviation values in the form of error bar plots, together with the median values as dotted lines.

As can be observed, OS changes very little among the different cases, while CG decreases and US increases as t becomes larger. This is because, as t increases, the probability of having non-closed contours is bigger, what gives rise to regions mixing pixels from different true regions, and therefore leads to non-correctly grouped pixels and under-segmentation. Up until $t = 4$, the percentage of correctly grouped pixels at the 90% level is above 95%, while over-segmentation is below 10% and under-segmentation does not involve more than a 5% of the image. As for Baddeley's measure, it remains almost constant for all the cases and the results indicate that the error in the localization of reflectance transitions is, on average, lower than one pixel.

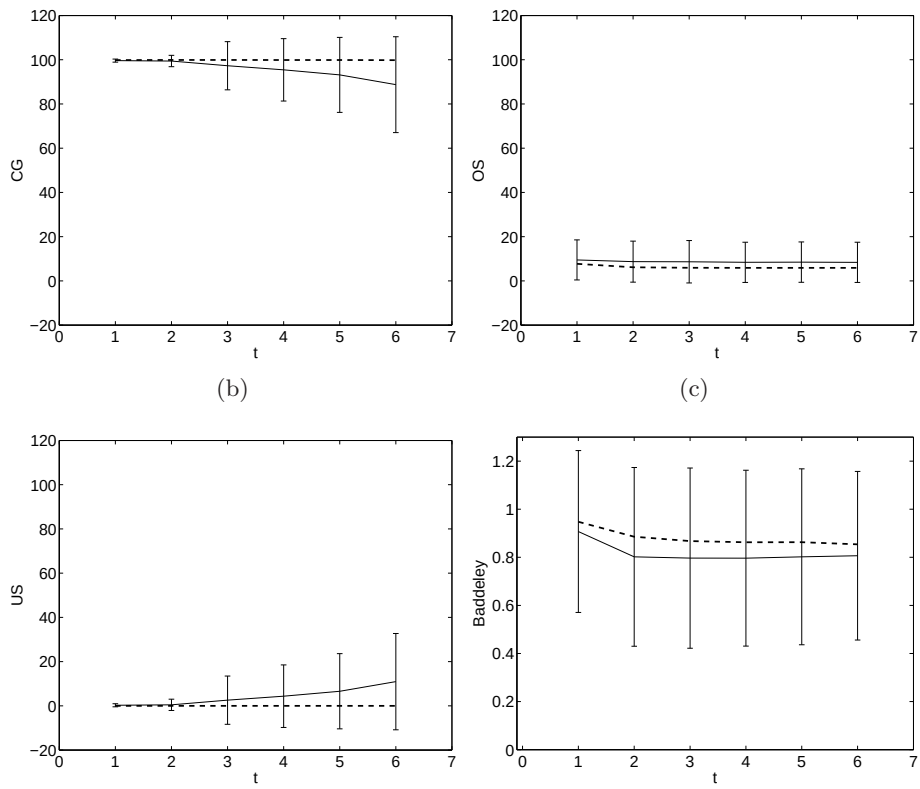
A more careful analysis of all these data through the 100% ranges also reveals that the performance associated to some images drops in a significant way: e.g. observe that, for $t = 3$, CG, OS and US attain extreme values such as 22.56%, 43.11% and 77.40%, respectively. However, the 90% ranges show a dramatic improvement in every measure, what indicates the presence of outliers. In other words, most times values of t up to 4 yield quite acceptable segmentation results, although some images can require a better and detailed configuration of parameters. This behaviour can also be deduced from the median values of the plots of figure 9.11: for instance, CG median is all the time 100%, what means that at least 50% of the images are perfectly segmented, within the 10% tolerance of pixels dispersion associated to CG(90).

In a second experiment, C^{3}S was compared against KLN and GEV under normal noise conditions. The corresponding results can be found in figure 9.12. As can be observed, C^{3}S outperforms both KLN and GEV for the three values considered for t, 2, 3 and 4. The improvement in performance is clear for the case of GEV, while the differences are rather narrow for the case of KLN. Besides, observe that C^{3}S localizes reflectance transitions in a more accurate way. The parameters used for KLN and GEV corresponded to the ones providing better performance in a series of trials. The respective values can be found in tables 9.3 and 9.4.

Finally, all three algorithms were faced against increasing noise conditions. The procedure for generating the different sets of noisy synthetic images was the same followed for comparing C^{3}E and S&G in chapter 8. As for that experiment, intensity uncertainties were re-calculated for C^{3}S for every

	$t \rightarrow$	1	2	3	4	5	6
CG	avg - std	99.62 - 0.69	99.42 - 2.57	97.29 - 10.89	95.45 - 14.11	93.17 - 16.96	88.75 - 21.67
	100% range	[94.42,100.00]	[75.98,100.00]	[22.56,100.00]	[22.56,100.00]	[22.56,100.00]	[13.76,100.00]
	90% range	[98.84,100.00]	[98.34,100.00]	[78.38,100.00]	[59.61,100.00]	[53.74,100.00]	[33.72,100.00]
OS	avg - std	9.47 - 9.05	8.69 - 9.27	8.65 - 9.59	8.39 - 9.09	8.47 - 9.12	8.38 - 9.09
	100% range	[0.00,43.11]	[0.00,43.11]	[0.00,43.11]	[0.00,43.11]	[0.00,43.11]	[0.00,43.11]
	90% range	[0.00,26.41]	[0.00,27.80]	[0.00,29.31]	[0.00,25.12]	[0.00,25.12]	[0.00,25.12]
US	avg - std	0.25 - 0.69	0.45 - 2.56	2.56 - 10.91	4.36 - 14.14	6.58 - 17.01	10.94 - 21.78
	100% range	[0.00, 5.58]	[0.00,23.81]	[0.00,77.40]	[0.00,77.40]	[0.00,77.40]	[0.00,86.24]
	90% range	[0.00, 1.00]	[0.00, 1.62]	[0.00,21.45]	[0.00,40.34]	[0.00,46.25]	[0.00,66.19]
Baddeley	avg - std	0.91 - 0.34	0.80 - 0.37	0.80 - 0.37	0.80 - 0.37	0.80 - 0.37	0.81 - 0.35
	100% range	[0.21, 1.67]	[0.01, 1.58]	[0.00, 1.58]	[0.00, 1.58]	[0.00, 1.58]	[0.00, 1.58]
	90% range	[0.34, 1.43]	[0.08, 1.34]	[0.06, 1.33]	[0.08, 1.33]	[0.06, 1.33]	[0.07, 1.33]

(a)



(b) (c) (d) (e)

Fig. 9.11. Experimental results for C^3S and noisy synthetic images: (a) every entry of the table consists of average - standard deviation, [minimum,maximum] and [5% percentile, 95% percentile]; (b,c,d,e) plots of average values and standard deviations for $t = 1..6$ for, respectively, CG, OS, US and Baddeley performance measures; the dashed line corresponds to the median.

	C^3S			KLN	GEV
	$t = 2$	$t = 3$	$t = 4$		
CG	99.42 - 2.57	97.29 - 10.89	95.45 - 14.11	92.89 - 11.71	74.24 - 22.70
	[75.98,100.00]	[22.56,100.00]	[22.56,100.00]	[27.75,100.00]	[6.95,96.03]
OS	8.69 - 9.27	8.65 - 9.59	8.39 - 9.09	9.19 - 12.54	30.06 - 20.85
	[0.00,43.11]	[0.00,43.11]	[0.00,43.11]	[0.00,55.43]	[1.28,100.00]
US	0.45 - 2.56	2.56 - 10.91	4.36 - 14.14	6.42 - 11.70	23.70 - 23.34
	[0.00,23.81]	[0.00,77.40]	[0.00,77.40]	[0.00,72.00]	[1.34,93.03]
Baddeley	0.80 - 0.37	0.80 - 0.37	0.80 - 0.37	1.15 - 0.30	1.67 - 0.27
	[0.01, 1.58]	[0.00, 1.58]	[0.00, 1.58]	[0.26, 1.88]	[1.08, 2.45]

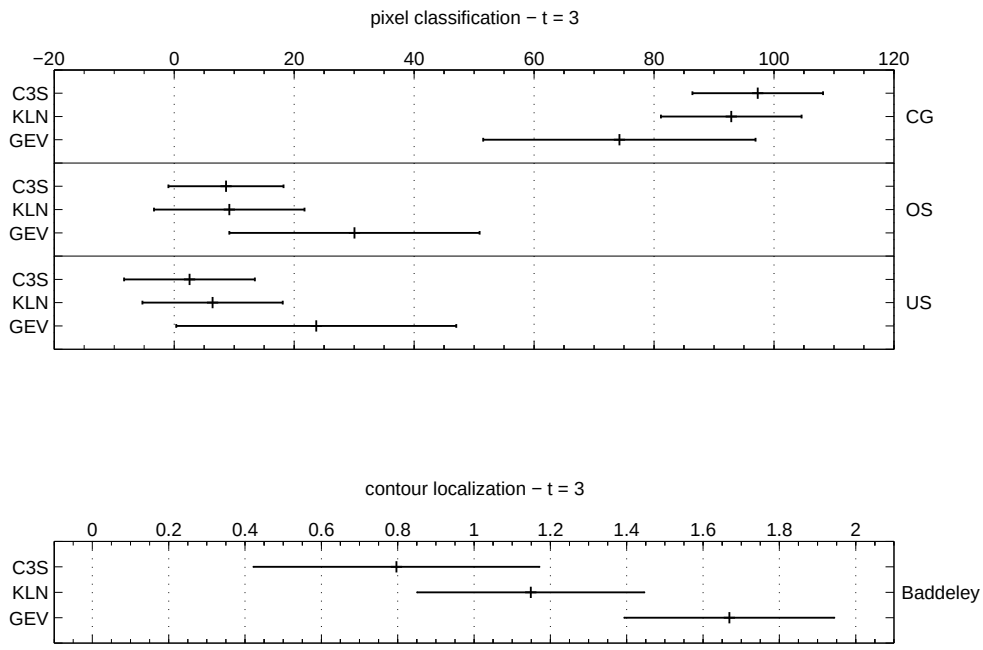


Fig. 9.12. Comparison between C^3S, KLN and GEV under normal noise conditions by means of synthetic images: [top] every entry of the table contains average - standard deviation [minimum,maximum]; [bottom] plots of average and standard deviation values for $t = 3$.

Table 9.3. Parameters for KLN during the comparison with C^3S by means of synthetic images. (The description of the parameters corresponds to the one used in the original paper [129, table 6].)

parameter description	**value**				
	noise×1	noise×2	noise×3	noise×4	noise×5
window size (for initial grouping)			10 × 10		
minimal area size			100 pixels		
camera noise	2	3	6	9	15
cylinder width	6	8	16	24	30
minimal intensity			15		

Table 9.4. Parameters for GEV during the comparison with C^3S by means of synthetic images.

parameter description	value
standard deviation for gaussian derivative operators	1.0
maximum standard deviation for homogeneous regions (normalized between 0 and 255)	5.0
Tlow for hysteresis thresholding	0.9
Thigh for hysteresis thresholding	0.9

set. The parameters of KLN were accordingly modified in order to obtain the best performance under the different operating conditions (see table 9.3). As for GEV, curiously, it produced better results if the parameters were left constant and equal to the ones given in table 9.4.

The results for this experiment can be found in table 9.5 and in figure 9.13. This time C^3S outperforms KLN and GEV under all noise conditions only for $t = 2$; for the other two values considered for t, sometimes KLN or even GEV exhibit better performance. On the other hand, C^3S and KLN produce worse results as noise increases, as could be easily guessed. Observe that C^3S seems to show a lower tolerance to noise with regard to C^3E, since, as noise is increased, the trend towards lower performance is steeper for the former; KLN also exhibits this behaviour. GEV, however, is almost insensitive to noise, yielding, in general, lower performance than the other two but all the time at the same level. Once again, C^3S localizes reflectance transitions clearly better than KLN and GEV.

9.4.2 Experiments with real images

This section comments on the performance of C^3S against real images using the same strategy followed for C^3E in chapter 8. That is to say, a varied set of real images for which a ground truth has been manually generated will be used to provide quantitative information, while results for additional images will also be provided to illustrate better the performance of C^3S. As on other occasions, comparative results will also be presented, including this time not only KLN and GEV algorithms, but also the non-physics-based C&M algorithm. On the other hand, this time the *missing rate* e_R^m, the *false alarm rate* e_R^f and the *overall performance* p_R measures by Huang and Dom [106] will also be provided to help in the comparison and the discussion about the performance of C^3S.

Figure 9.14 presents again the six real images for which a ground truth is available, in order to make easier referring to them throughout this section (they already appeared in figure 8.19).

In a first experiment, the image of figure 9.14(a) was again used as a test bench for showing the performance evolution of C^3S alone as parameter t was varied. As can be observed, this time parameter t becomes critical since a wrong value leads to severe under-segmentations. In this particular case, $t = 2$ results to be the best election since, for $t = 3$, the amount of under-segmentation experiments a dramatic increase, while, for $t = 1$, the amount of over-segmentation attains about 50% of the image. This fact is confirmed by p_R, which achieves its maximum for $t = 2$.

In a second experiment, C^3S, KLN, GEV and C&M were faced against all the images of figure 9.14, as was also done in chapter 8 to test C^3E. As was also commented on that occasion, part of those images can slightly deviate from the assumed reflection model, but, as before, this is considered useful since allows observing the degree of tolerance of the different algorithms to the so-called *scene noise* introduced in section 4.4.3. The quantitative results of the experiment are shown in table 9.6 and in figure 9.16. While, again, only parameter t was allowed to take free values for C^3S, best results for KLN and GEV are provided, being the corresponding parameters the ones of, respectively, tables 9.7 and 9.8; as for C&M, the under-segmentation option was selected in all cases, so that no other parameter needed to be set up. Finally, lighting parameters which were available for C^3S specularity merging stage are indicated in table 9.9; when they were not available, white light was assumed.

As a global result of this experiment, figures 9.16(a) and (c) show that C^3S was able to produce the results with the largest percentages of correctly grouped pixels and lowest levels of under-segmentation,

Table 9.5. Comparison between C^3S, KLN and GEV under increasing noise conditions by means of synthetic images: every entry of the table consists of average - standard deviation [minimum,maximum].

noise	measure	C^3S			KLN	GEV
		$t = 2$	$t = 3$	$t = 4$		
×1	CG	99.42 - 2.57 [75.98,100.00]	97.29 - 10.89 [22.56,100.00]	95.45 - 14.11 [22.56,100.00]	92.89 - 11.71 [27.75,100.00]	74.24 - 22.70 [6.95,96.03]
	OS	8.69 - 9.27 [0.00,43.11]	8.65 - 9.59 [0.00,43.11]	8.39 - 9.09 [0.00,43.11]	9.19 - 12.54 [0.00,55.43]	30.06 - 20.85 [1.28,100.00]
	US	0.45 - 2.56 [0.00,23.81]	2.56 - 10.91 [0.00,77.40]	4.36 - 14.14 [0.00,77.40]	6.42 - 11.70 [0.00,72.00]	23.70 - 23.34 [1.34,93.03]
	Baddeley	0.80 - 0.37 [0.01, 1.58]	0.80 - 0.37 [0.00, 1.58]	0.80 - 0.37 [0.00, 1.58]	1.15 - 0.30 [0.26, 1.88]	1.67 - 0.27 [1.08, 2.45]
×2	CG	99.00 - 4.81 [58.49,100.00]	94.85 - 13.70 [33.67,100.00]	92.35 - 16.61 [16.71,100.00]	95.55 - 9.93 [62.91,100.00]	77.32 - 17.38 [26.46,95.91]
	OS	9.58 - 9.50 [0.00,38.61]	9.04 - 9.13 [0.00,37.65]	9.14 - 9.28 [0.00,37.65]	6.13 - 11.03 [0.00,52.98]	27.91 - 21.54 [0.23,100.00]
	US	0.90 - 4.83 [0.00,41.49]	5.03 - 13.69 [0.00,66.32]	7.49 - 16.61 [0.00,83.29]	4.21 - 9.77 [0.00,37.00]	20.43 - 17.90 [1.35,72.62]
	Baddeley	0.73 - 0.34 [0.00, 1.42]	0.73 - 0.34 [0.00, 1.30]	0.73 - 0.34 [0.00, 1.32]	1.35 - 0.33 [0.04, 1.87]	1.61 - 0.27 [1.07, 2.42]
×3	CG	96.51 - 11.37 [29.89,100.00]	86.96 - 21.93 [28.99,100.00]	76.01 - 28.38 [13.95,100.00]	88.06 - 16.93 [37.95,100.00]	79.36 - 15.80 [30.21,95.81]
	OS	8.99 - 9.63 [0.00,38.06]	8.55 - 9.42 [0.00,38.06]	8.53 - 9.42 [0.00,38.06]	8.28 - 11.91 [0.00,45.72]	30.36 - 19.78 [1.36,100.00]
	US	3.35 - 11.39 [0.00,70.11]	12.88 - 21.97 [0.00,71.01]	23.65 - 28.60 [0.00,86.05]	11.21 - 16.95 [0.00,61.91]	18.32 - 16.41 [1.35,69.10]
	Baddeley	0.74 - 0.31 [0.00, 1.52]	0.75 - 0.31 [0.00, 1.35]	0.80 - 0.32 [0.00, 1.64]	1.28 - 0.33 [0.42, 2.24]	1.66 - 0.29 [0.87, 2.65]
×4	CG	90.34 - 21.18 [18.75,100.00]	78.21 - 27.98 [9.89,100.00]	67.39 - 30.49 [1.84,100.00]	71.39 - 30.60 [0.20,100.00]	80.75 - 18.29 [4.97,96.88]
	OS	11.83 - 10.86 [0.00,40.46]	11.61 - 10.49 [0.00,40.46]	10.76 - 10.16 [0.00,40.01]	12.49 - 19.15 [0.00,100.00]	34.16 - 22.55 [0.61,90.96]
	US	9.46 - 21.20 [0.00,81.18]	21.40 - 28.16 [0.00,90.11]	32.19 - 30.66 [0.00,98.16]	28.03 - 30.93 [0.00,99.80]	16.96 - 18.87 [0.00,94.89]
	Baddeley	0.81 - 0.28 [0.00, 1.33]	0.83 - 0.29 [0.00, 1.44]	0.85 - 0.27 [0.00, 1.51]	1.63 - 0.34 [0.84, 2.56]	1.77 - 0.35 [1.15, 2.87]
×5	CG	86.83 - 20.34 [29.55,100.00]	69.35 - 30.99 [3.28,100.00]	56.18 - 35.27 [2.86,100.00]	63.56 - 27.99 [7.94,100.00]	77.59 - 18.82 [19.67,94.49]
	OS	9.81 - 9.78 [0.00,37.21]	9.54 - 9.62 [0.00,37.21]	9.33 - 9.75 [0.00,37.21]	18.40 - 17.31 [0.00,64.59]	34.35 - 20.10 [4.63,84.88]
	US	12.86 - 20.47 [0.00,70.45]	30.29 - 31.22 [0.00,96.72]	43.48 - 35.54 [0.00,97.14]	35.52 - 28.42 [0.00,92.06]	20.15 - 19.37 [1.70,80.03]
	Baddeley	0.81 - 0.31 [0.00, 1.39]	0.86 - 0.32 [0.00, 1.64]	0.91 - 0.28 [0.00, 1.47]	1.55 - 0.43 [0.82, 2.79]	1.87 - 0.36 [1.10, 3.01]

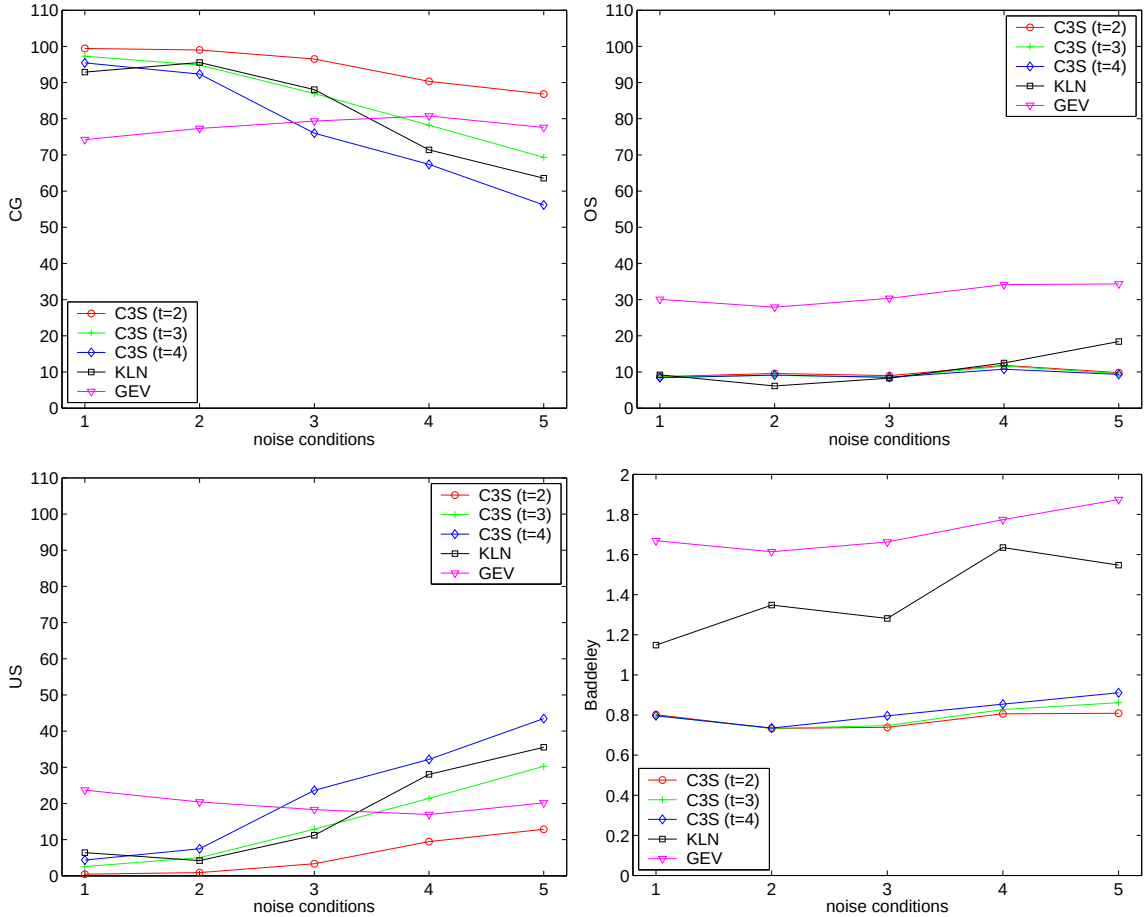


Fig. 9.13. Comparison between C^{3}S, KLN and GEV under increasing noise conditions by means of synthetic images: plots of average values for the performance measures.

specially for $t = 2$. As for the degree of over-segmentation (figure 9.16(b)), both C^{3}S and KLN seem to show better results than the rest, with more measurements in the leftmost part of the graph. The combined measure by Huang and Dom, p_R, however, does not show a preference for any of the algorithms (figure 9.16(d)). Finally, the accuracy in the localization of reflectance transitions (figure 9.16(e)) seems to be better in C^{3}S than in the rest, with more measurements for the former in the leftmost part of the graph again.

A more detailed analysis of the performance data by means of table 9.6 reveals that C^{3}S outperforms the other algorithms most times (i.e. the best performance measurement is most times in the columns corresponding to C^{3}S). This can be easily seen through the rows labelled $\sum \#best$, which accumulate, for every column of the table, the number of times that algorithm attained the best scoring. As can be observed, for image 9.14(a), (b) and (c), C^{3}S always gets the largest number of best scorings, while for (d) is KLN and for (e) is C&M; in the case of image (f), C^{3}S and KLN obtain the same number of best scorings. In a global count through the lowermost row of the table, TOTAL, C^{3}S also exhibits the best behaviour.

In a second part of this experiment, a finer tuning of C^{3}S parameters was allowed to observe whether better quantitative or qualitative results could be obtained. The new parameter values, in case they led to a better segmentation, are indicated in table 9.10, as well as the performance measures resulting from this finer tuning; in case no better results were obtained, the table indicates the value of t leading to a larger number of best scorings. The best segmentations for C^{3}S, KLN, GEV and C&M are then shown in figures 9.17 and 9.18. As can be observed in table 9.10, in the case of figure 9.14(a), the finer

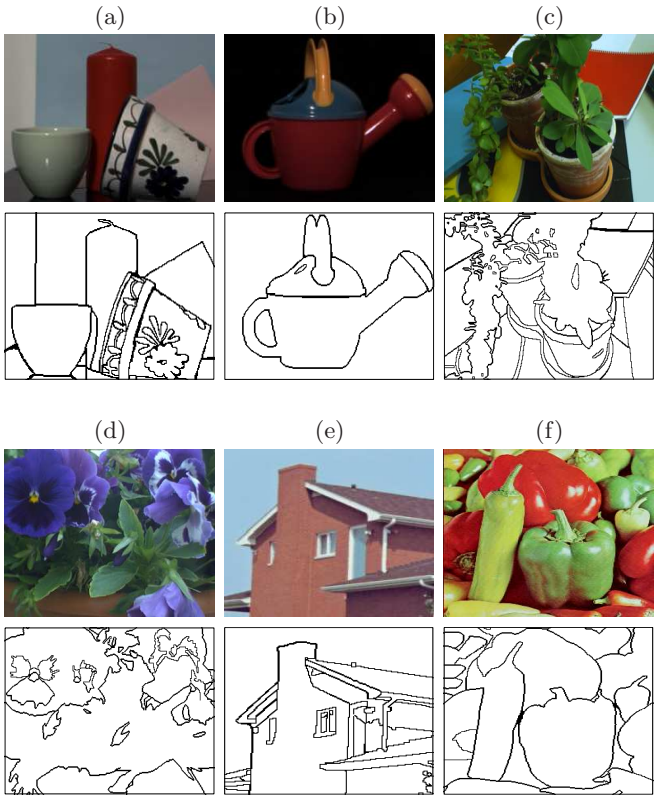


Fig. 9.14. Real test images and their manually-generated ground truths.

$t \rightarrow$	1	2	3	4	5	6
CG	96.62	95.84	64.49	28.60	38.61	50.28
OS	46.09	26.21	9.57	9.58	0.55	0.00
US	3.05	3.82	35.35	71.20	61.24	49.66
e_R^m	0.10	0.04	0.02	0.02	0.00	0.00
e_R^f	0.01	0.01	0.12	0.31	0.26	0.35
p_R	0.95	0.97	0.93	0.83	0.87	0.82
Baddeley	1.02	0.80	1.04	1.19	1.68	1.80

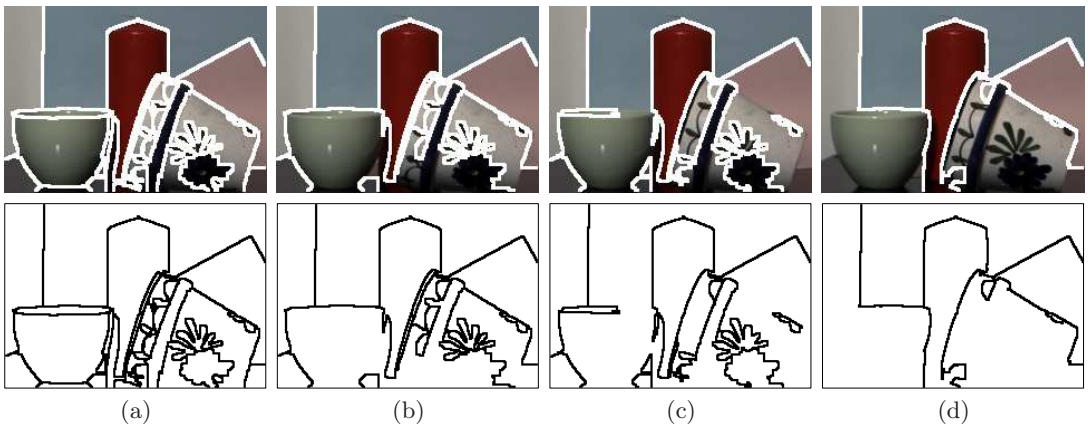


Fig. 9.15. Experimental results for C^3S and real image 9.14(a): [top] CG, OS, US, Huang and Dom, and Baddeley's measures for $t = 1..6$; [bottom] results for (a) t=2, (b) t=3, (c) t=4 and (d) t=5.

Table 9.6. Comparison between C^{3}S, KLN, GEV and C&M by means of real images: $\underline{\text{x}}$ means the best in the row; $\sum \#best$ is the sum of times that algorithm was best according to a certain measurement, for the corresponding image; and TOTAL is the number of times that algorithm attained a best scoring.

		C^{3}S			KLN	GEV	C&M
		$t=2$	$t=3$	$t=4$			
9.14(a)	CG	$\underline{95.84}$	64.49	28.60	70.45	42.91	25.60
	OS	26.21	$\underline{9.57}$	9.58	10.76	59.54	48.17
	US	$\underline{3.82}$	35.35	71.20	29.16	55.32	74.36
	e_R^m	0.04	$\underline{0.02}$	0.02	0.03	0.21	0.18
	e_R^f	$\underline{0.01}$	0.12	0.31	0.11	0.28	0.25
	p_R	$\underline{0.97}$	0.93	0.83	0.93	0.76	0.78
	Baddeley	$\underline{0.80}$	1.04	1.19	1.50	1.90	1.43
	$\sum \#best$	5	2	0	0	0	0
9.14(b)	CG	99.30	99.36	$\underline{99.55}$	98.91	98.11	92.50
	OS	6.67	$\underline{0.00}$	$\underline{0.00}$	17.42	$\underline{0.00}$	69.95
	US	0.03	$\underline{0.00}$	$\underline{0.00}$	0.56	$\underline{0.00}$	7.38
	e_R^m	0.01	0.01	$\underline{0.01}$	0.06	0.01	0.14
	e_R^f	0.01	0.01	$\underline{0.00}$	0.01	0.02	0.01
	p_R	0.99	0.99	$\underline{0.99}$	0.97	0.99	0.92
	Baddeley	0.56	0.50	$\underline{0.49}$	1.51	0.63	1.06
	$\sum \#best$	0	2	7	0	2	0
9.14(c)	CG	$\underline{89.78}$	69.91	79.29	43.39	62.54	74.50
	OS	57.00	56.77	$\underline{56.43}$	63.08	60.42	63.30
	US	$\underline{8.77}$	29.28	18.01	54.93	34.05	25.10
	e_R^m	0.44	0.33	$\underline{0.16}$	0.33	0.20	0.33
	e_R^f	$\underline{0.04}$	0.07	0.11	0.25	0.18	0.07
	p_R	0.76	0.80	$\underline{0.86}$	0.71	0.81	0.80
	Baddeley	1.81	1.56	$\underline{1.42}$	2.10	1.79	1.60
	$\sum \#best$	3	0	4	0	0	0
9.14(d)	CG	$\underline{97.34}$	96.53	96.61	95.91	85.73	94.29
	OS	90.54	88.03	$\underline{83.57}$	89.61	83.62	91.92
	US	1.08	1.73	1.09	$\underline{0.00}$	6.83	5.52
	e_R^m	0.76	0.66	0.64	$\underline{0.27}$	0.45	0.76
	e_R^f	0.02	0.02	0.02	0.04	0.09	$\underline{0.01}$
	p_R	0.61	0.66	0.67	$\underline{0.84}$	0.73	0.61
	Baddeley	2.38	1.84	$\underline{1.63}$	2.21	1.83	2.10
	$\sum \#best$	1	0	2	3	0	1
9.14(e)	CG	92.41	92.01	89.20	74.25	80.11	$\underline{96.44}$
	OS	16.62	9.57	$\underline{9.13}$	13.01	54.33	13.03
	US	6.63	7.03	9.90	25.52	14.38	$\underline{3.39}$
	e_R^m	0.07	0.06	$\underline{0.03}$	0.05	0.11	0.06
	e_R^f	0.02	0.02	0.04	0.19	0.09	$\underline{0.01}$
	p_R	0.95	0.96	0.96	0.88	0.90	$\underline{0.97}$
	Baddeley	0.94	0.77	$\underline{0.75}$	1.21	1.43	0.88
	$\sum \#best$	0	0	3	0	0	4
9.14(f)	CG	92.70	$\underline{93.13}$	88.66	27.12	10.77	56.53
	OS	96.48	94.45	93.08	$\underline{66.15}$	67.73	95.19
	US	6.72	$\underline{6.07}$	10.92	72.88	89.13	43.32
	e_R^m	0.50	0.49	0.33	$\underline{0.29}$	0.29	0.36
	e_R^f	$\underline{0.02}$	0.02	0.04	0.28	0.43	0.15
	p_R	0.74	0.74	$\underline{0.82}$	0.72	0.64	0.74
	Baddeley	2.34	2.00	$\underline{1.62}$	1.82	1.68	1.77
	$\sum \#best$	1	2	2	2	0	0
TOTAL		10	6	18	5	2	5

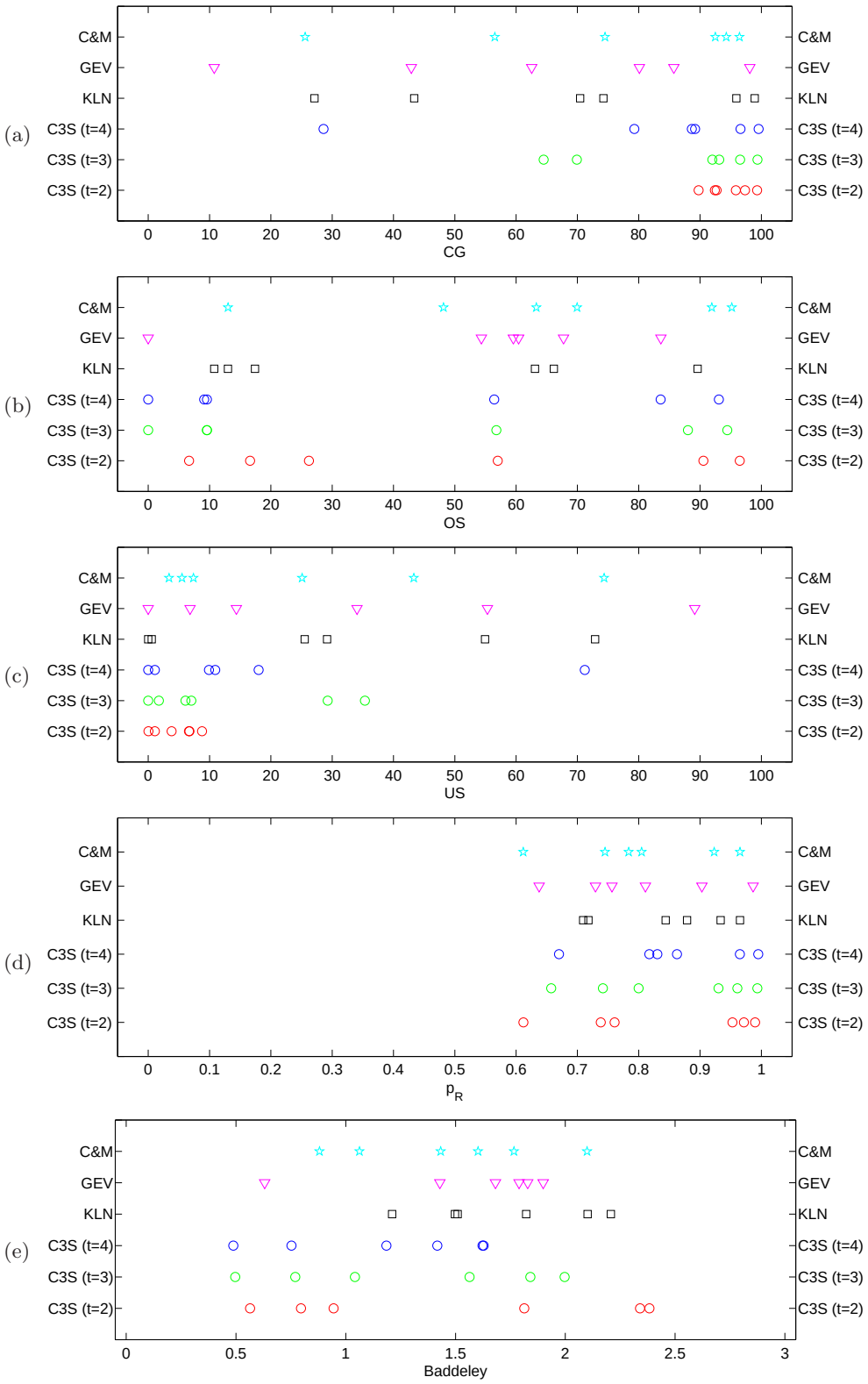


Fig. 9.16. Comparison between C^{3}S, KLN, GEV and C&M by means of real images: (a) CG, (b) OS, (c) US, (d) p_R and (e) Baddeley. (In every graph, every 1-dimensional axis corresponds to the measures for one algorithm.)

Table 9.7. Parameter values for KLN during the experiments with the real images of figure 9.14. (The description of the parameters corresponds to the one used in the original paper [129, table 6].)

parameter description	**image**					
	9.14(a)	9.14(b)	9.14(c)	9.14(d)	9.14(e)	9.14(f)
window size (for initial grouping)	10×10					
minimal area size	100	100	100	500	100	500
camera noise	3	5	5	5	3	8
cylinder width	6	10	10	10	8	20
minimal intensity	20					

Table 9.8. Parameter values for GEV during the experiments with the real images of figure 9.14.

parameter description	**image**					
	9.14(a)	9.14(b)	9.14(c)	9.14(d)	9.14(e)	9.14(f)
standard deviation for gaussian derivative operators	1.0					
maximum standard deviation for homogeneous regions (normalized between 0 and 255)	5	15	10	10	5	15
Tlow for hysteresis thresholding	0.4					
Thigh for hysteresis thresholding	0.6	0.75	0.8	0.8	0.75	0.7

Table 9.9. Estimates of directional illumination strength, L_d^c, for images of figure 9.14.

image	9.14(a)	9.14(b)	9.14(c)	9.14(d)	9.14(e)	9.14(f)
(L_d^R, L_d^G, L_d^B)	(193.00,187.50,169.50)	(196.25,160.00,134.75)	not available (200.00,200.00,200.00) used instead			

Table 9.10. Parameters and performance measures for C^{3}S best segmentations of images of figure 9.14.

image	parameters	CG	OS	US	e_R^m	e_R^f	p_R	Baddeley
9.14(a)	$t_1 = 3$, $t_2 = 2$, $t_3 = 3$	93.04	25.56	5.56	0.04	0.02	0.97	0.87
9.14(b)	$t = 4$	see table 9.6						
9.14(c)	$t = 4$	see table 9.6						
9.14(d)	$t = 4$	see table 9.6						
9.14(e)	$t = 3$	see table 9.6						
9.14(f)	$t = 4$, $\tau_{sdim} = 3.0$ $\tau_{split} = 100.0$	52.06	70.16	47.59	0.26	0.09	0.83	1.09

tuning improves OS at the expense of worsening CG and Baddeley, what leads to a visually apparent better segmentation. In the case of figure 9.14(f), it was necessary to increase τ_{sdim}, to tolerate the larger width of specularities of this image, as well as τ_{split}, because of the non-linearity of matte clusters of this standard image (this was not necessary, however, for the other standard image of figure 9.14(e), probably because of the almost non-existent scene curvature). As a consequence, this finer tuning reduced the level of over-segmentation and improved the combined measure by Huang and Dom, p_R, as well as the localization of reflectance transitions.

In a more qualitative level, figures 9.17 and 9.18 show that C^{3}S is able to produce sound segmentations for the varied set of images considered in the test bench, what means that, although some performance measures does not situate C^{3}S at the optimum side on some occasions, the visual appearance of the result is quite reasonable. Also observe that C^{3}S produces best results mostly for $t = 3$ and $t = 4$. It is also noteworthy that GEV exhibits serious problems to produce acceptable segmentations for the images of figure 9.14 except for image 9.14(b). The different trials performed for this algorithm

throughout this experiment have shown that this is due to the use of the Delaunay triangulation to steer the splitting stage of the algorithm, but, more importantly, because of the global threshold used to control the level of regions homogeneity, what prevents the merging stage from leading to low levels of over- and under-segmentation[2]. As for KLN, a certain amount of spurious regions tend to appear in every image; this is, however, not very serious since they could be easily removed by means of the region usefulness criteria commented on chapter 4. Besides, KLN seems to be in trouble with the specularities of image 9.14(b); the glossiness of image 9.14(f) is neither well handled by KLN, but this is a standard image for which there is not any guarantee of corresponding to the model of image formation assumed by this algorithm. Finally, C&M behaves as expected with this type of images: where scene curvature becomes noticeable, it tends to produce over-segmentation (figures 9.14(a), (b) and (f)); in other cases, such as, for instance, the standard *house* image (figure 9.14(e)) an excellent behaviour is exhibited.

To finish this section of experimental results with real images, figures 9.20, 9.21, 9.22, and 9.23 show the segmentations produced by, respectively, C^3S, KLN, GEV and C&M for the set of real images shown in figure 9.19, which was also used to test C^3E in section 8.9.2. They are reproduced here again to make the comparison easier. Remember that no ground truth data are available for the images of this set, so that the discussion will also this time be performed at the qualitative level.

As can be observed from these figures, the behaviour already exhibited by the different algorithms persists in this experiment. Besides, C^3S exhibits a nice performance with the textured cup of image (d) of figure 9.19, as well as for the standard images *lena* and *sailboat* (respectively, images 9.19(g) and (h)). A shortcoming of C^3S can, however, be noticed in the image (f) of figure 9.19, where C^3S is not able to deal with small image details, such as the chain in the upper left-hand corner of the image, what is clearly overcome by C&M. KLN seems to have problems with the white wall of image (f) of figure 9.19, as well as with the specularities of image (b); the textured cup, however, is also relatively well handled by this algorithm. This time GEV also deserves special attention, since it has been necessary to resort to the RGB colour space to obtain relatively reasonable segmentations for many of the images of this set; as can be observed in figure 9.22, the results obtained from GEV when working in the $C_1C_2C_3$ space were almost useless on some occasions[3].

9.5 Conclusions

A curvature-insensitive segmentation algorithm called C^3S (*Colour Channels Coupling-based Segmentation*) has been proposed. Although it is also based on the Dichromatic Reflection Model, like many other physics-based segmentation algorithms, C^3S is radically different to other approaches also making use of it. C^3S consists of three main stages: the first one scans the image looking for reflectance transitions, a task that is performed by the already introduced C^3E edge detection algorithm; the second stage takes as a basis the non-thinned edge map produced by C^3E to make regions grow until a first segmentation of the image is achieved; and the third stage refines this former segmentation by merging those regions whose common border does not make sense from a perceptual point of view. C^3S is, therefore, based on colour channel coupling analysis, which is the fact that makes it different from other physics-based segmentation approaches. As an added value, typical problems of the classical approaches are avoided:

[2] The author provided experimental results about the adaptability of the Delaunay triangulation to the geometry of the scene, but the experiment only included synthetic images consisting of objects with mostly polygonal contours [68]. On the other hand, successful results for a real image of plastic toys were also provided, which coincides with the good segmentation produced for image 9.14(b). Therefore, it is not surprising the behaviour observed for GEV throughout this second experiment, given the difficulty of the images considered against the ones appearing in the paper by Gevers.

[3] Note that when GEV works in the RGB space leaves its condition as a physics-based algorithm, since in this space it does not take into account the physics of image formation to tolerate optical effects such as the shading. It is through the $C_1C_2C_3$ and $L_1L_2L_3$ colour spaces proposed by the author how these effects are counteracted.

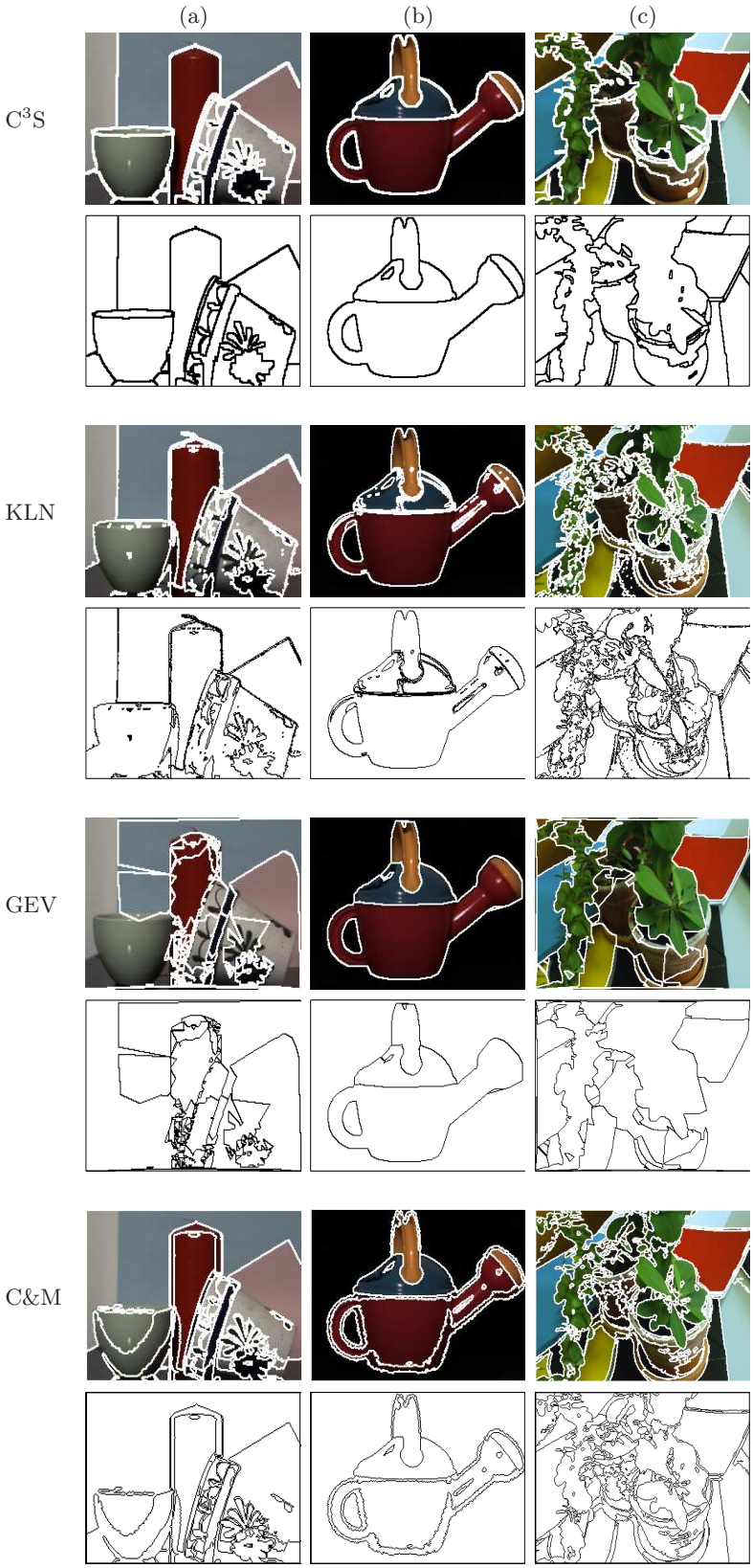


Fig. 9.17. Best segmentation results for C^3S, KLN, GEV and C&M for images in figure 9.14(a-c): the first row of every algorithm presents the original images with region contours superimposed, while the second row shows region contours.

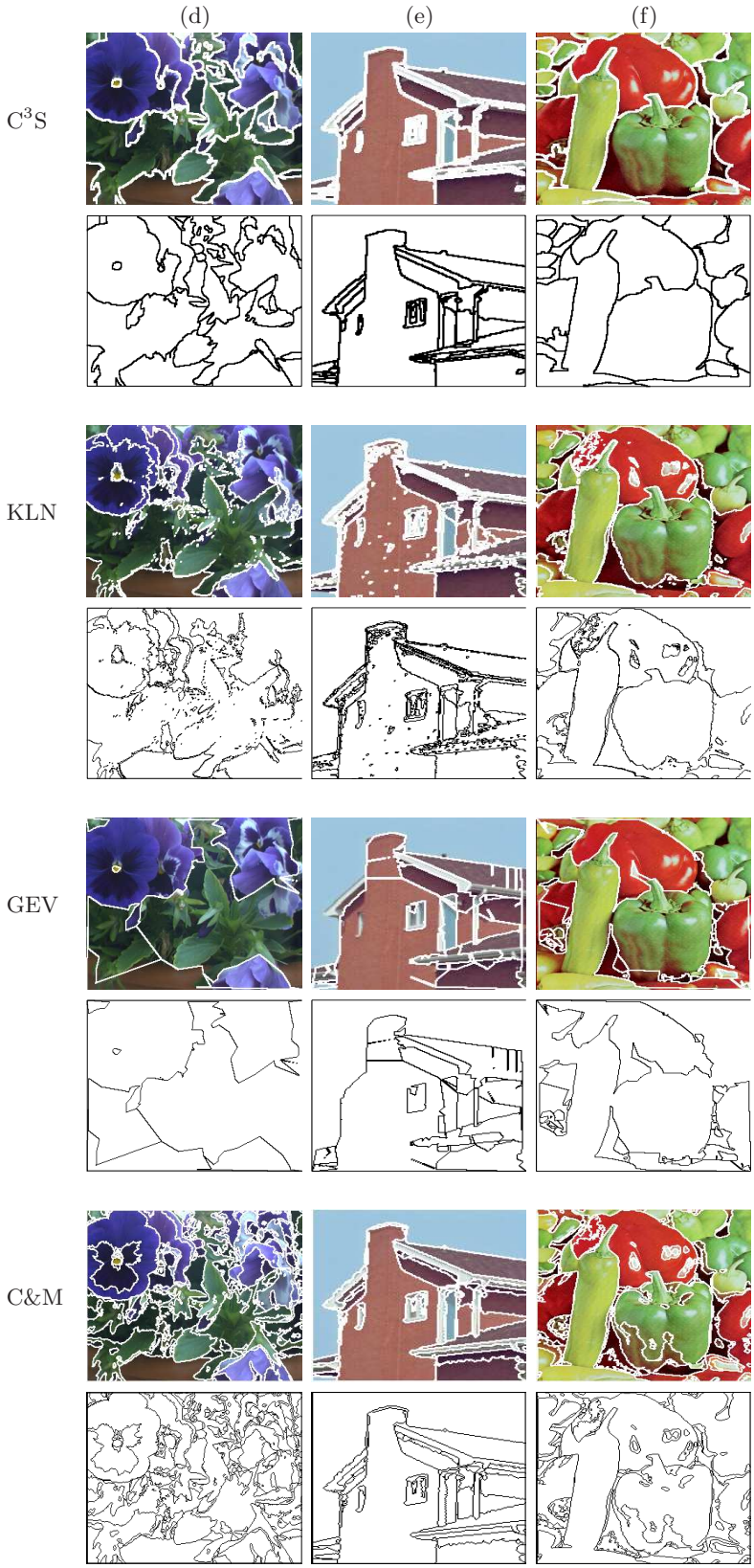


Fig. 9.18. Best segmentation resuls for C^3S, KLN, GEV and C&M for images in figure 9.14(d-f): the first row of every algorithm presents the original images with region contours superimposed, while the second row shows region contours.

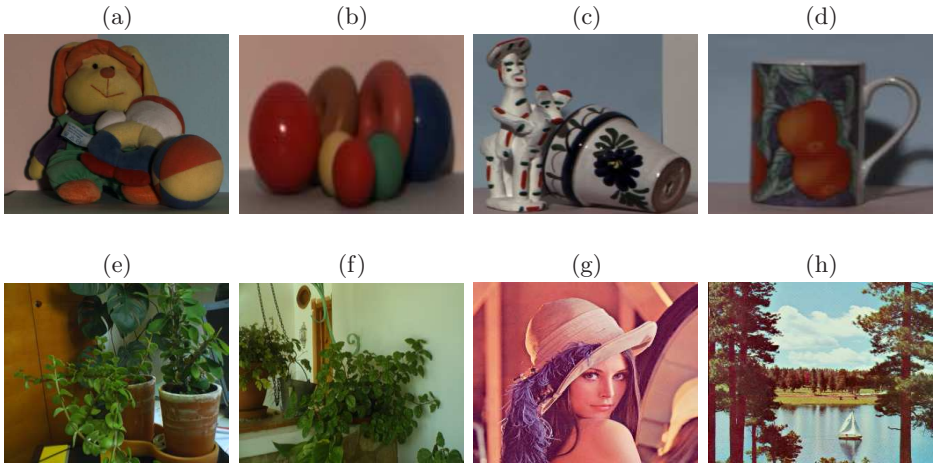


Fig. 9.19. Real images used in the third experiment to compare the performance of C^3S, KLN, GEV and C&M.

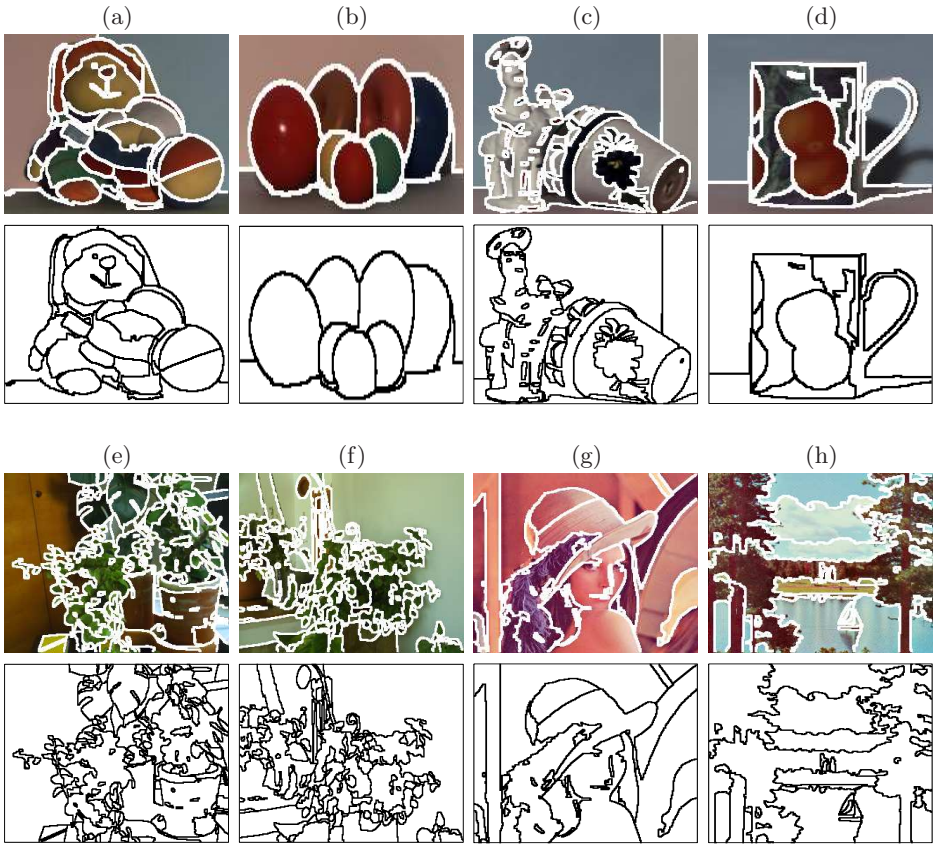


Fig. 9.20. Best segmentation results for C^3S corresponding to the third experiment with real images: the first and third rows are for the original images with region contours superimposed, while the second and fourth rows are for the region contours only.

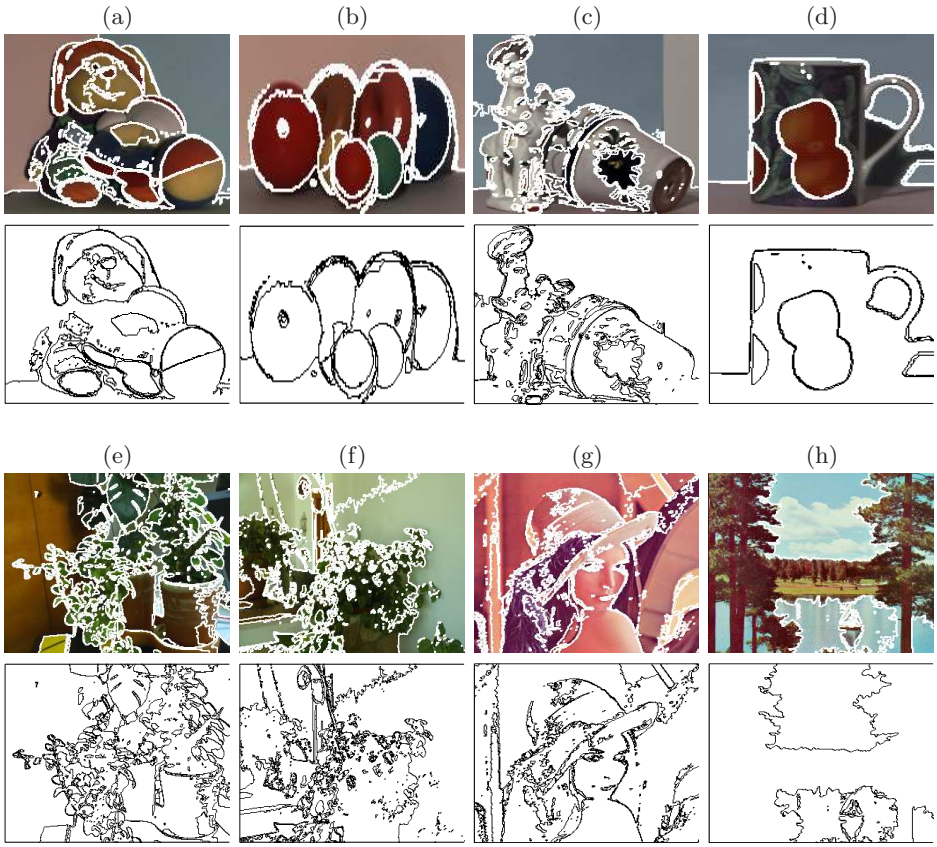


Fig. 9.21. Best segmentation results for KLN corresponding to the third experiment with real images: the first and third rows are for the original images with region contours superimposed, while the second and fourth rows are for the region contours only.

- there is no need to numerically estimate scene material reflectances, which becomes quite difficult if not impossible as soon as the image formation model incorporates complex optical phenomena such as specularities or inter-reflections;
- pixel clustering in colour space is not needed, so that deviations of the cluster shapes due to image noise do not influence the segmentation process; and
- numerical instabilities of photometric invariants due to their formulation or due to noise are avoided.

Experimental results for noisy synthetic and real images corresponding to a varied set of scenes have been provided and discussed. C^3S has been compared with two well-known physics-based segmentation algorithms, one by Klinker et al. [129] (KLN) and the other one by Gevers [68] (GEV). The non-physics-based segmentation algorithm by Comaniciu and Meer [29] (C&M) has also been incorporated during the comparison with real images. Several conclusions can be drawn from the results of the different experiments performed:

- The scene constraints concerning the fulfillment of the coupling properties were not so hard as seemed at first sight, what could already be observed in the experimental results of C^3E.
- The edge maps produced by C^3E have resulted to be useful to obtain adequate segmentations of images. In fact, C^3S has proved to be able to deal with scenes with glossy curved objects and different surface materials.
- Despite the *a priori* large number of parameters of C^3S, quite acceptable results have been obtained reducing them to just the number of intensity uncertainties t used to counteract image noise. It has

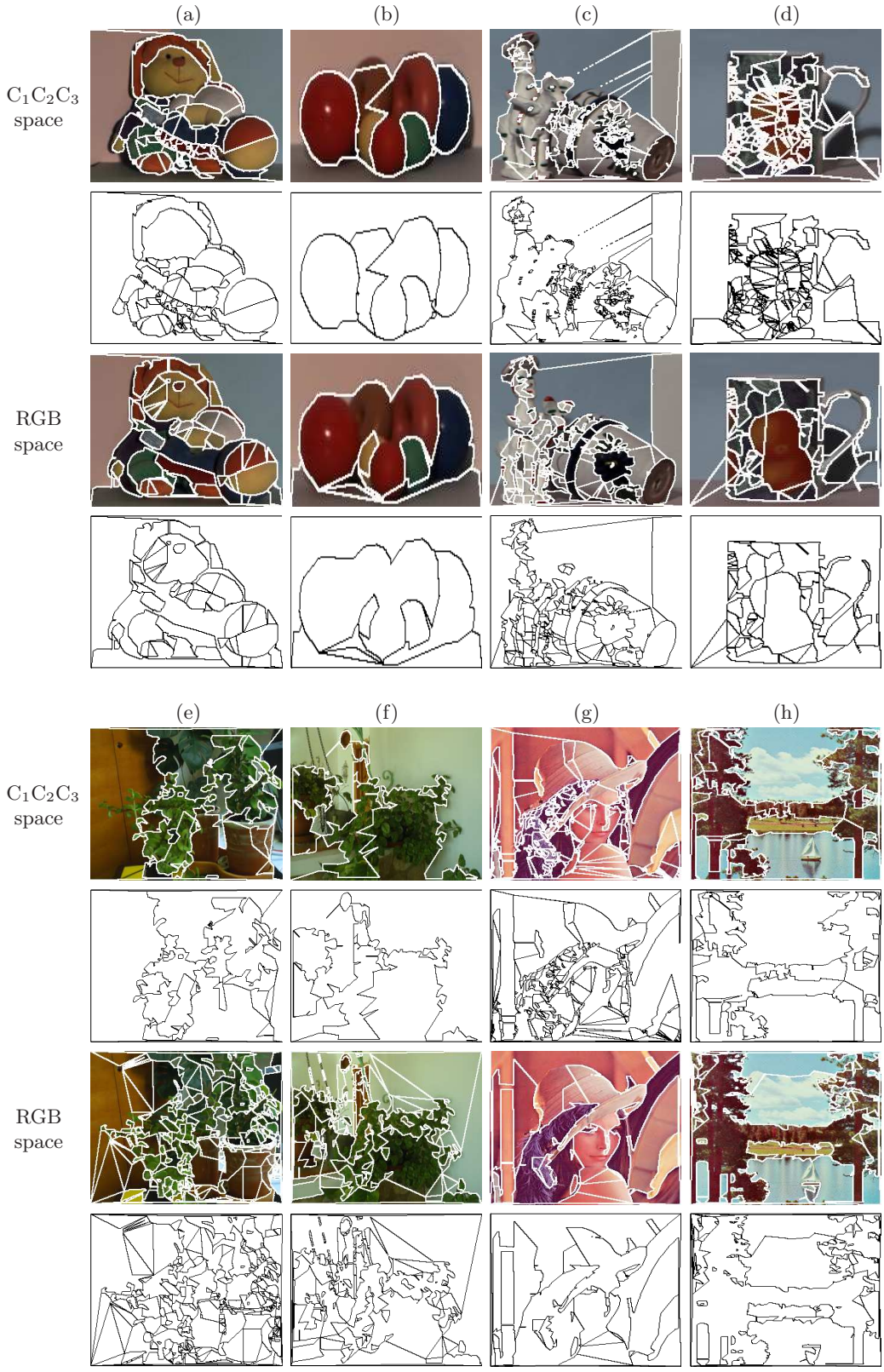


Fig. 9.22. Best segmentation results for GEV corresponding to the third experiment with real images: 1st, 3rd, 5th and 7th rows are for the original images with region contours superimposed, while 2nd, 4th, 6th and 8th rows are for the region contours only.

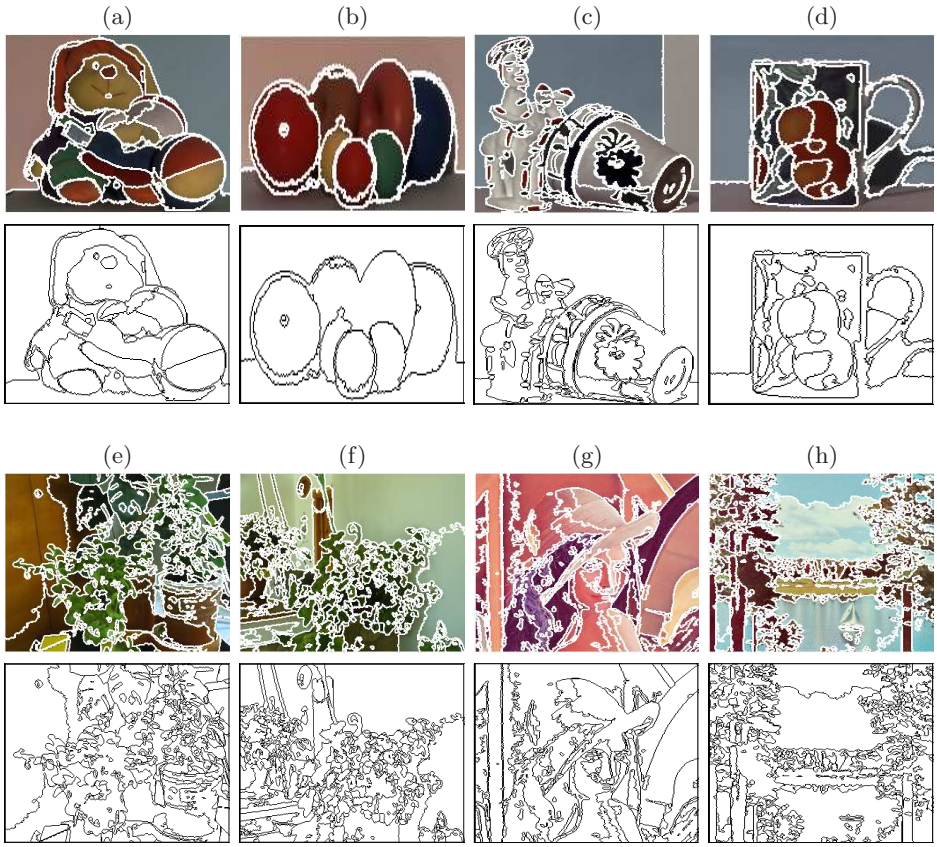


Fig. 9.23. Best segmentation results for C&M corresponding to the third experiment with real images: the first and third rows are for the original images with region contours superimposed, while the second and fourth rows are for the region contours only.

been observed that for most images $t = 3$ or $t = 4$ produced sound results. A finer tuning, however, can improve the quality of the segmentation in some cases.

- Under normal noise conditions, C^3S has shown a better performance than the other algorithms for synthetic images. When processing real images, the differences are narrower, but still C^3S produces the best results on most occasions, independently of the type of scene.
- The tolerance to noise of C^3S has found to be dependent on the number of uncertainties. In particular, referring to the classification aspect of segmentation, C^3S with $t = 2$ has shown to be the algorithm with the largest tolerance, but for larger values of t KLN has resulted to be better than C^3S. On the other hand, regarding the localization of reflectance transitions, C^3S has performed better than the other segmentation algorithms irrespective of t. GEV has shown the worst performance, but, at the same time, this keeps almost the same as the noise contaminating the image is increased.
- Intensity uncertainties have proved to be useful in another computer vision task.
- The evaluation of segmentation performance by means of CG, OS and US has resulted satisfactory, although Huang and Dom performance measures e_R^m, e_R^f and p_R have found to be useful as secondary measures in some experiments.

On the other hand, some segmentation mistakes have been detected when scenes contained small details. On those cases, C^3S tends to join those small details with the region surrounding them. This behaviour stems from the edge-finding nature of C^3S because of the thickness of edges in the edge map. In this sense, a pixel-based segmentation approach is more likely to succeed. Therefore, an approach

combining the advantages of C^3S and a pixel-based segmentation strategy could constitute a good source of improvement in a future work.

As for the executions times, both KLN and C&M have needed on the order of 1-2 minutes to produce a segmentation result, while in the case of C^3S and GEV the execution time has oscillated between less than a minute and up to 15 minutes depending on the image (all the executions were performed over a Pentium IV @ 3GHz machine). This is because both algorithms include iterative processes which must converge to yield a segmentation. In the case of GEV, the splitting stage implies considering several triangulation scenarios for every edge point in order to take the triangulation best describing the scene, so that the larger the number of edges, the larger the execution time. In the case of C^3S, although the computation of the edge map reduces to very simple operations between a pixel and its neighbourhood and therefore does mean a very reduced execution time —around one second on most images, as was already commented on in chapter 8—, the growing and merging stages, above all the latter, involve a large amount of more or less complex operations. In particular, the merging stage computation time depends on the number of evaluations of ζ_{pcm} between pairs of regions. The worst case happens when the number of final regions is low, but the point of departure, produced by the region growing stage, consists of an important amount of regions. This is the typical case for images of natural scenes. A supporting hardware could, probably, alleviate the problem.

10

Conclusions and Future Work

This chapter finishes the main body of this documentation by presenting and discussing a set of conclusions which the work performed have led to. Further work and lines of research are also advanced.

10.1 Conclusions

This section gathers and summarizes the different sets of conclusions which have been drawn from the different contributions of this thesis and which have been provided at the end of the corresponding chapters of this documentation. By way of first summary of the work performed, figure 10.1 replicates the graph of figure 4.7 labelling each component with the name and chapter of the related contribution of this thesis.

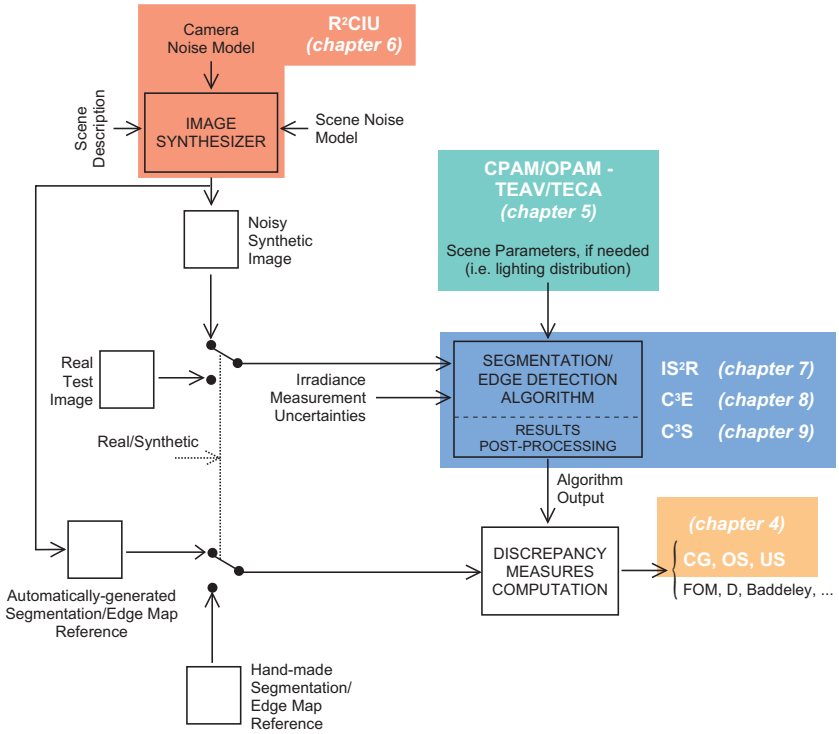


Fig. 10.1. Diagram of the performance evaluation setup proposed in chapter 4 labelled with the contributions of this thesis.

Table 10.1. Incorporation of IS^2R, C^3E and C^3S to the summary of physics-based segmentation approaches of table 3.1.

alg	GL	CO	A	B	I	S	IR	WI	NIR	IWC	out	GA	PB	more ...
IS^2R	✓	✓	✓	✓		✓					IS	region	reflectance	use of intensity uncertainties; set of images limited
C^3E	✓	✓	✓	✓							EM	edge	coupling	use of intensity uncertainties; NIR and proportional ambient illumination, but not essential
C^3S	✓	✓	✓	✓	✓						IS	region	coupling	use of intensity uncertainties; NIR and proportional ambient illumination, but not essential

In the following, conclusions which can be drawn from a global point of view are distinguished from particular conclusions for every contribution of this thesis, which, in turn, are put in sections apart. Besides, the particular conclusions are, to a certain extent, given in the form of summaries of the ones given at the end of the different chapters, with the aim of emphasizing the most relevant facts.

10.1.1 Global conclusions

From a global point of view, this thesis has achieved the following objectives:

(1) The segmentation problem has been studied from within the physics-based vision paradigm, contributing with a quite exhaustive review of image segmentation and edge detection approaches agreeing with the same paradigm. The main strategies have also been identified and used to develop a comprehensive taxonomy of physics-based segmentation approaches. As a result of the analysis, several shortcomings of the so far proposed algorithms have been enumerated, which, on the other hand, have confirmed the validity and the coherence of the goals of this thesis.

(2) Several approaches have been explored and, as a consequence, three algorithms named IS^2R, C^3E and C^3S have resulted, whose main features can be found summarized in table 10.1 following the same structure as the table introduced in chapter 3 to highlight the characteristics of the algorithms considered as the state of the art. In the table, C^3E and C^3S have been considered to be able to handle also gray-level images, although no results have been given, because, from a theoretical level, their capabilities in this respect have not been cut off. On the other hand, a new category called *coupling* has been created for both algorithms since none of the previously introduced classes are fully appropriate because:
   - no reflectance is estimated,
   - there is not an analysis of clusters in any colour space,
   - they are not based on studying the continuity of the image surface, and, finally,
   - no invariants are used.

(3) All three algorithms:
   - consider ambient illumination since their theoretical inception;
   - ensure the full coverage of reflectance transitions, at least from a theoretical point of view;
   - work at the signal level, except for the assumption of smooth surfaces; and
   - adapt the respective thresholds on the basis of intensity uncertainties.

(4) The different proposals have been tested from different points of view:
   - using synthetic images with and without noise to test, respectively, the theoretical correctness of the algorithm and their degree of robustness;

- using real images of a varied set of scenes and different surface materials to confirm their validity in less controlled conditions; and
- comparing them with other well-known image segmentation and edge detection algorithms.

Furthermore, it is worth noting the fact that the real images used in the experiments are more complex than the ones that have traditionally been used by researchers in physics-based segmentation. In particular, images of natural scenes have been incorporated into the sets of test images.

During the experiments, C^3E and C^3S have compared favourably with three physics-based and two non-physics-based recognized methods belonging to the current state of the art. As for IS^2R, it has become patent its theoretical correctness but also a limited scope because of the approach followed, which implies, among others, a perfect accommodation to the model of image formation assumed. Nevertheless, the experimental results showed it is able to deal with images which which other algorithms have serious problems. Finally, IS^2R presents the added value of also providing shape information apart from the expected segmentation results.

(5) Apart from the image segmentation and edge detection algorithms, this thesis has contributed with several methods for determining the lighting parameters of the scene (OPAM/CPAM - TEAV/TECA) and an algorithm for estimating the parameters of a noise model related with the operation of a CCD camera (R^2CIU). On the one hand, OPAM/CPAM-TEAV/TECA have proved useful for the segmentation algorithms making use of it: IS^2R, which needed both the direction of the light source and the chromatic information of the ambient and directional sources of light, and C^3S, which just needed the chromatic values of the directional light source for the detection of specularities. On the other hand, R^2CIU has resulted useful not only for estimating the parameters of a noise model related with the operation of a CCD camera, but also for computing intensity uncertainties to be incorporated in the different segmentation and edge detection algorithms proposed.

10.1.2 Conclusions for chapter 4

Chapter 4 proposed a framework for image segmentation and edge detection, together with an experimental setup aiming at optimizing the conditions of operation for physics-based segmentation algorithms. At the end of this thesis, this setup can be said to have satisfied the expectations put on it, since it has allowed to successfully develop and test physics-based algorithms, what is proved by the agreement between the experiments with synthetic images and the experiments with real images.

Besides, this chapter proposed three measures of performance for image segmentation algorithms called CG, OS and US. Throughout the different experiments in which they have been used, they have proved to behave satisfactorily, although, in some experiments, their combination with the Huang and Dom performance measures e_R^m, e_R^f and p_R has also found to be useful.

10.1.3 Conclusions for chapter 5

A total of four methods for estimating scene lighting parameters have been presented throughout chapter 5. Two of them assume orthographic projection (OPAM) while the other two assume perspective projection (CPAM). Within every projection case, two algorithms for estimating the lighting orientation have been discussed and compared, TEAV and TECA, which differ in the way how the tilt of the illumination is determined. In all cases, the corresponding estimation method has been developed in connection with the camera noise model presented in section 2.4.4, which has allowed identifying the noise sources corrupting digital pixel values in order to counteract their effects throughout the estimation process.

An extensive battery of tests has been carried out in order to characterize as accurately as possible the scope of application of the lighting parameters estimation methods presented. More precisely: (1) all the variants have been tested against noiseless synthetic images under different geometrical conditions; (2) several robustness tests have been conducted, which included images with Gaussian noise, changes in the lighting strengths maintaining the lighting direction constant and changes in the position of the calibration object keeping constant the lighting strength; and, finally, (3) all the methods have been

compared with other well-known estimators with similar features and scope. As a general conclusion, the combination CPAM-TEAV has proved to be the best election, although the other variants provided estimations which could also be considered quite accurate. The combinations with OPAM require to respect the fact that orthographic projection are assumed and, therefore, the closer the calibration sphere to the optical axis, the better.

Despite the limitations of the lighting estimation methods proposed in chapter 5 as for the number of light sources, several advantages make them attractive in single-source cases against multi-source methods: (1) in general, its complexity is lower than the multi-source methods, (2) a previous geometric calibration of the camera is not required and the estimation of the surface normal vectors of the calibration sphere is embedded within the method itself, (3) all the image pixels belonging to the sphere projection are used in the estimation of the lighting parameters, and (4) no threshold or parameter needs to be set up.

Finally, the combination of CPAM-TEAV with a segmentation algorithm has produced satisfactory results. In particular, it is noteworthy the behaviour of the lighting estimation method in reference to the estimates produced in chapter 7, where very consistent estimations resulted when analyzing the same conditions of illumination over gray-level and colour images (figures 7.12 and 7.16, and figures 7.18 and 7.21).

10.1.4 Conclusions for chapter 6

Chapter 6 proposes an algorithm called R^2CIU (*Robust Radiometric Calibration and Intensity Uncertainty estimation*) for radiometric camera calibration. It allows estimating the distribution parameters of the noise sources of the camera noise model proposed by Healey and Kondepudy in [93]. Besides, two applications of the radiometric calibration of CCD cameras have also been discussed, namely synthesis of noisy images for testing purposes, and estimation of intensity uncertainties, both according to the radiometric performance of a particular camera, or, in other words, a particular set of camera noise model parameters. Both have proved to be useful within the image segmentation and edge detection framework proposed in chapter 4.

More precisely, R^2CIU is based on uniform reflectance calibration cards. Besides, the fact that all the pixels of the captured image can participate in the estimation makes R^2CIU a robust estimator of the noise model parameters. It has been compared against itself using two sets of calibration images taken at different time instants with slight deviations in the parameters estimated for both sets. It also compares favourably with the algorithm of Healey and Kondepudy [93], showing a similar behaviour but a simpler use.

10.1.5 Conclusions for chapter 7

This chapter has presented a segmentation algorithm called IS^2R (*Image Segmentation by Scene Reconstruction*) insensitive to scene curvature. To this end, reflectance values of the scene objects are estimated by means of singular points and the scene shape is progressively reconstructed in order to predict intensity values to be compared with real intensity values in a so-called consistency test, what allows making a decision about whether a pixel must join a certain region or not. Experimental results for both synthetic and real images have been provided.

During the experiments performed, both colour and gray-level images have been considered and no difference in performance has been noticed. On the other hand, algorithm parameters have resulted to be quite stable, except those related with spatial dimensions, specially for the determination of singular points. In this case, some fine tuning has been needed.

Although the experimental results prove IS^2R is able to deal with scenes with curved objects, its general applicability has been compromised by the limited sort of images which can correctly be segmented. In particular, IS^2R is not very tolerant to deviations of the theoretical image formation model, namely specularities, reflectance non-uniformities and small details. One of the clear points for this limitation is the SFS algorithm which is embedded in IS^2R for estimating shape information. Nevertheless, the very basis of IS^2R is also the cause for its low performance in certain cases, since the

use of the prediction of the intensity values means a strong agreement must exist between images and the image formation model, what not always takes place.

10.1.6 Conclusions for chapter 8

A study about what happens at the signal level when reflectance changes take place has been presented in chapter 8. As it has been shown, colour channels keep coupled in uniform reflectance areas, while the coupling is broken in a number of ways at pixels involved in reflectance transitions. This coupling has been expressed in the form of a set of properties which are fulfilled by all the pixels corresponding to the same surface material (i.e. same reflectance), so that, when any of them is not satisfied, the only possible cause of the unfulfillment reduces to a change in reflectance. The satisfaction of those properties has been studied for different instantiations of the image formation model, which have allowed determining the constraints to be met by the scene so that these properties can be effectively used to locate reflectance transitions. The coupling properties have in turn been used to derive a number of types of edges which can be easily detected by analyzing the way how the colour channels are related among them. This study has finally derived in the development of, on the one hand, an edge detector, and, on the other hand, an image segmentation algorithm, both of which do not exhibit the typical drawbacks of some of the approaches surveyed in chapter 3:

- there is no need to numerically estimate scene material reflectances, what becomes quite difficult if not impossible as soon as the image formation model incorporates complex optical phenomena such as specularities or inter-reflections;
- pixel clustering in colour space is not needed, so that deviations of the cluster shapes due to image noise do not influence the segmentation process; and
- numerical instabilities of photometric invariants due to their formulation or due to noise are avoided.

Chapter 8 has also proposed the edge detector named C^3E (*Colour Channel Coupling-based Edge detection*) rooted on the results of the colour channel coupling study. Experiments with synthetic and real images have been presented, showing the power of the approach for dealing with scenes with curved objects and different surface materials. However, in general, it is not tolerant to highlights. This means that, from a theoretical point of view, it is not possible to avoid specularity edges in all cases, although it is clear that they appear as a result of the colour channel coupling analysis only under some circumstances.

Besides, C^3E has been compared in a number of tests with the physics-based edge detection algorithm by Stokman and Gevers [79, 262] and with the recognized non-physics-based edge detection algorithm by Meer and Georgescu [168]. From a global point of view, C^3E can be said to exhibit a good tolerance to noise thanks to the use of intensity uncertainties for threshold selection and it has also outperformed the other two algorithms involved in the comparison.

10.1.7 Conclusions for chapter 9

As a natural continuation of chapter 8, chapter 9 has proposed a curvature-insensitive segmentation algorithm called C^3S (*Colour Channels Coupling-based Segmentation*). Although it is also based on the Dichromatic Reflection Model, like many other physics-based segmentation algorithms, C^3S is radically different to other approaches also making use of it because of its roots on colour channel coupling analysis. C^3S consists of three main stages: the first one scans the image looking for reflectance transitions, what is performed by the already introduced C^3E edge detection algorithm; the second stage takes as a basis the non-thinned edge map produced by C^3E to make regions grow until a first segmentation of the image is achieved; and the third stage refines this former segmentation by merging those regions whose common border does not make sense from a perceptual point of view.

Experimental results for noisy synthetic and real images corresponding to a varied set of scenes have been provided and discussed. C^3S has been compared with two well-known physics-based segmentation algorithms, one by Klinker et al. [129] and the other one by Gevers [68]. The non-physics-based

segmentation algorithm by Comaniciu and meer [29] has also been incorporated during the comparison with real images. Among the several conclusions which can be drawn from the results of the different experiments performed, it is worth noting the general better performance of C^3S with regard to the other algorithms when processing both noisy synthetic images and real images. On the other hand, some segmentation mistakes have been detected when scenes contained small details. On those cases, C^3S tends to join those small details with the region surrounding them.

10.2 Future Work

This thesis has focused on the image segmentation problem and, after the consideration of an ample set of questions, have resulted in several contributions, some of them directly related and some others indirectly related with new segmentation approaches. During both the time of development and the experiments, a number of hints for improvement and newer ideas have arisen. While the former refer to a set of misbehaviours which have been detected in the algorithms, the latter correspond to new strategies which, at least at the moment of inception, intuition said they could lead to better performance or to extend the scope of application. Both kinds of suggestions are invitations to follow this research line in the future. They can be found summarized in the following:

Regarding the lighting estimation methods CPAM/OPAM-TEAV/TECA ...

- The most relevant drawback of these methods is that they assume a model of illumination consisting of an only directional light source. On some occasions, this model is enough because, taking as an example the particular case of this thesis, the segmentation algorithm making use of the estimations assumes the same sort of illumination. However, a model with a larger number of directional light sources would allow extending the scope of application of the segmentation algorithm, in case it could effectively benefit from the enhanced model. In the case of this thesis, this would not be specially difficult provided that the shape estimation stage of IS^2R was able to cope with the new model. In this sense, the incorporation of, for instance, the SFS method of [279] should be considered.

Regarding R^2CIU ...

- The imperfections of the camera optics are not taken into account in R^2CIU. However, during the estimation of the noise parameter values, a special care has been taken to avoid the effects of such imperfections by reducing the population of image cells involved in the process to a central window of 100×100 pixels and by properly focusing the camera. Therefore, although computing the estimates from about 10,000 cells is statistically meaningful enough, it would be interesting to be able to include all the cells of the CCD in the computation of the noise parameters. Furthermore, the uncertainties estimated for every intensity level of every colour channel do not include the alterations coming from the optics. However, it has to be proved whether a new version of R^2CIU incorporating an enhanced noise model including the imperfections of the optics would significantly improve the quality of the uncertainties estimated.
- Due to their low cost and enhanced functionality, optical CMOS imaging sensors are being incorporated in every time more vision cameras. Although there is a certain overlapping between CMOS- and CCD-based imaging devices, the operation of the new sensor should be carefully studied and an adapted noise model suggested, together with, perhaps, new methods for estimating the parameters of the resulting noise models.

Regarding C^3S and C^3E ...

- As has been mentioned several times, C^3S and C^3E show a tendency to miss some small details in the image. This is because of the thickness of the edges resulting from the analysis of the coupling between colour channels. The generation of edges of such thickness is coherent with the fact that, because of camera aliasing, the reflectance transition spans along several pixels. Nevertheless, though taking into account that those pixels are not valid in the sense that violate at least one of the coupling properties, a refinement step for locating precisely the point where

the reflectance change takes place would probably reduce the incidence of thick edges in the posterior segmentation and edge detection results.

- LOG zero-crossings are detected in C^3S and C^3E working on every colour channel independently. This fact leads to the detection of the same zero-crossing at adjacent pixels in the same scanline because of noise, what contributes to the edge thickness in the edge map. Di Zenzo suggested in [37] a colour gradient which could be incorporated in C^3E to alleviate this problem. Nevertheless, a careful analysis should follow in order to be sure no edges are missed with regard to the previous approach.
- Although the computation times of the edge map for C^3E are quite reduced, the ones for C^3S can be quite large for some images. As was already indicated in chapter 9, this was mostly due to the application of the PCM algorithm during the merging stage. Therefore, the optimization of this algorithm or its replacement by another one with similar performance would be an interesting improvement.
- The simplicity in the calculations needed to generate the edge maps of C^3E suggests the idea of implementing in hardware this part of C^3S. This would allow obtaining edge maps for real-time applications and also would reduce the execution time of C^3S.
- The best way to extend the scope of application of a physics-based segmentation algorithm is to extend the model of image formation. Clearly, although both C^3S and C^3E have shown their power over a varied set of scenes, there is still more place for enhancements. On the one hand, the models considered in this thesis have all the time assumed non-attenuating propagation media, while, for instance, in an underwater environment, this is of little use. On the other hand, surfaces have been considered opaque and non-emissive, so that effects such as transparency, incandescence, fluoroscence or fosforescence have not been taken into account. Finally, reflectance has been required to be uniform for a certainly large amount of pixels, which would be violated in scenes with textured objects. All these modifications of the working conditions would certainly mean reconsider the approaches followed in C^3S and C^3E, specially the tolerance to textured objects.
- Several researchers have shown the power of integrating region and boundary information to compute the segmentation of an image [53, 172, 173]. In a certain way, this integration is also done by C^3S in the sense that the point of departure for the segmentation is an edge map, and, once it has been computed, the region information is the protagonist during the rest of the process. However, the techniques mentioned in [53, 172, 173] and others published with posteriority could perhaps provide further improvements in C^3S performance.

A

Multi-line Least Squares Fitting

A.1 Common-Origin Multi-Line Least Squares Fitting

Given M sets of points $(f(x_{k,i}), y_{k,i}), k \in \{1..M\}, i \in \{1..N\}$, where the sets are known to lie in M possibly different straight lines sharing a common origin (i.e. the same intercept with the Y axis), the best slopes a_k and the best common origin b, in the least squares sense, can be expressed as the minimizers of:

$$\Phi = \sum_i \left(\frac{y_{1,i} - (a_1 f(x_{1,i}) + b)}{\sigma_{1,i}} \right)^2 + \cdots + \sum_i \left(\frac{y_{M,i} - (a_M f(x_{M,i}) + b)}{\sigma_{M,i}} \right)^2 , \tag{A.1}$$

where $\sigma_{k,i}$ is the uncertainty in $y_{k,i}$.

For simplicity, but without compromising the generality of the method, the problem will be restricted in the following to the case of 3 sets of data ($M = 3$). In this way, equation A.1 can be expressed in matrix form as:

$$\Phi = \left(\overbrace{\begin{pmatrix} \frac{y_{1,1}}{\sigma_{1,1}} \\ \vdots \\ \frac{y_{1,N}}{\sigma_{1,N}} \\ \hline \frac{y_{2,1}}{\sigma_{2,1}} \\ \vdots \\ \frac{y_{2,N}}{\sigma_{2,N}} \\ \hline \frac{y_{3,1}}{\sigma_{3,1}} \\ \vdots \\ \frac{y_{3,N}}{\sigma_{3,N}} \end{pmatrix}}^{Y} - \overbrace{\begin{pmatrix} \frac{f(x_{1,1})}{\sigma_{1,1}} & 0 & 0 & \frac{1}{\sigma_{1,1}} \\ & \vdots & & \\ \frac{f(x_{1,N})}{\sigma_{1,N}} & 0 & 0 & \frac{1}{\sigma_{1,N}} \\ \hline 0 & \frac{f(x_{2,1})}{\sigma_{2,1}} & 0 & \frac{1}{\sigma_{2,1}} \\ & \vdots & & \\ 0 & \frac{f(x_{2,N})}{\sigma_{2,N}} & 0 & \frac{1}{\sigma_{2,N}} \\ \hline 0 & 0 & \frac{f(x_{3,1})}{\sigma_{3,1}} & \frac{1}{\sigma_{3,1}} \\ & \vdots & & \\ 0 & 0 & \frac{f(x_{3,N})}{\sigma_{3,N}} & \frac{1}{\sigma_{3,N}} \end{pmatrix}}^{X} \overbrace{\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ b \end{pmatrix}}^{p} \right)^2 . \tag{A.2}$$

As usual, the minimum of Φ is obtained differenciating Φ respect to p, equating to 0 and solving the resultant linear system:

$$\frac{\partial \Phi}{\partial p} = 2X^T (Y - Xp) = 0 \Rightarrow \overbrace{\left(X^T X\right)}^{A} p = \overbrace{X^T Y}^{B} \tag{A.3}$$

where:

$$A = \begin{pmatrix} \sum \frac{f^2(x_{1,i})}{\sigma_{1,i}^2} & 0 & 0 & \sum \frac{f(x_{1,i})}{\sigma_{1,i}^2} \\ 0 & \sum \frac{f^2(x_{2,i})}{\sigma_{2,i}^2} & 0 & \sum \frac{f(x_{2,i})}{\sigma_{2,i}^2} \\ 0 & 0 & \sum \frac{f^2(x_{3,i})}{\sigma_{3,i}^2} & \sum \frac{f(x_{3,i})}{\sigma_{3,i}^2} \\ \sum \frac{f(x_{1,i})}{\sigma_{1,i}^2} & \sum \frac{f(x_{2,i})}{\sigma_{2,i}^2} & \sum \frac{f(x_{3,i})}{\sigma_{3,i}^2} & \left(\sum \frac{1}{\sigma_{1,i}^2} + \sum \frac{1}{\sigma_{2,i}^2} + \sum \frac{1}{\sigma_{3,i}^2}\right) \end{pmatrix}$$

$$B = \begin{pmatrix} \sum \frac{f(x_{1,i})y_{1,i}}{\sigma_{1,i}^2} \\ \sum \frac{f(x_{2,i})y_{2,i}}{\sigma_{2,i}^2} \\ \sum \frac{f(x_{3,i})y_{3,i}}{\sigma_{3,i}^2} \\ \left(\sum \frac{y_{1,i}}{\sigma_{1,i}^2} + \sum \frac{y_{2,i}}{\sigma_{2,i}^2} + \sum \frac{y_{3,i}}{\sigma_{3,i}^2}\right) \end{pmatrix} \tag{A.4}$$

The uncertainties in the straight lines parameters can now be determined by error propagation of $\sigma_{k,i}$ through the expressions leading to a_1, a_2, a_3 and b [272]. By way of example, using Cramer's rule for calculating b, the uncertainty in b, σ_b, can be shown to satisfy:

$$\begin{aligned} \sigma_b^2 = & \left(A_{22}A_{33}\sum_{i=1}^{N}\left(A_{11}\frac{1}{\sigma_{1,i}} - A_{14}\frac{f(x_{1,i})}{\sigma_{1,i}}\right)\right)^2 |A|^{-2} \\ & + \left(A_{11}A_{33}\sum_{i=1}^{N}\left(A_{22}\frac{1}{\sigma_{2,i}} - A_{24}\frac{f(x_{2,i})}{\sigma_{2,i}}\right)\right)^2 |A|^{-2} \\ & + \left(A_{11}A_{22}\sum_{i=1}^{N}\left(A_{33}\frac{1}{\sigma_{3,i}} - A_{34}\frac{f(x_{3,i})}{\sigma_{3,i}}\right)\right)^2 |A|^{-2} \end{aligned} \tag{A.5}$$

A.2 Common-Slope Multi-Line Least Squares Fitting

Given M sets of points $(f(x_{k,i}), y_{k,i}), k \in \{1..M\}, i \in \{1..N\}$, where the sets are known to lie in M possibly different straight lines sharing a common slope, the best slope a and the best intercepts with the Y axis b_k, in the least squares sense, can be expressed as the minimizers of:

$$\Phi = \sum_i \left(\frac{y_{1,i} - (af(x_{1,i}) + b_1)}{\sigma_{1,i}}\right)^2 + \cdots + \sum_i \left(\frac{y_{M,i} - (af(x_{M,i}) + b_M)}{\sigma_{M,i}}\right)^2, \tag{A.6}$$

where $\sigma_{k,i}$ is the uncertainty in $y_{k,i}$.

For simplicity, but without compromising the generality of the method, the problem will be restricted in the following to the case of 3 sets of data ($M = 3$). In this way, equation A.6 can be expressed in matrix form as:

$$\Phi = \left(\overbrace{\begin{pmatrix} \frac{y_{1,1}}{\sigma_{1,1}} \\ \vdots \\ \frac{y_{1,N}}{\sigma_{1,1}} \\ \hline \frac{y_{2,1}}{\sigma_{2,1}} \\ \vdots \\ \frac{y_{2,N}}{\sigma_{2,1}} \\ \hline \frac{y_{3,1}}{\sigma_{3,1}} \\ \vdots \\ \frac{y_{3,N}}{\sigma_{3,1}} \end{pmatrix}}^{Y} - \overbrace{\begin{pmatrix} \frac{f(x_{1,1})}{\sigma_{1,1}} & \frac{1}{\sigma_{1,1}} & 0 & 0 \\ & \vdots & & \\ \frac{f(x_{1,N})}{\sigma_{1,N}} & \frac{1}{\sigma_{1,N}} & 0 & 0 \\ \hline \frac{f(x_{2,1})}{\sigma_{2,1}} & 0 & \frac{1}{\sigma_{2,1}} & 0 \\ & \vdots & & \\ \frac{f(x_{2,N})}{\sigma_{2,N}} & 0 & \frac{1}{\sigma_{2,N}} & 0 \\ \hline \frac{f(x_{3,1})}{\sigma_{3,1}} & 0 & 0 & \frac{1}{\sigma_{3,1}} \\ & \vdots & & \\ \frac{f(x_{3,N})}{\sigma_{3,N}} & 0 & 0 & \frac{1}{\sigma_{3,N}} \end{pmatrix}}^{X} \overbrace{\begin{pmatrix} a \\ b_1 \\ b_2 \\ b_3 \end{pmatrix}}^{p} \right)^2 . \tag{A.7}$$

As usual, the minimum of Φ is obtained differenciating Φ respect to p, equating to 0 and solving the resultant linear system:

$$\frac{\partial \Phi}{\partial p} = 2X^T\left(Y - Xp\right) = 0 \Rightarrow \overbrace{\left(X^T X\right)}^{A} p = \overbrace{X^T Y}^{B} \tag{A.8}$$

where:

$$A = \begin{pmatrix} \left(\sum \frac{f^2(x_{1,i})}{\sigma_{1,i}^2} + \sum \frac{f^2(x_{2,i})}{\sigma_{2,i}^2} + \sum \frac{f^2(x_{3,i})}{\sigma_{3,i}^2}\right) & \sum \frac{f(x_{1,i})}{\sigma_{1,i}^2} & \sum \frac{f(x_{2,i})}{\sigma_{2,i}^2} & \sum \frac{f(x_{3,i})}{\sigma_{3,i}^2} \\ \sum \frac{f(x_{1,i})}{\sigma_{1,i}^2} & \sum \frac{1}{\sigma_{1,i}^2} & 0 & 0 \\ \sum \frac{f(x_{2,i})}{\sigma_{2,i}^2} & 0 & \sum \frac{1}{\sigma_{2,i}^2} & 0 \\ \sum \frac{f(x_{3,i})}{\sigma_{3,i}^2} & 0 & 0 & \sum \frac{1}{\sigma_{3,i}^2} \end{pmatrix}$$

$$B = \begin{pmatrix} \left(\sum \frac{f(x_{1,i})y_{1,i}}{\sigma_{1,i}^2} + \sum \frac{f(x_{2,i})y_{2,i}}{\sigma_{2,i}^2} + \sum \frac{f(x_{3,i})y_{3,i}}{\sigma_{3,i}^2}\right) \\ \sum \frac{y_{1,i}}{\sigma_{1,i}^2} \\ \sum \frac{y_{2,i}}{\sigma_{2,i}^2} \\ \sum \frac{y_{3,i}}{\sigma_{3,i}^2} \end{pmatrix} \tag{A.9}$$

The uncertainties in the straight lines parameters can now be determined by error propagation of $\sigma_{k,i}$ through the expressions leading to a, b_1, b_2 and b_3 [272]. By way of example, using Cramer's rule for calculating a, the uncertainty in a, σ_a, can be shown to satisfy:

$$\begin{aligned} \sigma_a^2 = & \left(A_{33}A_{44}\sum_{i=1}^{N}\left(A_{22}\frac{f(x_{1,i})}{\sigma_{1,i}} - A_{12}\frac{1}{\sigma_{1,i}}\right)\right)^2 |A|^{-2} \\ & + \left(A_{22}A_{44}\sum_{i=1}^{N}\left(A_{33}\frac{f(x_{2,i})}{\sigma_{2,i}} - A_{13}\frac{1}{\sigma_{2,i}}\right)\right)^2 |A|^{-2} \\ & + \left(A_{22}A_{33}\sum_{i=1}^{N}\left(A_{44}\frac{f(x_{3,i})}{\sigma_{3,i}} - A_{14}\frac{1}{\sigma_{3,i}}\right)\right)^2 |A|^{-2} \end{aligned} \tag{A.10}$$

B

Proofs of chapter 8

Proof of proposition 8.1.

On the one hand:

$$\Delta D_1 = D^{c_1}(i_1, j_1) - D^{c_2}(i_1, j_1) = m_b(i_1, j_1)\,(C_b^{c_1} - C_b^{c_2}) = m_b(i_1, j_1)\Delta C_b$$

Analogously:

$$\Delta D_2 = D^{c_1}(i_2, j_2) - D^{c_2}(i_2, j_2) = m_b(i_2, j_2)\,(C_b^{c_1} - C_b^{c_2}) = m_b(i_2, j_2)\Delta C_b$$

Therefore, taking into account that $m_b(i, j) \geq 0, \forall(i, j)$, $\Delta D_1 \geq 0$ if and only if $\Delta D_2 \geq 0$, and the same applies for $\leq$. Notice the sign of both ΔD_1 and ΔD_2 coincide with the sign of ΔC_b, which is exclusively dependent on the particular surface material and the colour of the illuminant, which is assumed to be constant throughout the scene.

□

Proof of proposition 8.2.

Taking into account the hypotheses:

$$\left(\frac{dD^{c_1}(i, j)}{d\xi}\right) = m_b'(i, j)C_b^{c_1}$$
$$\left(\frac{dD^{c_2}(i, j)}{d\xi}\right) = m_b'(i, j)C_b^{c_2},$$

where $m_b'(i, j)$ is the first-order derivative along direction ξ of $m_b(i, j)$.

Therefore, the sign of both $\left(\frac{dD^{c_1}(i,j)}{d\xi}\right)$ and $\left(\frac{dD^{c_2}(i,j)}{d\xi}\right)$ depend exclusively on the sign of $m_b'(i, j)$ as $C_b^{c_1} \geq 0$ and $C_b^{c_2} \geq 0$.

□

Proof of proposition 8.3.

First, let us introduce the following notation:

$$\Delta m_b = m_b(i_1, j_1) - m_b(i_2, j_2)$$
$$\Delta C_b = C_b^{c_1} - C_b^{c_2}$$

Now, let us consider cases (1) and (2):

(1)($\Rightarrow$) Let us prove first from left to right the double implication. Two cases have to be considered:

$$\text{(a)} \quad \begin{cases} D^{c_1}(i_1, j_1) \geq D^{c_1}(i_2, j_2) \Rightarrow m_b(i_1, j_1) \geq m_b(i_2, j_2) \Rightarrow \Delta m_b \geq 0 \\ D^{c_1}(i_1, j_1) \geq D^{c_2}(i_1, j_1) \Rightarrow C_b^{c_1} \geq C_b^{c_2} \Rightarrow \Delta C_b \geq 0 \end{cases}$$

Therefore:

$$(D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1)) - (D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2)) = \Delta m_b \Delta C_b \geq 0$$

(b) $\begin{cases} D^{c_1}(i_1,j_1) \leq D^{c_1}(i_2,j_2) \Rightarrow m_b(i_1,j_1) \leq m_b(i_2,j_2) \Rightarrow \Delta m_b \leq 0 \\ D^{c_1}(i_1,j_1) \leq D^{c_2}(i_1,j_1) \Rightarrow C_b^{c_1} \leq C_b^{c_2} \Rightarrow \Delta C_b \leq 0 \end{cases}$

Therefore:

$$(D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1)) - (D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2)) = \Delta m_b \Delta C_b \geq 0$$

(⇐) Let us prove now from right to left the double implication:

$$(D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1)) - (D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2)) = \Delta m_b \Delta C_b \geq 0$$
$$\Rightarrow \begin{cases} \text{(a)} \begin{cases} \Delta m_b \geq 0 \Rightarrow D^{c_1}(i_1,j_1) \geq D^{c_1}(i_2,j_2) \\ \Delta C_b \geq 0 \Rightarrow D^{c_1}(i_1,j_1) \geq D^{c_2}(i_1,j_1) \end{cases} \\ \text{or} \\ \text{(b)} \begin{cases} \Delta m_b \leq 0 \Rightarrow D^{c_1}(i_1,j_1) \leq D^{c_1}(i_2,j_2) \\ \Delta C_b \leq 0 \Rightarrow D^{c_1}(i_1,j_1) \leq D^{c_2}(i_1,j_1) \end{cases} \end{cases}$$

(2)(⇒) Let us prove first from left to right the double implication. Two cases have to be considered:

(a) $\begin{cases} D^{c_1}(i_1,j_1) \geq D^{c_1}(i_2,j_2) \Rightarrow m_b(i_1,j_1) \geq m_b(i_2,j_2) \Rightarrow \Delta m_b \geq 0 \\ D^{c_1}(i_1,j_1) \leq D^{c_2}(i_1,j_1) \Rightarrow C_b^{c_1} \leq C_b^{c_2} \Rightarrow \Delta C_b \leq 0 \end{cases}$

Therefore:

$$(D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1)) - (D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2)) = \Delta m_b \Delta C_b \leq 0$$

(b) $\begin{cases} D^{c_1}(i_1,j_1) \leq D^{c_1}(i_2,j_2) \Rightarrow m_b(i_1,j_1) \leq m_b(i_2,j_2) \Rightarrow \Delta m_b \leq 0 \\ D^{c_1}(i_1,j_1) \geq D^{c_2}(i_1,j_1) \Rightarrow C_b^{c_1} \geq C_b^{c_2} \Rightarrow \Delta C_b \geq 0 \end{cases}$

Therefore:

$$(D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1)) - (D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2)) = \Delta m_b \Delta C_b \leq 0$$

(⇐) Let us prove now from right to left the double implication:

$$(D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1)) - (D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2)) = \Delta m_b \Delta C_b \leq 0$$
$$\Rightarrow \begin{cases} \text{(a)} \begin{cases} \Delta m_b \geq 0 \Rightarrow D^{c_1}(i_1,j_1) \geq D^{c_1}(i_2,j_2) \\ \Delta C_b \leq 0 \Rightarrow D^{c_1}(i_1,j_1) \leq D^{c_2}(i_1,j_1) \end{cases} \\ \text{or} \\ \text{(b)} \begin{cases} \Delta m_b \leq 0 \Rightarrow D^{c_1}(i_1,j_1) \leq D^{c_1}(i_2,j_2) \\ \Delta C_b \geq 0 \Rightarrow D^{c_1}(i_1,j_1) \geq D^{c_2}(i_1,j_1) \end{cases} \end{cases}$$

□

Proof of lemma 8.4.

On the one hand, if interface reflection is not involved in the reflected light, ambient reflectance simplifies to body reflectance and thus:

$$\begin{aligned} C_a^{c_1} - C_a^{c_2} &= q_0^{c_1} \int_\Lambda L_a(\lambda)\rho_b(\lambda)\tau^{c_1}(\lambda)s(\lambda)\ d\lambda - q_0^{c_2} \int_\Lambda L_a(\lambda)\rho_b(\lambda)\tau^{c_2}(\lambda)s(\lambda)\ d\lambda \\ &= \alpha \left(q_0^{c_1} \int_\Lambda L_d(\lambda)\rho_b(\lambda)\tau^{c_1}(\lambda)s(\lambda)\ d\lambda - q_0^{c_2} \int_\Lambda L_d(\lambda)\rho_b(\lambda)\tau^{c_2}(\lambda)s(\lambda)\ d\lambda \right) \end{aligned}$$

On the other hand:

$$C_b^{c_1} - C_b^{c_2} = q_0^{c_1} \int_\Lambda L_d(\lambda)\rho_b(\lambda)\tau^{c_1}(\lambda)s(\lambda)\ d\lambda - q_0^{c_2} \int_\Lambda L_d(\lambda)\rho_b(\lambda)\tau^{c_2}(\lambda)s(\lambda)\ d\lambda$$

Therefore, given that $\alpha \geq 0$, $C_a^{c_1} - C_a^{c_2} \geq 0$ if and only if $C_b^{c_1} - C_b^{c_2} \geq 0$ and $C_a^{c_1} - C_a^{c_2} \leq 0$ if and only if $C_b^{c_1} - C_b^{c_2} \leq 0$.

□

Proof of proposition 8.5.

On the one hand:

$$\begin{aligned} \Delta D_1 = D^{c_1}(i_1, j_1) - D^{c_2}(i_1, j_1) &= (C_a^{c_1} - C_a^{c_2}) + m_b(i_1, j_1)\,(C_b^{c_1} - C_b^{c_2}) \\ &= \Delta C_a + m_b(i_1, j_1)\Delta C_b \end{aligned}$$

Analogously:

$$\begin{aligned} \Delta D_2 = D^{c_1}(i_2, j_2) - D^{c_2}(i_2, j_2) &= (C_a^{c_1} - C_a^{c_2}) + m_b(i_2, j_2)\,(C_b^{c_1} - C_b^{c_2}) \\ &= \Delta C_a + m_b(i_2, j_2)\Delta C_b \end{aligned}$$

Therefore, taking into account that $m_b(i, j) \geq 0, \forall (i, j)$ and that ΔC_a and ΔC_b have the same sign by lemma 8.4, $\Delta D_1 \geq 0$ if and only if $\Delta D_2 \geq 0$, and the same applies for $\leq$. Notice the sign of both ΔD_1 and ΔD_2 coincide with the common sign of ΔC_a and ΔC_b.

□

Proof of proposition 8.6.

Since at (i, j) does not take place at a material change by hypothesis, the ambient reflection components vanishes in the first-order derivatives. As a result, the derivatives of the image formation model including ambient illumination and body reflection coincide with the derivatives of the image formation model consisting of just body reflection. Therefore, the same proof as of proposition 8.2 applies here.

□

Proof of proposition 8.7.

First, let us introduce the following notation:

$$\begin{aligned} \Delta m_b &= m_b(i_1, j_1) - m_b(i_2, j_2) \\ \Delta C_b &= C_b^{c_1} - C_b^{c_2} \end{aligned}$$

Now, the following derives directly from the image formation model $D^c = C_a^c + m_b(i, j)C_b^c$:

(a) $D^c(i_1, j_1) - D^c(i_2, j_2) = \Delta m_b C_b^c$,

(b) $D^{c_1}(i, j) - D^{c_2}(i, j) \geq 0 \Leftrightarrow \Delta C_b \geq 0$ and $D^{c_1}(i, j) - D^{c_2}(i, j) \leq 0 \Leftrightarrow \Delta C_b \leq 0$, by virtue of the hypothesis of proportionality between L_a and L_d and lemma 8.4, and

(c) $(D^{c_1}(i_1, j_1) - D^{c_2}(i_1, j_1)) - (D^{c_1}(i_2, j_2) - D^{c_2}(i_2, j_2)) = \Delta m_b \Delta C_b$.

Given the previous facts, the proof of proposition 8.5 also applies here.

□

Proof of proposition 8.8.

If the NIR formation model is assumed, then $\rho_i(\lambda) = \rho_i, \forall \lambda$. Therefore, according to equations 4.1 and 4.6:

$$\begin{aligned} D^c(i, j) &= m_b(i, j)C_b^c(i, j) + m_i(i, j)C_i^c(i, j) \\ &= m_b(i, j)C_b^c(i, j) + m_i(i, j)\rho_i(i, j)L_d^c \end{aligned}$$

where $L_d^c = q_0^c \int_\Lambda L_d(\lambda)\tau^c(\lambda)s(\lambda)d\lambda$ is the (scaled) radiance of directional illumination for colour channel c.

If L_d^c is assumed to be known for colour channels c_1 and c_2, then:

$$\begin{aligned}\Delta^* D &= D^{c_1}(i,j) - D^{c_2}(i,j)\frac{L_d^{c_1}}{L_d^{c_2}} \\ &= m_b(i,j)C_b^{c_1}(i,j) + m_i(i,j)\rho_i(i,j)L_d^{c_1} - m_b(i,j)C_b^{c_2}(i,j)\frac{L_d^{c_1}}{L_d^{c_2}} - m_i(i,j)\rho_i(i,j)L_d^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}} \\ &= m_b(i,j)\left(C_b^{c_1}(i,j) - C_b^{c_2}(i,j)\frac{L_d^{c_1}}{L_d^{c_2}}\right)\end{aligned}$$

As a consequence, on the one hand:

$$\begin{aligned}\Delta^* D_1 &= D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1)\frac{L_d^{c_1}}{L_d^{c_2}} \\ &= m_b(i_1,j_1)\left(C_b^{c_1} - C_b^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}}\right) = m_b(i_1,j_1)\Delta^* C_b\end{aligned}$$

Analogously, on the other hand:

$$\begin{aligned}\Delta^* D_2 &= D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2)\frac{L_d^{c_1}}{L_d^{c_2}} \\ &= m_b(i_2,j_2)\left(C_b^{c_1} - C_b^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}}\right) = m_b(i_2,j_2)\Delta^* C_b\end{aligned}$$

Therefore, taking into account that $m_b(i,j) \geq 0, \forall(i,j)$, $\Delta^* D_1 \geq 0$ if and only if $\Delta^* D_2 \geq 0$, and the same applies for $\leq$. Notice the sign of both $\Delta^* D_1$ and $\Delta^* D_2$ coincide with the sign of $\Delta^* C_b$, which is exclusively dependent on the particular surface material and the colour of the illuminant, which is assumed to be constant throughout the scene.

□

Proof of lemma 8.9.

On the one hand, since ambient reflectance is a linear combination of the body and interface reflectances:

$$\rho_a(\lambda) = \kappa_b\rho_b(\lambda) + \kappa_i\rho_i(\lambda),\ \ \kappa_b \geq 0, \kappa_i \geq 0$$

Therefore:

$$\begin{aligned}&C_a^{c_1} - C_a^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}} = \\ &q_0^{c_1}\int_\Lambda L_a(\lambda)\rho_a(\lambda)\tau^{c_1}(\lambda)s(\lambda)\ d\lambda - q_0^{c_2}\left(\int_\Lambda L_a(\lambda)\rho_a(\lambda)\tau^{c_2}(\lambda)s(\lambda)\ d\lambda\right)\frac{L_d^{c_1}}{L_d^{c_2}} = \\ &q_0^{c_1}\kappa_b\int_\Lambda L_a(\lambda)\rho_b(\lambda)\tau^{c_1}(\lambda)s(\lambda)\ d\lambda - q_0^{c_2}\kappa_b\left(\int_\Lambda L_a(\lambda)\rho_b(\lambda)\tau^{c_2}(\lambda)s(\lambda)\ d\lambda\right)\frac{L_d^{c_1}}{L_d^{c_2}} + \\ &q_0^{c_1}\kappa_i\int_\Lambda L_a(\lambda)\rho_i(\lambda)\tau^{c_1}(\lambda)s(\lambda)\ d\lambda - q_0^{c_2}\kappa_i\left(\int_\Lambda L_a(\lambda)\rho_i(\lambda)\tau^{c_2}(\lambda)s(\lambda)\ d\lambda\right)\frac{L_d^{c_1}}{L_d^{c_2}}\end{aligned}$$

Now, if the NIR model applies here, $\rho_i(\lambda) = \rho_i$, and if the ambient lighting is proportional to the directional illumination, $L_a(\lambda) = \alpha L_d(\lambda)$:

$$
\begin{aligned}
&C_a^{c_1} - C_a^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}} = \\
&q_0^{c_1}\kappa_b\int_\Lambda L_a(\lambda)\rho_b(\lambda)\tau^{c_1}(\lambda)s(\lambda)\ d\lambda - q_0^{c_2}\kappa_b\left(\int_\Lambda L_a(\lambda)\rho_b(\lambda)\tau^{c_2}(\lambda)s(\lambda)\ d\lambda\right)\frac{L_d^{c_1}}{L_d^{c_2}} + \\
&q_0^{c_1}\kappa_i\rho_i\int_\Lambda L_a(\lambda)\tau^{c_1}(\lambda)s(\lambda)\ d\lambda - q_0^{c_2}\kappa_i\rho_i\left(\int_\Lambda L_a(\lambda)\tau^{c_2}(\lambda)s(\lambda)\ d\lambda\right)\frac{L_d^{c_1}}{L_d^{c_2}} = \\
&\alpha\kappa_b\left[\overbrace{q_0^{c_1}\int_\Lambda L_d(\lambda)\rho_b(\lambda)\tau^{c_1}(\lambda)s(\lambda)\ d\lambda}^{C_b^{c_1}} - \overbrace{q_0^{c_2}\left(\int_\Lambda L_d(\lambda)\rho_b(\lambda)\tau^{c_2}(\lambda)s(\lambda)\ d\lambda\right)}^{C_b^{c_2}}\frac{L_d^{c_1}}{L_d^{c_2}}\right] + \\
&\alpha\kappa_i\rho_i\left[\overbrace{q_0^{c_1}\int_\Lambda L_d(\lambda)\tau^{c_1}(\lambda)s(\lambda)\ d\lambda}^{L_d^{c_1}} - \overbrace{\left(q_0^{c_2}\int_\Lambda L_d(\lambda)\tau^{c_2}(\lambda)s(\lambda)\ d\lambda\right)}^{L_d^{c_2}}\frac{L_d^{c_1}}{L_d^{c_2}}\right] = \\
&\alpha\kappa_b\left(C_b^{c_1} - C_b^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}}\right)
\end{aligned}
$$

Therefore, given that $\alpha \geq 0$ and $\kappa_b \geq 0$, $C_a^{c_1} - C_a^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}} \geq 0$ if and only if $C_b^{c_1} - C_b^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}} \geq 0$ and $C_a^{c_1} - C_a^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}} \leq 0$ if and only if $C_b^{c_1} - C_b^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}} \leq 0$.

□

Proof of proposition 8.10.

If the NIR model holds:

$$D^c(i,j) = C_a^c + m_b(i,j)C_b^c + m_i(i,j)\rho_i L_d^c$$

Therefore, on the one hand:

$$
\begin{aligned}
\Delta^* D_1 &= D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1)\frac{L_d^{c_1}}{L_d^{c_2}} \\
&= \left(C_a^{c_1} - C_a^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}}\right) + m_b(i_1,j_1)\left(C_b^{c_1} - C_b^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}}\right) + m_i(i_1,j_1)\left(\rho_i L_d^{c_1} - \rho_i L_d^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}}\right) \\
&= \Delta^* C_a + m_b(i_1,j_1)\Delta^* C_b
\end{aligned}
$$

Analogously:

$$\Delta^* D_2 = D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2)\frac{L_d^{c_1}}{L_d^{c_2}} = \Delta^* C_a + m_b(i_2,j_2)\Delta^* C_b$$

Now, given the proportionality between ambient and directional illumination, and resorting to the proof of lemma 8.9, clearly $\Delta^* C_a \geq 0$ if and only if $\Delta^* C_b \geq 0$, and analogously for case $\leq$. Therefore, since $m_b(i,j) \geq 0, \forall(i,j)$, then the signs of both $\Delta^* D_1$ and $\Delta^* D_2$ coincide with the common sign of $\Delta^* C_a$ and $\Delta^* C_b$.

□

Proof of proposition 8.11.

First, let us introduce the following notation:

$$
\begin{aligned}
\Delta^* C_a(i,j) &= C_a^{c_1}(i,j) - C_a^{c_2}(i,j)\frac{L_d^{c_1}}{L_d^{c_2}} \\
\Delta^* C_b(i,j) &= C_b^{c_1}(i,j) - C_b^{c_2}(i,j)\frac{L_d^{c_1}}{L_d^{c_2}}
\end{aligned}
$$

(1) On the one hand:

$$
\begin{aligned}
D^{c_1}(i_1,j_1) \; - \; D^{c_2}(i_1,j_1)\frac{L_d^{c_1}}{L_d^{c_2}} \geq 0 \\
\Leftrightarrow C_a^{c_1}(i_1,j_1) - C_a^{c_2}(i_1,j_1)\frac{L_d^{c_1}}{L_d^{c_2}} + \\
m_b(i_1,j_1)\left(C_b^{c_1}(i_1,j_1) - C_b^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}}\right) + m_i(i_1,j_1)\left(\rho_i L_d^{c_1} - \rho_i L_d^{c_2}\frac{L_d^{c_1}}{L_d^{c_2}}\right) \geq 0 \\
\Leftrightarrow \Delta^* C_a(i_1,j_1) + m_b(i_1,j_1)\Delta^* C_b(i_1,j_1) \geq 0
\end{aligned}
$$

Due to the proportionality between ambient illumination and directional illumination, $\Delta^* C_a = \alpha \Delta^* C_b$ (see the proof of lemma 8.9). Therefore, since $m_b(i,j) \geq 0, \forall (i,j)$:

$$\Delta^* C_a(i_1,j_1) + m_b(i_1,j_1)\Delta^* C_b(i_1,j_1) \geq 0 \Leftrightarrow \Delta^* C_b(i_1,j_1) \geq 0$$

On the other hand:

$$D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2)\frac{L_d^{c_1}}{L_d^{c_2}} \leq 0 \Leftrightarrow \Delta^* C_a(i_2,j_2) + m_b(i_2,j_2)\Delta^* C_b(i_2,j_2) \leq 0$$

By the same reasons as before:

$$\Delta^* C_a(i_2,j_2) + m_b(i_2,j_2)\Delta^* C_b(i_2,j_2) \leq 0 \Leftrightarrow \Delta^* C_b(i_2,j_2) \leq 0$$

(2) See the proof of case (1) just exchanging (i_1,j_1) and (i_2,j_2).

□

Proof of proposition 8.12.

First, let us introduce the following notation:

$$
\begin{aligned}
C_a^c(i_k,j_k) = C_{a,k}^c \\
C_b^c(i_k,j_k) = C_{b,k}^c \\
m_b(i_k,j_k) = m_{b,k}
\end{aligned}
$$

(1) On the one hand:

$$D^{c_1}(i_1,j_1) - D^{c_1}(i_2,j_2) = C_{a,1}^{c_1} + m_{b,1}C_{b,1}^{c_1} - C_{a,2}^{c_1} - m_{b,2}C_{b,2}^{c_1} > 0$$

Due to the proportionality between ambient and directional illumination, $C_a^c(i,j) = \alpha C_b^c(i,j)$ (see the proof of lemma 8.4). Therefore:

$$
\begin{aligned}
C_{a,1}^{c_1} + m_{b,1}C_{b,1}^{c_1} - C_{a,1}^{c_1} - m_{b,2}C_{b,2}^{c_1} > 0 \\
\Rightarrow (\alpha + m_{b,1})C_{b,1}^{c_1} > (\alpha + m_{b,2})C_{b,2}^{c_1} \\
\Rightarrow \frac{\alpha + m_{b,1}}{\alpha + m_{b,2}} > \frac{C_{b,2}^{c_1}}{C_{b,1}^{c_1}}
\end{aligned}
$$

On the other hand:

$$D^{c_2}(i_1,j_1) - D^{c_2}(i_2,j_2) = C_{a,1}^{c_2} + m_{b,1}C_{b,1}^{c_2} - C_{a,2}^{c_2} - m_{b,2}C_{b,2}^{c_2} < 0$$

As before:

$$
\begin{aligned}
C_{a,1}^{c_2} + m_{b,1}C_{b,1}^{c_2} - C_{a,1}^{c_2} - m_{b,2}C_{b,2}^{c_2} < 0 \\
\Rightarrow (\alpha + m_{b,1})C_{b,1}^{c_2} < (\alpha + m_{b,2})C_{b,2}^{c_2} \\
\Rightarrow \frac{\alpha + m_{b,1}}{\alpha + m_{b,2}} < \frac{C_{b,2}^{c_2}}{C_{b,1}^{c_2}}
\end{aligned}
$$

Finally:

$$\frac{C_{b,2}^{c_1}}{C_{b,1}^{c_1}} < \frac{\alpha + m_{b,1}}{\alpha + m_{b,2}} < \frac{C_{b,2}^{c_2}}{C_{b,1}^{c_2}} \Rightarrow \frac{C_{b,1}^{c_2}}{C_{b,1}^{c_1}} < \frac{C_{b,2}^{c_2}}{C_{b,2}^{c_1}}$$

(2) In a similar way as in case (1), the following derives directly from the hypothesis:

$$\frac{\alpha + m_{b,1}}{\alpha + m_{b,2}} < \frac{C_{b,2}^{c_1}}{C_{b,1}^{c_1}}$$
$$\frac{\alpha + m_{b,1}}{\alpha + m_{b,2}} > \frac{C_{b,2}^{c_2}}{C_{b,1}^{c_2}}$$

Therefore:

$$\frac{C_{b,2}^{c_1}}{C_{b,1}^{c_1}} > \frac{\alpha + m_{b,1}}{\alpha + m_{b,2}} > \frac{C_{b,2}^{c_2}}{C_{b,1}^{c_2}} \Rightarrow \frac{C_{b,1}^{c_2}}{C_{b,1}^{c_1}} > \frac{C_{b,2}^{c_2}}{C_{b,2}^{c_1}}$$

□

Proof of proposition 8.13.

First, let us introduce the following notation:

$$C_a^c(i_k, j_k) = C_{a,k}^c$$
$$C_b^c(i_k, j_k) = C_{b,k}^c$$
$$m_b(i_k, j_k) = m_{b,k}$$

(1) Due to the proportionality between ambient and directional illumination, $C_a^c(i,j) = \alpha C_b^c(i,j)$ (see the proof of lemma 8.4). Therefore:

$$\begin{aligned} D^{c_1}(i_1, j_1) \geq D^{c_1}(i_2, j_2) &\Rightarrow C_{a,1}^{c_1} + m_{b,1} C_{b,1}^{c_1} \geq C_{a,2}^{c_1} + m_{b,2} C_{b,2}^{c_1} \\ &\Rightarrow (\alpha + m_{b,1}) C_{b,1}^{c_1} \geq (\alpha + m_{b,2}) C_{b,2}^{c_1} \\ &\Rightarrow \frac{\alpha + m_{b,1}}{\alpha + m_{b,2}} \geq \frac{C_{b,2}^{c_1}}{C_{b,1}^{c_1}} \end{aligned}$$

Analogously:

$$D^{c_2}(i_1, j_1) \geq D^{c_2}(i_2, j_2) \Rightarrow \frac{\alpha + m_{b,1}}{\alpha + m_{b,2}} \geq \frac{C_{b,2}^{c_2}}{C_{b,1}^{c_2}}$$

On the other hand:

$$\begin{aligned} D^{c_1}(i_1, j_1) \geq D^{c_2}(i_1, j_1) &\Rightarrow C_{a,1}^{c_1} + m_{b,1} C_{b,1}^{c_1} \geq C_{a,1}^{c_2} + m_{b,1} C_{b,1}^{c_2} \\ &\Rightarrow (\alpha + m_{b,1}) C_{b,1}^{c_1} \geq (\alpha + m_{b,1}) C_{b,1}^{c_2} \Rightarrow C_{b,1}^{c_1} \geq C_{b,1}^{c_2} \end{aligned}$$

Analogously:

$$D^{c_1}(i_2, j_2) \geq D^{c_2}(i_2, j_2) \Rightarrow C_{b,2}^{c_1} \geq C_{b,2}^{c_2}$$

Using the last hypothesis:

$$\begin{aligned} D^{c_1}(i_1, j_1) \; - \; D^{c_2}(i_1, j_1) &< D^{c_1}(i_2, j_2) - D^{c_2}(i_2, j_2) \\ &\Rightarrow (\alpha + m_{b,1}) \overbrace{\left(C_{b,1}^{c_1} - C_{b,1}^{c_2}\right)}^{\geq 0} < (\alpha + m_{b,2}) \overbrace{\left(C_{b,2}^{c_1} - C_{b,2}^{c_2}\right)}^{\geq 0} \\ &\Rightarrow \alpha + m_{b,1} < \alpha + m_{b,2} \Rightarrow \frac{\alpha + m_{b,1}}{\alpha + m_{b,2}} < 1 \end{aligned}$$

Therefore:

$$\left.\begin{array}{l}\frac{\alpha+m_{b,1}}{\alpha+m_{b,2}} \geq \frac{C_{b,2}^{c_1}}{C_{b,1}^{c_1}}\\ \frac{\alpha+m_{b,1}}{\alpha+m_{b,2}} \geq \frac{C_{b,2}^{c_2}}{C_{b,1}^{c_2}}\\ \frac{\alpha+m_{b,1}}{\alpha+m_{b,2}} < 1\end{array}\right\} \Rightarrow \left\{\begin{array}{l}C_{b,1}^{c_1} > C_{b,2}^{c_1}\\ C_{b,1}^{c_2} > C_{b,2}^{c_2}\end{array}\right.$$

(2) Taking into account the proof for case (1):

$$\left.\begin{array}{l}D^{c_1}(i_1,j_1) \leq D^{c_1}(i_2,j_2) \Rightarrow \frac{\alpha+m_{b,1}}{\alpha+m_{b,2}} \leq \frac{C_{b,2}^{c_1}}{C_{b,1}^{c_1}}\\ D^{c_2}(i_1,j_1) \leq D^{c_2}(i_2,j_2) \Rightarrow \frac{\alpha+m_{b,1}}{\alpha+m_{b,2}} \leq \frac{C_{b,2}^{c_2}}{C_{b,1}^{c_2}}\\ D^{c_1}(i_1,j_1) \leq D^{c_2}(i_1,j_1) \Rightarrow C_{b,1}^{c_1} \leq C_{b,1}^{c_2}\\ D^{c_1}(i_2,j_2) \leq D^{c_2}(i_2,j_2) \Rightarrow C_{b,2}^{c_1} \leq C_{b,2}^{c_2}\\ D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1) < D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2)\\ \quad \Rightarrow (\alpha+m_{b,1})\overbrace{\left(C_{b,1}^{c_1} - C_{b,1}^{c_2}\right)}^{\leq 0} < (\alpha+m_{b,2})\overbrace{\left(C_{b,1}^{c_1} - C_{b,1}^{c_2}\right)}^{\leq 0}\\ \quad \Rightarrow \alpha+m_{b,1} > \alpha+m_{b,2} \Rightarrow \frac{\alpha+m_{b,1}}{\alpha+m_{b,2}} > 1\end{array}\right\} \Rightarrow \left\{\begin{array}{l}C_{b,1}^{c_1} < C_{b,2}^{c_1}\\ C_{b,1}^{c_2} < C_{b,2}^{c_2}\end{array}\right.$$

(3) Taking into account the proof for case (1):

$$\left.\begin{array}{l}D^{c_1}(i_1,j_1) \geq D^{c_1}(i_2,j_2) \Rightarrow \frac{\alpha+m_{b,1}}{\alpha+m_{b,2}} \geq \frac{C_{b,2}^{c_1}}{C_{b,1}^{c_1}}\\ D^{c_2}(i_1,j_1) \geq D^{c_2}(i_2,j_2) \Rightarrow \frac{\alpha+m_{b,1}}{\alpha+m_{b,2}} \geq \frac{C_{b,2}^{c_2}}{C_{b,1}^{c_2}}\\ D^{c_1}(i_1,j_1) \leq D^{c_2}(i_1,j_1) \Rightarrow C_{b,1}^{c_1} \leq C_{b,1}^{c_2}\\ D^{c_1}(i_2,j_2) \leq D^{c_2}(i_2,j_2) \Rightarrow C_{b,2}^{c_1} \leq C_{b,2}^{c_2}\\ D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1) > D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2)\\ \quad \Rightarrow (\alpha+m_{b,1})\overbrace{\left(C_{b,1}^{c_1} - C_{b,1}^{c_2}\right)}^{\leq 0} > (\alpha+m_{b,2})\overbrace{\left(C_{b,1}^{c_1} - C_{b,1}^{c_2}\right)}^{\leq 0}\\ \quad \Rightarrow \alpha+m_{b,1} < \alpha+m_{b,2} \Rightarrow \frac{\alpha+m_{b,1}}{\alpha+m_{b,2}} < 1\end{array}\right\} \Rightarrow \left\{\begin{array}{l}C_{b,1}^{c_1} > C_{b,2}^{c_1}\\ C_{b,1}^{c_2} > C_{b,2}^{c_2}\end{array}\right.$$

(4) Taking into account the proof for case (1):

$$\left.\begin{array}{l}D^{c_1}(i_1,j_1) \leq D^{c_1}(i_2,j_2) \Rightarrow \frac{\alpha+m_{b,1}}{\alpha+m_{b,2}} \leq \frac{C_{b,2}^{c_1}}{C_{b,1}^{c_1}}\\ D^{c_2}(i_1,j_1) \leq D^{c_2}(i_2,j_2) \Rightarrow \frac{\alpha+m_{b,1}}{\alpha+m_{b,2}} \leq \frac{C_{b,2}^{c_2}}{C_{b,1}^{c_2}}\\ D^{c_1}(i_1,j_1) \geq D^{c_2}(i_1,j_1) \Rightarrow C_{b,1}^{c_1} \geq C_{b,1}^{c_2}\\ D^{c_1}(i_2,j_2) \geq D^{c_2}(i_2,j_2) \Rightarrow C_{b,2}^{c_1} \geq C_{b,2}^{c_2}\\ D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1) > D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2)\\ \quad \Rightarrow (\alpha+m_{b,1})\overbrace{\left(C_{b,1}^{c_1} - C_{b,1}^{c_2}\right)}^{\geq 0} > (\alpha+m_{b,2})\overbrace{\left(C_{b,1}^{c_1} - C_{b,1}^{c_2}\right)}^{\geq 0}\\ \quad \Rightarrow \alpha+m_{b,1} > \alpha+m_{b,2} \Rightarrow \frac{\alpha+m_{b,1}}{\alpha+m_{b,2}} > 1\end{array}\right\} \Rightarrow \left\{\begin{array}{l}C_{b,1}^{c_1} < C_{b,2}^{c_1}\\ C_{b,1}^{c_2} < C_{b,2}^{c_2}\end{array}\right.$$

□

Proof of proposition 8.14.

First, let us introduce the following notation:

$$\Delta C_a(i,j) = C_a^{c_1}(i,j) - C_a^{c_2}(i,j)$$
$$\Delta C_b(i,j) = C_b^{c_1}(i,j) - C_b^{c_2}(i,j)$$

(1) On the one hand:

$$D^{c_1}(i_1,j_1) - D^{c_2}(i_1,j_1) > 0 \Leftrightarrow \Delta C_a(i_1,j_1) + m_b(i_1,j_1)\Delta C_b(i_1,j_1) > 0$$

Due to the proportionality between ambient illumination and directional illumination, $\Delta C_a = \alpha \Delta C_b$ (see the proof of lemma 8.4). Therefore, since $m_b(i,j) \geq 0, \forall (i,j)$:

$$\Delta C_a(i_1,j_1) + m_b(i_1,j_1)\Delta C_b(i_1,j_1) > 0 \Leftrightarrow \Delta C_b(i_1,j_1) > 0$$

On the other hand:

$$D^{c_1}(i_2,j_2) - D^{c_2}(i_2,j_2) < 0 \Leftrightarrow \Delta C_a(i_2,j_2) + m_b(i_2,j_2)\Delta C_b(i_2,j_2) < 0$$

By the same reasons as before:

$$\Delta C_a(i_2,j_2) + m_b(i_2,j_2)\Delta C_b(i_2,j_2) < 0 \Leftrightarrow \Delta C_b(i_2,j_2) < 0$$

(2) See the proof of case (1) just exchanging (i_1,j_1) and (i_2,j_2).

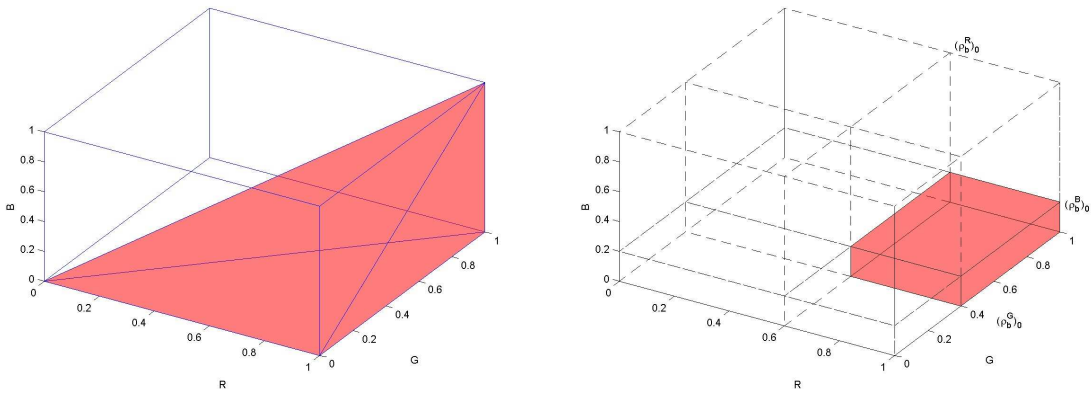
□

C

Reflectance Transitions Covered by the $\mathcal{C}$ Compatibility Relation

As propositions 8.11 and 8.14 show, a transition of type CHC means a change in the order relationship among the different components of the reflectance vector; that is to say, the reflectance for colour channel c_1 is above the reflectance for colour channel c_2 on one side of the transition, while on the other side this ordering is reversed. Observe that for a 3-colour-channel imaging sensor (i.e. a typical RGB camera) there are $3! = 6$ such possible orderings: RGB, RBG, GRB, GBR, BRG and BGR, where, in particular and by way of notation from now on, RGB comprises all the reflectance triplets $(\rho_b^R, \rho_b^G, \rho_b^B)$ such that $\rho_b^R > \rho_b^G > \rho_b^B$. Consequently, if the space of all possible reflectance triplets is considered, a certain ordering (i.e. the set of all the triplets $(\rho_b^R, \rho_b^G, \rho_b^B)$ whose components present the same ordering) corresponds to a sixth of this space. Figure C.1(a) shows the space of all possible reflectance triplets over a cube of unit side, and also the volume corresponding to all the reflectance triplets such that $\rho_b^R > \rho_b^G > \rho_b^B$, or RGB zone.

Therefore, a reflectance change, let us say, from RGB zone to RBG zone, represents $1/6 \times 1/6 = 1/36$ of all possible reflectance transitions. This can also be derived integrating the hypervolume of all possible transitions from one zone to a different one. By way of illustration, equation C.1 accounts for all possible transitions from RGB to RBG: (in equation C.1, $\big((\rho_b^R)_1, (\rho_b^G)_1, (\rho_b^B)_1\big)$ are the variables used for integrating within the RGB volume, while $\big((\rho_b^R)_2, (\rho_b^G)_2, (\rho_b^B)_2\big)$ are for integrating inside the RBG volume.)



(a) (b)

Fig. C.1. Space of reflectance triplets for a 3-colour-channel imaging sensor: in (a) the coloured volume corresponds to the subspace of reflectance triplets where $\rho_b^R > \rho_b^G > \rho_b^B$ (RGB zone), while in (b) the coloured volume is the subspace of reflectance triplets $\big(\rho_b^R, \rho_b^G, \rho_b^B\big)$, where, given $\big((\rho_b^R)_0, (\rho_b^G)_0, (\rho_b^B)_0\big)$, $(\rho_b^R)_0 < \rho_b^R$ and $(\rho_b^G)_0 < \rho_b^G$ and $(\rho_b^B)_0 > \rho_b^B$.

$$\int_0^1 \int_0^{(\rho_b^R)_1} \int_0^{(\rho_b^G)_1} \int_0^1 \int_0^{(\rho_b^R)_2} \int_0^{(\rho_b^B)_2} d(\rho_b^G)_2\, d(\rho_b^B)_2\, d(\rho_b^R)_2\, d(\rho_b^B)_1\, d(\rho_b^G)_1\, d(\rho_b^R)_1 = \frac{1}{36}\,. \tag{C.1}$$

On the other hand, the number of types of CHC reflectance changes corresponds to variations of 2 elements out of 6, $V_{2,6} = 6!/(6-2)! = 30$ (i.e. there are 6 possible orderings and a CHC reflectance change consists in a transition from any of these 6 to a different one, among the remaining 5). Therefore, a CHC reflectance change represents $30/36 \times 100 = 83.33\,\%$ of the total number of possible transitions. The other 6 cases correspond to transitions where there is no change in the ordering among the reflectance vector components (i.e. from RGB to RGB, from RBG to RBG, from GRB to GRB, etc).

As for NCD and NDD reflectance changes, it is not as easy as for CHC transitions to determine the percentage of reflectance transitions covered by them in the light of propositions 8.12 and 8.13. The problem is that these propositions are not double implications (remember the discussion about this point in section 8.4 and the cases of figures 8.3 and 8.5). Therefore, if, by way of example, reflectance changes from a given $\big((\rho_b^R)_0, (\rho_b^G)_0, (\rho_b^B)_0\big)$ to any $\big(\rho_b^R, \rho_b^G, \rho_b^B\big)$ such that, for any pair of colour channels c_1 and c_2:

$$(\rho_b^{c_1})_0 > \rho_b^{c_1} \text{ and } (\rho_b^{c_2})_0 > \rho_b^{c_2}, \text{ or } (\rho_b^{c_1})_0 < \rho_b^{c_1} \text{ and } (\rho_b^{c_2})_0 < \rho_b^{c_2} \quad \text{(see table 8.2)} \tag{C.2}$$

were accounted for, this would not mean accounting for exclusively reflectance changes of the NDD type. That is to say, if reflectance changes such as the ones of C.2 were counted, as before with CHC transitions, the final number would include more transitions than the wanted ones.

Moreover, the reflectance changes expressed in C.2 cover all the possible transitions. In effect, given a certain triplet of reflectances $\big((\rho_b^R)_0, (\rho_b^G)_0, (\rho_b^B)_0\big)$, the dashed lines of figure C.1(b) divide the space of reflectance triplets $\big(\rho_b^R, \rho_b^G, \rho_b^B\big)$ into the eight zones for which the reflectances for at least two colour channels are simultaneously larger or smaller than the respective components of $\big((\rho_b^R)_0, (\rho_b^G)_0, (\rho_b^B)_0\big)$, according to C.2. In particular, for the coloured volume of figure C.1(b):

$$(\rho_b^R)_0 < \rho_b^R \text{ and } (\rho_b^G)_0 < \rho_b^G \text{ and } (\rho_b^B)_0 > \rho_b^B\,. \tag{C.3}$$

Therefore, all the transitions from $\big((\rho_b^R)_0, (\rho_b^G)_0, (\rho_b^B)_0\big)$ to $\big(\rho_b^R, \rho_b^G, \rho_b^B\big)$ such that at least any two of ρ_b^R, ρ_b^G or ρ_b^B are simultaneously larger or smaller than the corresponding $(\rho_b^R)_0$, $(\rho_b^G)_0$ or $(\rho_b^B)_0$ involve the whole space of transitions from $\big((\rho_b^R)_0, (\rho_b^G)_0, (\rho_b^B)_0\big)$.

Therefore, by way of conclusion, it can be said that the compatibility relationship $\mathcal{C}$ covers more than the $83.33\,\%$ of the total of possible reflectance transitions. Unfortunately, the only upper bound which can be given about the percentage of reflectance transitions caught is just $100\,\%$.

D

The Possibilistic C-means Algorithm (PCM)

D.1 General Formulation of PCM

PCM is an iterative clustering method which assigns a possibility $u_{ij} \in [0,1]$ to every point of the data set x_i depending on its distance to everyone of the considered clusters, parameterized each by θ_j, and then recomputes the θ_j's using all the points of the data set weighted by their possibilities of pertaining to the cluster. After convergence, $\theta \equiv [\theta_1, \ldots, \theta_C]$ is the set of final cluster representatives and u_{ij} gives the degree of compatibility of x_i with the jth cluster representative, or, in other words, the possibility that x_i belongs to the jth cluster.

More formally, given a number of clusters, C, the set of data points, $X = [x_1, x_2, \ldots, x_N]$, and a distance function between data points and cluster representatives, $d(x_i, \theta_j)$, PCM minimizes the following cost function [133]:

$$J(\theta, U) = \sum_{j=1}^{C} \sum_{i=1}^{N} u_{ij}^q d(x_i, \theta_j) + \sum_{j=1}^{C} \eta_j \sum_{i=1}^{N} (1 - u_{ij})^q \,, \tag{D.1}$$

where q and η_j are parameters of PCM, and $U = [u_{ij}]_{N \times C}$ is such that:

- $u_{ij} \in [0,1]$,
- $\max_{j=1,\ldots,C}\{u_{ij}\} > 0, i = 1, \ldots, N$, and
- $0 < \sum_{i=1}^{N} u_{ij} \leq N, j = 1, \ldots, C$.

In this context, u_{ij} may be interpreted as the degree of compatibility of x_i with the jth cluster representative, what only depends on x_i and the cluster representative of the jth cluster. That is, it is independent of the possibilities that x_i belongs to the other clusters. On the other hand, the first term of equation D.1 is the sum of the weighted distances of points x_i to the current cluster representatives, while the second term avoids the trivial solution $u_{ij} = 0$ and minimizes the effects of outliers.

Differentiating D.1 with respect to u_{ij} and setting it to 0 leads to the equation:

$$u_{ij} = \frac{1}{1 + \left(\frac{d(x_i, \theta_j)}{\eta_j}\right)^{\frac{1}{q-1}}} \tag{D.2}$$

while, given the matrix U, the optimum θ_j is given by:

$$\sum_{i=1}^{N} u_{ij}^q \frac{\partial d(x_i, \theta_j)}{\partial \theta_j} = 0 \tag{D.3}$$

The PCM algorithmic scheme derives from equations D.2 and D.3, and is given in pseudocode D.1 [276]. As usual, $||\theta(t) - \theta(t-1)|| < \epsilon$ can be employed as a termination criterion. Based on the preceding generalized scheme, a number of possibilistic clustering algorithms can be defined by adapting the distance function to the problem at hand [133, 134].

Pseudocode D.1 Algorithmic formulation of PCM.

(1) **initialization**
  set η_j and q to appropriate values
  t := 0 and choose initial estimates for $\theta_j(t)$
(2) **repeat**
  for i := 1 **to** N
  for j := 1 **to** C
    $u_{ij}(t) := \frac{1}{1+\left(\frac{d(x_i,\theta_j(t))}{\eta_j}\right)^{\frac{1}{q-1}}}$
  endfor
  endfor
  $t := t + 1$
  for j := 1 **to** C
    solve $\sum_{i=1}^{N} u_{ij}^q(t-1)\frac{\partial d(x_i,\theta_j)}{\partial \theta_j} = \mathbf{0}$ for θ_j, and
    set $\theta_j(t)$ accordingly
  endfor
 until a termination criterion is met

D.2 Meaning and Importance of PCM Parameters

On the one hand, the η_j's determine the distance at which the possibilities u_{ij}'s become 0.5. In this sense, it is said they determine the zone of influence of any point: if the distance between a data point x_i and cluster θ_j is large compared with η_j, x_i will have little influence when updating θ_j. Because of this, η_j is referred to as the *scale*, *bandwidth* or *resolution* parameter for cluster θ_j. Consequently, in general, it is desirable that the η_j's are related to the overall size and shape of the respective cluster θ_j. In practice, it is chosen proportional to the average fuzzy intracluster distance of cluster θ_j:

$$\eta_j = \frac{\sum_{i=1}^{N} u_{ij}^q d(x_i, \theta_j)}{\sum_{i=1}^{N} u_{ij}^q}, \tag{D.4}$$

which implies running first the fuzzy C-means (FCM) clustering algorithm [276], from which a first estimation of matrix U is obtained. However, when the nature of the clusters is known, the η_j's can be fixed *a priori*, such as in the case of line or hyperplane clustering algorithms. In such cases, the clusters are expected to be thin lines or planes and the η_j's can be set to the respective expected thicknesses [133].

On the other hand, q determines the rate of decay of the possibility distribution and, thus, its shape. Figure D.1 plots u_{ij} as a function of $\frac{d(x_i,\theta_j)}{\eta_j}$ for several values of q. As can be seen, increasing values of q represent increased possibility of all points in the data set completely belonging to a given cluster. In this sense, it is said q determines the fuzziness of the final possibilistic partition and is referred to as the *fuzzifier*.

D.3 PCM as a Mode-Seeking Algorithm

Contrarily to the behaviour of other clustering algorithms, such as FCM, which partitions the data set in the specified number of clusters, PCM is a *mode-seeking* algorithm. That is to say, cluster prototypes are automatically attracted to dense regions in feature space as iterations proceed, and, therefore, each final cluster corresponds to a dense region in the data set [134].

The most important consequence of this property is that, provided the θ_j's are initially placed in different dense regions of the data set, the number of clusters need not be known *a priori*. In effect, if PCM is run for C_1 clusters when the data set naturally contains $C_2 < C_1$, some of the C_1 clusters will coincide with others, so that, probably, only C_2 different clusters will have been reported. If $C_2 > C_1$, PCM can potentially find C_1 good clusters out of the real C_2. In this sense, PCM recognizes the structure of the data better.

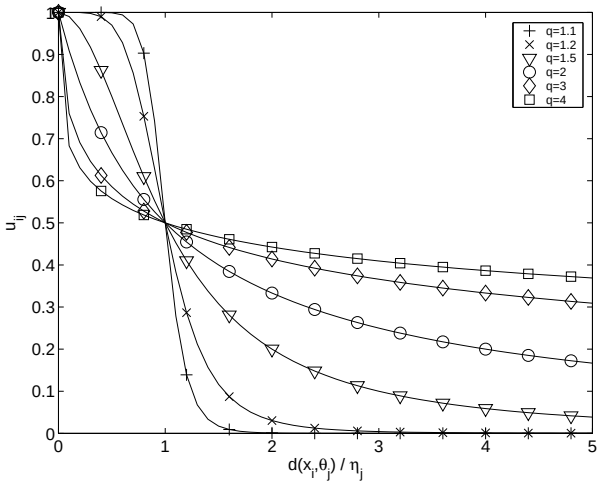


Fig. D.1. Decay of u_{ij} with regard to the fuzzifier q.

References

1. D. Alexander and B. Buxton. Modelling of single mode distributions of colour data using directional statistics. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 319–324, 1997.
2. Y. Aloimonos. Purposive and qualitative active vision. In *Proceedings of the European Conference on Computer Vision*, 1990. (Workshop on Control of Perception in Active Vision).
3. Y. Aloimonos and A. Rosenfeld. A response to "Ignorance, myopia, and naiveté in computer vision systems" by R.C. Jain and T.O. Binford. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 53(1):120–124, 1991.
4. C. Alsina and E. Trillas. *Lecciones de Álgebra y Geometría*. Gustavo Gili, 1984.
5. G. Amelio, M. Tompsett, and G. Smith. Experimental verification of the charge coupled device concept. *Bell Syst. Tech. J.*, 49:593–600, April 1970.
6. A. Baddeley. An error metric for binary images. In W. Förstner and S. Ruwiedel, editors, *Robust Computer Vision: Quality of Vision Algorithms*, pages 59–78, Karlsruhe, 1992. Wichmann. (Proceedings of the International Workshop on Robust Computer Vision).
7. R. Bajcsy, S. W. Lee, and A. Leonardis. Color image segmentation with detection of highlights and local illumination induced by inter-reflections. Technical Report GRASP LAB 182 MS-CIS-89-39, Department of Computer and Information Science, University of Pennsylvania, 1989.
8. R. Bajcsy, S. W. Lee, and A. Leonardis. Color image segmentation with detection of highlights and local illumination induced by inter-reflections. In *Proceedings of the IEEE International Conference on Pattern Recognition*, volume I, pages 785–790, 1990.
9. R. Bajcsy, S. W. Lee, and A. Leonardis. Color image segmentation with detection of highlights and local illumination induced by inter-reflections (reprint). In G. E. Healey, S. A. Shafer, and L. B. Wolff, editors, *Physics-based vision: principles and practice*, volume on Color, pages 199–204. Jones and Bartlett Publishers, 1992.
10. R. Bajcsy, S. W. Lee, and A. Leonardis. Detection of diffuse and specular interface reflections and inter-reflections by color image segmentation. *International Journal of Computer Vision*, 17(3):241–272, 1996.
11. D. Ballard and C. Brown. Computer vision. In *Prentice Hall*, 1982.
12. M. Beauchemin and K. Thomson. The evaluation of segmentation results and the overlapping area matrix. *International Journal of Remote Sensing*, 18(18):3895–3899, 1997.
13. P. Beckmann and A. Spizzichino. *The Scattering of Electromagnetic Waves from Rough Surfaces*. Macmillan, New York, 1963.
14. P. J. Besl and R. C. Jain. Segmentation through symbolic surface descriptions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 77–85, 1986.
15. P. J. Besl and R. C. Jain. Segmentation through variable-order surface fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(2):167–192, 1988.
16. P. J. Besl and R. C. Jain. Segmentation through variable-order surface fitting (reprint). In R. Kasturi and R. C. Jain, editors, *Computer vision*, volume Principles, pages 144–169. IEEE Computer Society Press, 1991.
17. M. Bichsel and A. P. Pentland. A simple algorithm for shape from shading. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 459–465, 1992.
18. M. Borsotti, P. Campadelli, and R. Schettini. Quantitative evaluation of color image segmentation results. *Pattern Recognition Letters*, 19:741–747, 1998.

19. C.-S. Bouganis and M. Brookes. Multiple light source detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(4):509–514, 2004.
20. K. W. Bowyer, C. Kranenburg, and S. Dougherty. Edge detector evaluation using empirical ROC curves. *Computer Vision and Image Understanding*, 84(1):77–103, 2001.
21. W. Boyle and G. Smith. Charge coupled semiconductor devices. *Bell Syst. Tech. J.*, 49:587–593, April 1970.
22. G. Brelstaff and A. Blake. Detecting specular reflections using Lambertian constraints. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 297–302, 1988.
23. M. Brooks and W. Chojnacki. Direct computation of shape from shading. In *Proceedings of the IEEE International Conference on Pattern Recognition*, pages 114–119, 1994.
24. M. Brooks and B. Horn. Shape and source from shading. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 932–936, 1985.
25. J. F. Canny. Finding edges and lines in images. Technical Report 720, MIT AI Laboratory, June 1983.
26. H. Cheng, X. Jiang, Y. Sun, and J. Wang. Color image segmentation: advances and prospects. *Pattern Recognition*, 34:2259–2281, 2001.
27. Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:790–799, 1995.
28. S. I. Cho, H. Saito, and S. Ozawa. A divide-and-conquer strategy in recovering shape of book surface from shading. In *Proceedings of the IAPR International Conference on Image Analysis and Processing*, volume 2, pages 262–269, 1997.
29. D. Comaniciu and P. Meer. Robust analysis of feature spaces: Color image segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1997. (The code is publicly available at http://www.caip.rutgers.edu/~comanici/segm_images.html).
30. D. Comaniciu and P. Meer. Mean-shift analysis and applications. In *Proceedings of the IEEE International Conference on Computer Vision*, 1999. (The code of the segmentation algorithm is publicly available at http://www.caip.rutgers.edu/riul/research/code.html).
31. Commission Internationale de l'Eclairage. *International Lighting Vocabulary*, 1970. (3rd. edition, Publ. CIE 17 (E-1.1)).
32. M. C. Cooper. The tractability of segmentation and scene analysis. *Proceedings of the IEEE International Conference on Computer Vision*, 30(1):27–42, 1998.
33. C. G. Cullen. *An Introduction to Numerical Linear Algebra*. PWS Publishing Company, 1994.
34. K. J. Dana and B. van Ginneken. Reflectance and texture of real-world surfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 151–157, 1997.
35. G. Deng and J. Pinoli. Differentiation-based edge-detection using the logarithmic image-processing model. *Journal of Mathematical Imaging and Vision*, 8(2):161–180, March 1998.
36. A. Desolneux, L. Moisan, and J.-M. Morel. *Seeing, Thinking and Knowing*, chapter Gestalt Theory and Computer Vision, pages 71–101. A. Carsetti editor, Kluwer Academic Publishers, 2004.
37. S. di Zenzo. A note on the gradient of a multiimage. *Computer Vision, Graphics, and Image Processing*, 33(1):116–125, 1986.
38. P. Dupuis and J. Oliensis. Direct method for reconstructing shape from shading. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 453–458, 1992.
39. Eastman Kodak Company - Image Sensor Solutions. *CCD primer: Charge-Coupled Device (CCD) Image Sensors (MTD/PS-0218)*, 2001. Revision 1.0, May 29 [http://www.kodak.com/global/en/digital/ccd/publications/applicationNotes.jhtml].
40. Eastman Kodak Company - Image Sensor Solutions. *CCD Primer: Conversion of Light (Photons) to Electronic Charge (MTD/PS-0217)*, 2001. Revision 1.0, May 29 [http://www.kodak.com/global/en/digital/ccd/publications/applicationNotes.jhtml].
41. Eastman Kodak Company - Image Sensor Solutions. *Color Correction for Image Sensors (MTD/PS-0233)*, 2003. Revision 2.0, January 16 [http://www.kodak.com/global/en/digital/ccd/publications/applicationNotes.jhtml].
42. Eastman Kodak Company - Image Sensor Solutions. *Color Correction for Image Sensors (MTD/PS-0534-2)*, 2003. Revision 2.0, October 27 [http://www.kodak.com/global/en/digital/ccd/publications/applicationNotes.jhtml].
43. G. Edwards, C. Taylor, and T. Cootes. Improving identification performance by integrating evidence from sequences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume I, pages 486–491, 1999.
44. A. Escalera. *Visión por computador: fundamentos y métodos*. Prentice-Hall, 2001.
45. H. Farid and E. Adelson. Separating reflections from images by use of independent component analysis. *Journal of the Optical Society of America - A*, 16(9):2136–2145, 1999.

46. N. Fernandez-Garcia, R. Medina-Carnicer, A. Carmona-Poyato, F. Madrid-Cuevas, and M. Prieto-Villegas. Characterization of empirical discrepancy evaluation measures. *Pattern Recognition Letters*, 25(1):35–47, 2004.
47. G. D. Finlayson, S. D. Hordley, and M. S. Drew. Removing shadows from images. In *Proceedings of the European Conference on Computer Vision*, volume IV, pages 823–836, 2002.
48. G. D. Finlayson, S. D. Hordley, and M. S. Drew. Removing shadows from images using retinex. In *Color Imaging Conference*, pages 73–79, 2002.
49. A. Fitzgibbon and R. Fisher. A buyer's guide to conic fitting. In *Proceedings of the British Machine Vision Conference*, 1995.
50. A. Fitzgibbon, M. Pilu, and R. Fisher. Direct least square fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):476–480, may 1999.
51. L. Forbes and B. Draper. Inconsistencies in edge detector evaluation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume II, pages 398–404, 2000.
52. D. Forsyth. A novel algorithm for colour constancy. *International Journal of Computer Vision*, 5(1):5–36, 1990.
53. J. Freixenet, X. Muñoz, D. Raba, J. Martí, and X. Cufí. Yet another survey on image segmentation: region and boundary information integration. In *Proceedings of the European Conference on Computer Vision*, volume III, pages 408–422, 2002.
54. K. Fu and J. Mui. A survey on image segmentation. *Pattern Recognition*, 13:3–16, 1981.
55. P. Fua and Y. G. Leclerc. Object-centered surface reconstruction: combining multi-image stereo and shading. *International Journal of Computer Vision*, 16(1):35–56, 1995.
56. B. V. Funt and M. S. Drew. Color space analysis of mutual illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(12):1319–1326, 1993.
57. J. Gerbrands, E. Backer, and W. van der Hoeven. Quantitative evaluation of edge detection by dynamic programming. In *Proceedings of Pattern Recognition in Practice*, pages 91–99, 1986.
58. R. Gershon, A. D. Jepson, and J. K. Tsotsos. Ambient illumination and the determination of material changes. *Journal of the Optical Society of America - A*, 3(10):1700–1707, 1986.
59. R. Gershon, A. D. Jepson, and J. K. Tsotsos. The effects of ambient illumination on the structure of shadows in chromatic images. Technical Report RBVC-TR-86-9, Department of Computer Science, University of Toronto, (Ontario - Canada), 1986.
60. R. Gershon, A. D. Jepson, and J. K. Tsotsos. Ambient illumination and the determination of material changes (reprint). In G. E. Healey, S. A. Shafer, and L. B. Wolff, editors, *Physics-based vision: principles and practice*, volume on Color, pages 101–108. Jones and Bartlett Publishers, 1992.
61. J.-M. Geusebroek. *Color and Geometrical Structure in Images*. PhD thesis, University of Amsterdam, Amsterdam, The Netherlands, 2000.
62. J.-M. Geusebroek, R. Boomgaard, A. Smeulders, and A. Dev. Color and scale: the spatial structure of color images. In *Proceedings of the European Conference on Computer Vision*, volume 1, pages 331–341, 2000.
63. J.-M. Geusebroek, R. Boomgaard, A. Smeulders, and H. Geerts. Color invariance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(12):1338–1350, 2001.
64. J.-M. Geusebroek, A. Dev, R. Boomgaard, A. Smeulders, F. Cornelissen, and H. Geerts. Color invariant edge detection. In *Scale-Space Theories in Computer Vision*, pages 459–464, 1999.
65. J.-M. Geusebroek, A. Dev, R. Boomgaard, A. Smeulders, F. Cornelissen, and H. Geerts. A scalespace approach to color invariance. In *Proceedings of the fifth annual conference of the Advanced School for Computing and Imaging*, pages 383–389, 1999.
66. J.-M. Geusebroek, A. Smeulders, and R. Boomgaard. Measurement of color invariants. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 50–57, 2000.
67. J.-M. Geusebroek, A. Smeulders, and R. Boomgaard. Measurement of color invariants (reprint). In A. M. Vossepoel and F. M. Vos, editors, *Fourth Quinquennial Review 1996-2000*, pages 129–138. Dutch Society for Pattern Recognition and Image Processing, 2001.
68. T. Gevers. Adaptive image segmentation by combining photometric invariant region and edge information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):848–852, 2002.
69. T. Gevers. Reflectance-based classification of color edges. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 856–861, 2003.
70. T. Gevers. Robust segmentation and tracking of colored objects in video. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(6):776–781, 2004.
71. T. Gevers, S. Ghebreab, and A. W. Smeulders. Color invariant snakes. In *Proceedings of the British Machine Vision Conference*, pages 578–588, 1998.

72. T. Gevers and F. C. Groen. Segmentation of color images. In *Proceedings of the IAPR Scandinavian Conference on Image Analysis*, 1991.
73. T. Gevers and V. K. Kajcovski. Image segmentation by directed region subdivision. In *Proceedings of the IEEE International Conference on Pattern Recognition*, pages 342–346, 1994.
74. T. Gevers and A. W. Smeulders. Combining region splitting and edge detection through guided delaunay image subdivision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1021–1026, 1997.
75. T. Gevers and A. W. Smeulders. Edge steered region segmentation by photometric color invariant. In *Proceedings of the IAPR Scandinavian Conference on Image Analysis*, 1997.
76. T. Gevers, A. W. Smeulders, and H. Stokman. Photometric invariant region detection. In *Proceedings of the British Machine Vision Conference*, pages 659–668, 1998.
77. T. Gevers and H. Stokman. Reflectance based edge classification. In *Proceedings of Visual Inferface*, pages 25–32, 1999.
78. T. Gevers and H. Stokman. Classifying color transitions into shadow-geometry, illumination, highlight or material edges. In *Proceedings of the IEEE International Conference on Image Processing*, volume I, pages 521–524, 2000.
79. T. Gevers and H. Stokman. Classifying color edges in video into shadow-geometry, highlight, or material transitions. *IEEE Transactions on Multimedia*, 5(2):237–243, 2003.
80. T. Gevers and H. Stokman. Robust photometric invariant region detection in multispectral images. *International Journal of Computer Vision*, 53(2):135–151, 2003.
81. A. Goumeidane, M. Khamadja, B. Belaroussi, H. Benoit-Cattin, and C. Odet. New discrepancy measures for segmentation evaluation. In *Proceedings of the IEEE International Conference on Image Processing*, volume II, pages 411–414, 2003.
82. O. Graydon. Scientific CCD cameras grow in sophistication. *Opto & Laser Europe*, (103), February 2003.
83. R. Hall. *Illumination and color in computer generated imagery*. Springer-Verlag, 1989.
84. R. Haralick and L. Shapiro. Image segmentation techniques. *Computer Vision, Graphics, and Image Processing*, 29:100–132, 1985.
85. R. Haralick and L. Shapiro. *Computer and Robot Vision*. Addison-Wesley, 1992.
86. G. E. Healey. A color reflectance model and its use for segmentation. In *ICCV*, pages 460–466, 1988.
87. G. E. Healey. A parallel color algorithm for segmenting images of 3-D scenes. In *Proceedings of DARPA*, pages 1038–1042, 1989.
88. G. E. Healey. Using color for geometry-insensitive segmentation. *Journal of the Optical Society of America - A*, 6(6):920–937, 1989.
89. G. E. Healey. Using physical color models in 3-D machine vision. In *Proceedings of SPIE - Perceiving, Measuring and Using Color*, volume 1250, pages 264–275, 1990.
90. G. E. Healey. Segmenting images using normalized color. *IEEE Transactions on Systems, Man and Cybernetics*, 1992.
91. G. E. Healey. Segmenting images using normalized color (reprint). In G. E. Healey, S. A. Shafer, and L. B. Wolff, editors, *Physics-based vision: principles and practice*, volume on Color, pages 166–198. Jones and Bartlett Publishers, 1992.
92. G. E. Healey and T. O. Binford. Local shape from specularity. *Computer Vision, Graphics, and Image Processing*, 42:62–86, 1988.
93. G. E. Healey and R. Kondepudy. Radiometric CCD camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276, 1994.
94. G. E. Healey, S. A. Shafer, and L. B. Wolff, editors. *Physics-based vision: principles and practice*. Jones and Bartlett Publishers, 1992. (series consisting of three volumes).
95. G. E. Healey, S. A. Shafer, and L. B. Wolff, editors. *Physics-based vision: principles and practice*, volume on Color. Jones and Bartlett Publishers, 1992.
96. F. Heyden. Evaluation of edge detection algorithms. In *Proceedings of the IEEE International Conference on Image Processing*, pages 618–622, 1989.
97. J. Ho, B. V. Funt, and M. S. Drew. Separating a color signal into illumination and surface reflectance components: theory and aplications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10):966–977, 1990.
98. G. C. Holst. *CCD arrays, cameras, and displays*. SPIE press & JCD Publishing, 2nd edition, 1998.
99. A. Hoover, G. Jean-Baptiste, X. Jiang, P. Flynn, H. Bunke, D. Goldgof, K. Bowyer, D. Eggert, A. Fitzgibbon, and R. Fisher. An experimental comparison of range image segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7):673–689, 1996.
100. B. Horn. *Shape from shading: a method for obtaining the shape of a smooth opaque object from one view*. PhD thesis, MIT, Dep. of Elec. Eng, 1970.

101. B. Horn. Understanding image intensities. *Artificial Intelligence*, 8:201–231, 1977.
102. B. Horn. *Robot Vision*. MIT Press, 1986.
103. B. Horn and M. Brooks, editors. *Shape from shading*. MIT Press, 1989.
104. B. Horn and R. Sjoberg. Calculating the reflectance map. *Applied Optics*, 18(11):1770–1779, 1979.
105. D. R. Hougen and N. Ahuja. Estimation of the light source distribution and its use in integrated shape recovery from stereo and shading. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 148–155, 1993.
106. Q. Huang and B. Dom. Quantitative methods of evaluating image segmentation. In *Proceedings of the IEEE International Conference on Image Processing*, volume 3, pages 53–56, 1995.
107. JAI ®. *Specifications for JAI M-70 Progressive Scan RGB Color Camera*, 1998. [http://www.jai.dk].
108. R. C. Jain, R. Kasturi, and B. G. Schunck. *Machine Vision*. McGraw-Hill, 1995.
109. J. R. Janesick. *Scientific charge-coupled devices*. SPIE Press, 2001.
110. C. X. Jiang and M. O. Ward. Shadow identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 606–612, 1992.
111. C. X. Jiang and M. O. Ward. Shadow segmentation and classification in a constrained environment. *Computer Vision, Graphics, and Image Processing*, 59(2):213–225, 1994.
112. X. Jiang, K. Bowyer, Y. Morioka, S. Hiura, K. Sato, S. Inokuchi, M. Bock, C. Guerra, R. Loke, and J. du Buf. Some further results of experimental comparison of range image segmentation algorithms. In *Proceedings of the IEEE International Conference on Pattern Recognition*, volume IV, pages 877–881, 2000.
113. D. Judd and G. Wyszecki. *Color in Business, Science, and Industry*. John Wiley and Sons, 3rd edition, 1975.
114. B. Julesz. Textons, the elements of texture perception, and their interactions. *Nature*, 290:91–97, 1981.
115. T. Kanade, T. Binford, T. Poggio, and A. Rosenfeld. Vision. In *Annual Review of Computer Science 1990*, volume 4, pages 517–529. Annual Reviews Inc., 1990.
116. R. Kasturi and R. Jain. *Computer Vision: Principles*, volume 1. IEEE Press, 1991.
117. R. Kimmel and A. M. Bruckstein. Global shape from shading. *Computer Vision and Image Understanding*, 62(3):360–369, 1995.
118. R. Kimmel and J. A. Sethian. Optimal algorithm for shape from shading. In *Proceedings of the Asian Conference on Computer Vision*, 2000.
119. R. Klette, K. Schlüns, and A. Koschan. *Computer Vision: Three-Dimensional Data from Images*. Springer-Verlag, 1998.
120. G. J. Klinker. *A physical approach to color image understanding*. PhD thesis, Robotics Institute, Carnegie Mellon University, 1988. (Also published as tech. report CMU-CS-TR-88-161).
121. G. J. Klinker. *A physical approach to color image understanding*. A. K. Peters, 1993.
122. G. J. Klinker, S. A. Shafer, and T. Kanade. Measurement of gloss from color images. In *Proceedings of the Inter-Society Color Council (ISCC) Conference on Appearance*, pages 9–12, 1987.
123. G. J. Klinker, S. A. Shafer, and T. Kanade. Using a color reflection model to separate highlights from object color. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 145–150, 1987.
124. G. J. Klinker, S. A. Shafer, and T. Kanade. Using a color reflection model to separate highlights from object color. In *Proceedings of DARPA - Image Understanding Workshop*, pages 614–619, 1987.
125. G. J. Klinker, S. A. Shafer, and T. Kanade. Color image analysis with an intrinsic reflection model. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 292–296, 1988.
126. G. J. Klinker, S. A. Shafer, and T. Kanade. Image segmentation and reflection analysis through color. In *Proceedings of SPIE - Applications of Artificial Intelligence VI*, volume 937, 1988.
127. G. J. Klinker, S. A. Shafer, and T. Kanade. Image segmentation and reflection analysis through color. In *Proceedings of DARPA - Image Understanding Workshop*, pages 838–853, 1988.
128. G. J. Klinker, S. A. Shafer, and T. Kanade. The measurement of highlights in color images. *International Journal of Computer Vision*, 2(1):7–32, 1988.
129. G. J. Klinker, S. A. Shafer, and T. Kanade. A physical approach to color image understanding. *International Journal of Computer Vision*, 4(1):7–38, 1990.
130. G. J. Klinker, S. A. Shafer, and T. Kanade. The measurement of highlights in color images (reprint). In G. E. Healey, S. A. Shafer, and L. B. Wolff, editors, *Physics-based vision: principles and practice*, volume on Color, pages 309–334. Jones and Bartlett Publishers, 1992.
131. G. J. Klinker, S. A. Shafer, and T. Kanade. A physical approach to color image understanding (reprint). In G. E. Healey, S. A. Shafer, and L. B. Wolff, editors, *Physics-based vision: principles and practice*, volume on Color, pages 134–165. Jones and Bartlett Publishers, 1992.

132. V. Kravtchenko and J. J. Little. Efficient color object segmentation using the dichromatic reflection model. In *Proceedings of the Pacific Rim Conference on Communications, Computers and Signal Processing*, pages 90–94, 1999.
133. R. Krishnapuram and J. M. Keller. A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems*, 1(2):98–110, 1993.
134. R. Krishnapuram and J. M. Keller. The possibilistic C-means algorithm: insights and recommendations. *IEEE Transactions on Fuzzy Systems*, 4(3):385–393, 1996.
135. B. Kurt and M. Gökmen. Two dimensional generalized edge detector. In *Proceedings of the IAPR International Conference on Image Analysis and Processing*, pages 148–153, 1999.
136. L. Lam, S.-W. Lee, and C. Wuen. Thinning methodologies— a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9):869–885, 1992.
137. J. Lambert. *Photometria sive de mensura de gratibus luminis, colorum et umbrae*. Eberhard Klett, Augsberg, Germany, 1760.
138. C. Lawson and R. Hanson. *Solving least squares problems*. Prentice-Hall, 1974.
139. Y. G. Leclerc and A. F. Bobick. The direct computation of height from shading. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 552–558, 1991.
140. C.-H. Lee and A. Rosenfeld. Albedo estimation for scene segmentation. *Pattern Recognition Letters*, 1:155–160, March 1983.
141. C.-H. Lee and A. Rosenfeld. Improved methods of estimating shape from shading using the light source coordinate system. *Artificial Intelligence*, 26:125–143, 1985.
142. C.-H. Lee and A. Rosenfeld. Improved methods of estimating shape from shading using the light source coordinate system. In B. K. P. Horn and M. J. Brooks, editors, *Shape from Shading*, chapter 11, pages 323–347. The MIT Press, 1989.
143. H.-C. Lee. Method for computing the scene-illuminant chromaticity from specular highlights. *Journal of the Optical Society of America - A*, 3(10):1694–1699, 1986.
144. H.-C. Lee. Illuminant color from shading. In *Proceedings of SPIE Perceiving, Measuring, and Using Color*, volume 1250, pages 236–244, 1990.
145. H.-C. Lee, E. J. Breneman, and C. P. Schulte. Modeling light reflection for computer color vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(4):402–409, April 1990.
146. K. Lee and C. Kuo. Shape from shading with a generalized reflectance map model. *Computer Vision and Image Understanding*, 67(2):143–160, 1997.
147. M. Levine and A. Nazif. An experimental rule-based system for testing low level segmentation strategies. In K. Preston and L. Uhr, editors, *Multicomputers and Image Processing: Algorithms and Programs*, pages 149–160. Academic Press, New York, 1982.
148. M. Levine and A. Nazif. Dynamic measurement of computer generated image segmentations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7:155–164, 1985.
149. H. Lewis and M. Brown. A generalized confusion matrix for assessing area estimates from remotely sensed data. *International Journal of Remote Sensing*, 22(16):3223–3235, 2001.
150. Y. Lim and S. Lee. On the color image segmentation algorithms based on the thresholding and fuzzy C-means technique. *Pattern Recognition*, 23:935–952, 1990.
151. B. W. Lindgren. *Statistical Theory*. Chapman & Hall, 4th edition, 1993.
152. J. Liu and Y.-H. Yang. Multiresolution color image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(7):689–700, 1994.
153. L. Lucchese and S. Mitra. Color image segmentation: a state-of-the-art survey. In *Proceedings of the Indian National Science Academy*, volume 67-A-2, pages 207–221, 2001.
154. Q.-T. Luong, P. Fua, and Y. G. Leclerc. The radiometry of multiple images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):19–33, 2002.
155. J. Malik, S. Belongie, T. Leung, and J. Shi. Contour and texture analysis for image segmentation. *International Journal of Computer Vision*, 43(1):7–27, 2001.
156. L. Maloney and B. Wandell. Color constancy: a method for recovering surface spectral reflectance. *Journal of the Optical Society of America - A*, 3(1):29–33, 1986.
157. N. Max. Atmospheric illumination and shadows. *ACM SIGGRAPH Computer Graphics*, 20(4):117–124, 1986.
158. B. A. Maxwell. *Segmentation and Interpretation Using Multiple Physical Hypotheses of Image Formation*. PhD thesis, Robotics Institute, Carnegie Mellon University, 1996. (Also published as tech. report CMU-RI-TR-96-28).
159. B. A. Maxwell. A physics-based approach to interactive segmentation. In *Proceedings of the Visual Information and Information Systems (VISUAL) conference*, pages 517–524, 1999.

160. B. A. Maxwell and S. A. Shafer. A framework for segmentation using physical models of image formation. Technical Report CMU-RI-TR-93-29, Robotics Institute, Carnegie Mellon University, December 1993.
161. B. A. Maxwell and S. A. Shafer. A framework for segmentation using physical models of image formation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 361–368, 1994.
162. B. A. Maxwell and S. A. Shafer. A framework for segmentation using physical models of image formation. In *Proceedings of ARPA*, volume II, pages 1513–1522, 1994.
163. B. A. Maxwell and S. A. Shafer. Physics-based segmentation: looking beyond color. Technical Report CMU-RI-TR-95-37, Robotics Institute, Carnegie Mellon University, October 1995.
164. B. A. Maxwell and S. A. Shafer. Physics-based segmentation: looking beyond color. In *Proceedings of the Image Understanding Workshop*, 1996.
165. B. A. Maxwell and S. A. Shafer. Physics-based segmentation: moving beyond color. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 742–749, 1996.
166. B. A. Maxwell and S. A. Shafer. Physics-based segmentation of complex objects using multiple hypotheses of image formation. *Computer Vision and Image Understanding*, 65(2):269–295, 1997.
167. B. A. Maxwell and S. A. Shafer. Segmentation and interpretation of multicolored objects with highlights. *Computer Vision and Image Understanding*, 77(1):1–24, 2000.
168. P. Meer and B. Georgescu. Edge detection with embedded confidence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(12):1351–1365, 2001. (The code is publicly available at http://www.caip.rutgers.edu/riul/research/code.html).
169. V. Mezaris, I. Kompatsiaris, and M. Strintzis. Still image objective segmentation evaluation using ground truth. In B. Kovář, J. Přikryl, and M. Vlček, editors, *Proceedings of the 5th COST 276 Workshop on Information and Knowledge Management for Integrated Media Communication*, pages 9–14, 2003.
170. Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer, 1996.
171. Z. Michalewicz and D. B. Fogel. *How to Solve it: Modern Heuristics*. Springer, 2000.
172. X. Muñoz. *Image segmentation integrating colour, texture and boundary information*. PhD thesis, Universitat de Girona, Department of Electronics, Informatics and Automation, 2002.
173. X. Muñoz, J. Freixenet, X. Cufí, and J. Martí. Strategies for image segmentation combining region and boundary information. *Pattern Recognition Letters*, 24(1-3):383–400, 2003.
174. S. K. Nayar and R. M. Bolle. Computing reflectance ratios from an image. *Pattern Recognition*, 26(10):1529–1542, 1993.
175. S. K. Nayar and R. M. Bolle. Object recognition based on reflectance and geometry. *Proceedings of SPIE*, 2031:328–342, 1993.
176. S. K. Nayar and R. M. Bolle. Reflectance based recognition. In *Proceedings of DARPA*, pages 867–872, 1993.
177. S. K. Nayar and R. M. Bolle. Reflectance ratio: a photometric invariant for object recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 280–285, 1993.
178. S. K. Nayar and R. M. Bolle. Reflectance based object recognition. *International Journal of Computer Vision*, 17(3):219–240, 1996.
179. S. K. Nayar, X.-S. Fang, and T. Boult. Separation of reflection components using color and polarization. *International Journal of Computer Vision*, 21(3):163–186, 1997.
180. S. K. Nayar, K. Ikeuchi, and T. Kanade. Surface reflection: physical and geometrical perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7), July 1991.
181. S. Negahdaripour and C. H. Yu. On shape and range recovery from image shading for underwater applications. In J. Yuh, editor, *Underwater robotic vehicles: design and control*, chapter 8, pages 221–250. TSI Press, 1995.
182. R. Nevatia. Machine perception. In *Prentice-Hall*, 1982.
183. R. Nevatia. Image segmentation. In *Handbook of Pattern Recognition and Image Processing*, volume 86, pages 215–231, 1986.
184. F. Nicodemus, J. Richmond, J. Hsia, I. Ginsberg, and T. Limperis. *Geometrical considerations and nomenclature for reflectance*, volume 160 of *NBS Monograph*. National Bureau of Standards, October 1977.
185. T. Nishita and E. Nakamae. A shading model for atmospheric scattering considering the luminous intensity distribution of light sources. *ACM SIGGRAPH Computer Graphics*, 21(4):303–308, 1987.
186. C. L. Novak and S. A. Shafer. Anatomy of a color histogram. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 599–605, 1992.
187. C. L. Novak and S. A. Shafer. Color vision. In *Encyclopedia of Artificial Intelligence*, pages 192–202. J. Wiley and Sons, 1992.
188. C. L. Novak, S. A. Shafer, and R. G. Willson. Obtaining accurate color images for machine vision research. In *Proceedings of SPIE Perceiving, Measuring and Using Color*, volume 1250, pages 54–68, 1990.

189. C. Odet, B. Belaroussi, and H. Cattin. Scalable discrepancy measures for segmentation evaluation. In *Proceedings of the IEEE International Conference on Image Processing*, volume I, pages 785–788, 2002.
190. J. Oliensis. Shape from shading as a partially well-constrained problem. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 54(2):163–183, 1991.
191. J. Oliensis. Uniqueness in shape from shading. *International Journal of Computer Vision*, 6(2):75–104, 1991.
192. J. Oliensis and P. Dupuis. Direct method for reconstructing shape from shading. In *Proceedings of SPIE Geometric Methods in Computer Vision*, volume 1570, 1991.
193. J. Oliensis and P. Dupuis. A global algorithm for shape from shading. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 692–701, 1993.
194. C.-K. Ong and T. Matsuyama. Robust color segmentation using the dichromatic reflection model. In *Proceedings of the IEEE International Conference on Pattern Recognition*, pages 780–784, 1998.
195. C.-K. Ong and T. Matsuyama. Robust color segmentation using the dichromatic reflection model. *Information Processing Society of Japan Journal*, 40(8):3259–3266, 1999.
196. C. M. Onyango and J. A. Marchant. Physics-based colour image segmentation for scenes containing vegetation and soil. *Image and Vision Computing*, 19(8):523–538, 2001.
197. M. Oren and S. K. Nayar. Generalization of the Lambertian model and implications for machine vision. *International Journal of Computer Vision*, 14(3):227–251, 1995.
198. A. Ortiz. Aplicación de técnicas de visión por computador a entornos submarinos. Technical report, Departament de Matemàtiques i Informàtica. Universitat de les Illes Balears, 1998.
199. A. Ortiz and G. Oliver. A physics-based image segmentation technique for Lambertian scenes. Technical Report 2/99, Departament de Matemàtiques i Informàtica (Universitat de les Illes Balears), 1999.
200. A. Ortiz and G. Oliver. An adaptation for SFS algorithms to deal with multiple albedo images. Technical Report 1/2000, Departament de Matemàtiques i Informàtica (Universitat de les Illes Balears), 2000.
201. A. Ortiz and G. Oliver. Shape from Shading for Multiple Albedo Images. In *Proceedings of the IEEE International Conference on Pattern Recognition*, 2000.
202. A. Ortiz and G. Oliver. Albedo-based segmentation prior to estimating shape from shading in multi-albedo scenes. Technical Report 1/2002, Departament de Matemàtiques i Informàtica (Universitat de les Illes Balears), 2002.
203. A. Ortiz and G. Oliver. Image segmentation through physics-based quality maps. Technical Report 2/2002, Departament de Matemàtiques i Informàtica (Universitat de les Illes Balears), 2002.
204. A. Ortiz and G. Oliver. A method to discount ambient lighting and albedo variation before estimating shape from image shading. In *Proceedings of Advanced Concepts for Intelligent Vision Systems*, 2002.
205. A. Ortiz and G. Oliver. A qualitative approach to physics-based image segmentation. In *Proceedings of Advanced Concepts for Intelligent Vision Systems*, 2002.
206. A. Ortiz and G. Oliver. Detection of colour channels uncoupling for curvature-insensitive segmentation. In F. Perales, A. Campilho, N. P. de la Blanca, and A. Sanfeliu, editors, *Proceedings of the IAPR Iberian Conference on Pattern Recognition and Image Analysis*, volume 2652 of *Lecture Notes in Computer Science*, pages 664–672. Springer-Verlag, 2003.
207. A. Ortiz and G. Oliver. Estimation of scene lighting parameters and camera dark current. In I. Aguiló, L. c Valverde, and M. T. Escrig, editors, *Proceedings of the Catalan Conference on Artificial Intelligence*, volume 100 of *Artificial Intelligence Research and Development. Frontiers in Artificial Intelligence and Applications*, pages 199–210. IOS Press, 2003.
208. A. Ortiz and G. Oliver. Scene lighting parameters estimation and radiometric camera calibration. Technical Report A-2-2003, Departament de Matemàtiques i Informàtica (Universitat de les Illes Balears), 2003.
209. A. Ortiz and G. Oliver. Segmentation of images based on the detection of reflectance transitions. Technical Report A-3-2003, Departament de Matemàtiques i Informàtica (Universitat de les Illes Balears), 2003.
210. A. Ortiz and G. Oliver. Estimation of directional and ambient illumination parameters by means of a calibration object. In A. Campilho and M. Kamel, editors, *Proceedings of the IAPR International Conference on Image Analysis and Recognition*, volume 3211 of *Lecture Notes in Computer Science*, pages 672–679. Springer-Verlag, 2004.
211. A. Ortiz and G. Oliver. A physics-based colour edge detector. In J. Vitrià, P. Radeva, and I. Aguiló, editors, *Proceedings of the Catalan Conference on Artificial Intelligence*, volume 113 of *Artificial Intelligence Research and Development. Frontiers in Artificial Intelligence and Applications*, pages 201–208. IOS Press, 2004.
212. A. Ortiz and G. Oliver. Radiometric calibration of CCD sensors: dark current and fixed pattern noise estimation. In *Proceedings of the IEEE International Conference on Robotics and Automation*, volume 5, pages 4730–4735, 2004.

213. A. Ortiz and G. Oliver. Estimation of intensity uncertainties for computer vision applications. In J. Blanc-Talon, D. Popescu, W. Philips, and P. Scheunders, editors, *Proceedings of Advanced Concepts for Intelligent Vision Systems*, volume 3708 of *Lecture Notes in Computer Science*, pages 178–185. Springer Verlag, 2005.
214. A. Ortiz and G. Oliver. A novel segmentation strategy based on colour channels coupling. In F. Roli and S. Vitulano, editors, *Proceedings of the IAPR International Conference on Image Analysis and Processing*, volume 3617 of *Lecture Notes in Computer Science*, pages 328–335. Springer Verlag, 2005.
215. A. Ortiz and G. Oliver. R^2CIU: Robust Radiometric Calibration and Intensity Uncertainty estimation. Technical Report A-1-2005, Departament de Matemàtiques i Informàtica (Universitat de les Illes Balears), 2005.
216. N. Otsu. A threshold selection method from gray-level histogram. *IEEE Transactions on Systems, Man and Cybernetics*, 9:62–66, 1979.
217. G. Pajares and J. M. de la Cruz. Formas a partir de X. *Revista Electrónica de Visión por Computador*, 1(3), Julio 2000.
218. N. R. Pal and D. Bhandari. Image thresholding: some new techniques. *Signal Processing*, 33:139–158, 1993.
219. N. R. Pal and S. K. Pal. Entropic thresholding. *Signal Processing*, 16:97–108, 1989.
220. N. R. Pal and S. K. Pal. A review on image segmentation techniques. *Pattern Recognition*, 26(9):1277–1294, 1993.
221. A. W. Palowitch and J. S. Jaffe. Computer model for predicting underwater color images. In *Proceedings SPIE Underwater Imaging, Photography and Visibility*, volume 1537, pages 128–139, 1991.
222. J. B. Park. Efficient color representation for image segmentation under non-white illumination. In *Proceedings of SPIE - Intelligent Robots and Computer Vision XXI: Algorithms, Techniques, and Active Vision*, volume 5267, 2003.
223. J. B. Park and A. C. Kak. A truncated least squares approach to the detection of specular highlights in color images. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1397–1403, 2003.
224. A. P. Pentland. Local shading analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(2):170–187, 1984.
225. A. P. Pentland. Shape information from shading: a theory about human perception. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 404–413, 1988.
226. A. P. Pentland. Local shading analysis. In B. K. P. Horn and M. J. Brooks, editors, *Shape from Shading*, chapter 15, pages 443–487. The MIT Press, 1989.
227. F. Perez and C. Koch. Toward color image segmentation in analog VLSI: algorithm and hardware. *International Journal of Computer Vision*, 12(1):17–42, 1994.
228. B. Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18:311–317, 1975.
229. G. Pignalberi, R. Cucchiara, L. Cinque, and S. Levialdi. Tuning range image segmentation by genetic algorithm. *EURASIP Journal on Applied Signal Processing*, 2003(8):780–790, 2003.
230. M. Powell, S. Sarkar, and D. Goldgof. A simple strategy for calibrating the geometry of light sources. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9):1022–1027, 2001.
231. W. Pratt. *Digital Image Processing*. John Wiley and Sons, 1978.
232. W. Pratt. *Digital Image Processing*. John Wiley and Sons, 2nd edition, 1991.
233. V. Ramesh and R. Haralick. Performance characterization of edge detectors. In *Proceedings of SPIE Applications of AI X*, volume 1708, pages 252–266, 1992.
234. G. Rees, W. Wright, and P. Greenway. ROC method for the evaluation of multi-class segmentation/classification algorithms with infrared imagery. In *Proceedings of the British Machine Vision Conference*, 2002. (Poster session).
235. J. Reichman. Determination of absorption and scattering coefficients for nonhomogeneous media. 1: Theory. *Applied Optics*, 12:1811–1815, 1973.
236. M. Reichmann. Understanding raw files. http://www.luminous-landscape.com/tutorials/understanding-series/u-raw-files.shtml, 2005.
237. E. Riseman and M. Arbib. Computational techniques in the visual segmentation of static scenes. *Computer Vision, Graphics, and Image Processing*, 6(3):221–276, 1977.
238. L. Roberts. Machine perception of three-dimensional solids. In J. Tippett et al., editors, *Optical and Electro-Optical Information Processing*, pages 159–197. MIT Press, 1965.
239. R. Roman-Roldan, J. Gomez-Lopera, C. Atae-Allah, J. Martinez-Aroza, and P. Luque-Escamilla. A measure of quality for evaluating methods of segmentation and edge detection. *Pattern Recognition*, 34(5):969–980, May 2001.

240. A. Rosenfeld and L. S. Davis. Image segmentation and image models. *Proceedings of the IEEE*, 67(5), May 1979.
241. P. Rousseeuw and A. Leroy. *Robust Regression and Outlier Detection*. Wiley, 1987.
242. E. Rouy and A. Tourin. A viscosity solutions approach to shape-from-shading. *SIAM Journal on Numerical Analysis*, 29(3):867–884, 1992.
243. H. E. Rushmeier and K. E. Torrance. The zonal method for calculating light intensities in the presence of a participating medium. *ACM SIGGRAPH Computer Graphics*, 21(4):293–302, 1987.
244. P. Sahoo, S. Soltani, A. Wong, and Y. Chen. A survey on thresholding techniques. *Computer Vision, Graphics, and Image Processing*, 41:233–260, 1988.
245. E. Salvador, A. Cavallaro, and T. Ebrahimi. Shadow identification and classification using invariant color models. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 1545–1548, 2001.
246. E. Salvador, A. Cavallaro, and T. Ebrahimi. Cast shadow segmentation using invariant color features. *Computer Vision and Image Understanding*, 95(2):238–259, 2004.
247. D. Samaras and D. Metaxas. Incorporating illumination constraints in deformable models for shape from shading and light direction estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):247–264, 2003.
248. I. Sato, Y. Sato, and K. Ikeuchi. Illumination from shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(3):290–300, 2003.
249. K. Schlüns and M. Teschner. Analysis of 2D color spaces for highlight elimination in 3D shape reconstruction. In *Proceedings of the Asian Conference on Computer Vision*, volume II, pages 801–805, 1995.
250. K. Schlüns and M. Teschner. Fast separation of reflection components and its application in 3D shape recovery. In *Proceedings of the IS&T/SID Color Imaging Conference*, pages 48–51, 1995.
251. S. M. Selby, editor. *Standard Mathematical Tables*. CRC Press, 1975.
252. S. A. Shafer. Using color to separate reflection components. *COLOR Research and Application*, 10(4):210–218, 1985.
253. S. A. Shafer. (introduction to the chapter). In G. E. Healey, S. A. Shafer, and L. B. Wolff, editors, *Physics-based vision: principles and practice*, volume on Color, chapter Color interreflection, pages 349–350. Jones and Bartlett Publishers, 1992.
254. S. A. Shafer, T. Kanade, and G. J. Klinker. Physics-based models for early vision by machine. In *Proceedings of SPIE - Perceiving, Measuring, and Using Color*, volume 1250, pages 222–235, 1990.
255. S. A. Shafer, T. Kanade, G. J. Klinker, and C. L. Novak. Physical-based models for early vision by machine (reprint). In G. E. Healey, S. A. Shafer, and L. B. Wolff, editors, *Physics-based vision: principles and practice*, volume on Color, pages 351–364. Jones and Bartlett Publishers, 1992.
256. M. Shin, D. Goldgof, and K. Bowyer. Comparison of edge detector performance through use in an object recognition task. *Computer Vision and Image Understanding*, 84(1):160–178, 2001.
257. W. Skarbek and A. Koschan. Colour image segmentation — a survey. Technical Report 94-32, Institute for Technical Informatics, Technical University of Berlin, October 1994.
258. Sony Co. 1/2-inch Progressive Scan CCD Image Sensor ICX075AK: Technical Specifications. (E95706D72-PS).
259. B. Southall, B. Buxton, J. Marchant, and T. Hague. On the performance characterisation of image segmentation algorithms: a case study. In *Proceedings of the European Conference on Computer Vision*, volume II, pages 351–365, 2000.
260. L. Spirkovska. A summary of image segmentation techniques. NASA Technical Memorandum 104022, June 1993.
261. S. Stehman. Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, 62(1):77–89, 1997.
262. H. Stokman. *Robust Photometric Invariance in Machine Colour Vision*. PhD thesis, Intelligent Sensory Information Systems Research Group - Faculty of Science (University of Amsterdam), 2000.
263. H. Stokman and T. Gevers. Detection and classification of hyper-spectral edges. In *Proceedings of the British Machine Vision Conference*, pages 643–651, 1999.
264. H. Stokman and T. Gevers. Photometric invariant region detection in multi-spectral images. In *Proceedings of Vision Inferface*, pages 90–96, 1999.
265. H. Stokman and T. Gevers. Automatic thresholding of color edges. In *Proceedings of the Color Imaging Conference*, pages 178–183, 2000.
266. H. Stokman and T. Gevers. Error propagation due to color space transforms. In *Proceedings of SPIE*, 2000.
267. K. Strasters and J. Gerbrands. Three-dimensional image segmentation using a split, merge and group approach. *Pattern Recognition Letters*, 12:307–325, 1991.

268. R. T. Tan, K. Nishino, and K. Ikeuchi. Reflection components separation based on chromaticity and noise analysis. In *Proceedings of the IEEE International Workshop on Color and Photometric Methods in Computer Vision*, 2003.
269. J.-P. Tarel. Calibration radiométrique de caméra. Technical Report 2509, INRIA, mars 1995.
270. J.-P. Tarel. Une méthode de calibration radiométrique de caméra à focale variable. In *Proceedings of 10ème Congrés AFCET, Reconnaissance des Formes et Intelligence Artificielle, RFIA*, 1996.
271. G. Taubin. Estimation of planar curves, surfaces and nonplanar space curves defined by implicit equations with aplications to edge and range image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(11):1115–1138, 1991.
272. J. R. Taylor. *An Introduction to Error Analysis*. University Science Books, 2nd edition, 1997.
273. J. Tenenbaum, M. Fischler, and H. Barrow. Scene modeling: a structural basis for image description. *Computer Graphics and Image Processing*, 12:407–425, 1980.
274. The Imaging Source. *Background info: introduction to cameras*, 2001. [Notes on the web http://theimagingsource.com].
275. The Imaging Source. *Background info: introduction to lenses*, 2001. [Notes on the web http://theimagingsource.com].
276. S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, 1999.
277. Y. Tian, H. Tsui, and S. Yeung. Shape recovery from one image under multiple light sources. In *Proceedings of the Asian Conference on Computer Vision*, pages 216–223, 1998.
278. Y. Tian, H. Tsui, and S. Yeung. Spherical and cylindrical light source models for shape recovery. In *Proceedings of the Asian Conference on Computer Vision*, pages 224–231, 1998.
279. Y. Tian, H. Tsui, S. Yeung, and S. Ma. Shape from shading for multiple light sources. *Journal of the Optical Society of America - A*, 16(1):36–52, 1999.
280. S. Tominaga and B. Wandell. Standard surface-reflectance model and illuminant estimation. *Journal of the Optical Society of America - A*, 6(4):576–584, 1989.
281. K. Torrance and E. Sparrow. Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America*, 57:1105–1114, 1967.
282. P. Trahanias and A. Venetsanopoulos. Color edge detection using vector order statistics. *IEEE Transactions on Image Processing*, 2(2):259–264, April 1993.
283. P.-S. Tsai and M. Shah. A simple shape from shading algorithm. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 734–736, 1992.
284. P. W. M. Tsang and W. H. Tsang. Edge detection on object color. In *Proceedings of the IEEE International Conference on Image Processing*, volume III, pages 1049–1052, 1996.
285. W. H. Tsang and P. W. M. Tsang. Suppression of false edge detection due to specular reflection in color images. *Pattern Recognition Letters*, 18(2):165–171, 1997.
286. S. Ullman. *High-level vision: object recognition and visual cognition*. MIT Press, 1997.
287. S. Umeyama and G. Godin. Separation of diffuse and specular components of surface reflection by use of polarization and statistical analysis of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):639–647, 2004.
288. D. Vernon. Machine vision: automated visual inspection and robot vision. *Prentice Hall*, 1991.
289. Y. Wang and D. Samaras. Estimation of multiple illuminants from a single image of arbitrary known geometry. In *Proceedings of the European Conference on Computer Vision*, pages 273–288, 2002.
290. Z. Wang, K. Rao, and J. Ben-Arie. Optimal ramp edge-detection using expansion matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(11):1092–1097, November 1996.
291. J. Weijer, T. Gevers, and J.-M. Geusebroek. Color edge detection by photometric quasi-invariants. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1520–1525, 2003.
292. J. Weijer, T. Gevers, and J.-M. Geusebroek. Edge and corner detection by photometric quasi-invariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):625–630, 2005.
293. J. Weszka and A. Rosenfeld. Threshold evaluation techniques. *IEEE Transactions on Systems, Man and Cybernetics*, 8:622–629, 1978.
294. L. B. Wolff. A diffuse reflectance model for smooth dielectrics. *Journal of the Optical Society of America - A*, 11(11):2956–2968, 1994. (special issue on Physics Based Machine Vision).
295. L. B. Wolff, S. K. Nayar, and M. Oren. Improved diffuse reflection models for computer vision. *International Journal of Computer Vision*, 30(1):55–71, 1998.
296. W. Yasnoff and J. Bacus. Scene segmentation algorithm development using error measures. *Analytical and Quantitative Cytology*, 6:45–58, 1984.
297. W. Yasnoff, J. Mui, and J. Bacus. Error measures for scene segmentation. *Pattern Recognition*, 9:217–231, 1977.

298. Y. Yitzhaky and E. Peli. A method for objective edge detection evaluation and detector parameter selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):1027–1033, 2003.
299. Y. Yu, P. Debevec, J. Malik, and T. Hawkins. Inverse global illumination: recovering reflectance models of real scenes from photographs from. In *Siggraph99, Annual Conference Series*, pages 215–224, 1999.
300. R. Zhang and M. Shah. Height recovery from intensity gradient. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 508–513, 1994.
301. R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah. Analysis of shape from shading techniques. Technical report, Computer Science Department. University of Central Florida, 1994.
302. R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah. Analysis of shape from shading techniques. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1994.
303. R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah. Shape from shading: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):690–706, 1999.
304. Y. Zhang. Segmentation evaluation and comparision: a study of various algorithms. In B. G. Haskell and H.-M. Hang, editors, *Proceedings of SPIE Visual Communication and Image Processing*, volume 2094, pages 801–812, 1993.
305. Y. Zhang. Influence of image segmentation over feature measurement. *Pattern Recognition Letters*, 16:201–206, 1995.
306. Y. Zhang. A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8):1335–1346, 1996.
307. Y. Zhang and J. Gerbrands. Segmentation evaluation using ultimate measurement accuracy. In *Proceedings of SPIE*, volume 1657, pages 449–460, 1992.
308. Y. Zhang and J. Gerbrands. Objective and quantitative segmentation evaluation and comparison. *EURASIP Journal on Applied Signal Processing*, 39:43–54, 1994.
309. Y. Zhang and Y.-H. Yang. Multiple illuminant direction detection with application to image synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(8):915–920, 2001.
310. Z. Zhang. Parameter estimation techniques: a tutorial with application to conic fitting. Technical Report 2676, INRIA, October 1995.
311. Q. Zheng and R. Chellappa. Estimation of illuminant direction, albedo and shape from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):680–702, 1991.
312. W. Zhou and C. Kambhamettu. Estimation of illuminant direction and intensity of multiple light sources. In *Proceedings of the European Conference on Computer Vision*, pages 206–220, 2002.
313. S. C. Zhu and A. Yuille. Region competition: unifying snakes, region growing, and bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900, 1996.
314. S. Zucker. Algorithms for image segmentation. In *Digital Image Processing and Analysis*, pages 169–183, 1977.

Index

1/f noise, 30

absorption coefficient, 25
accuracy, 67
albedo
  diffuse, 16
ambient composite reflectance, 61
ambient lighting, 20
ambient reflectance, 21
amplifier noise, 30
angle of exitance, 19
angle of incidence, 19
Approximate Color Reflectance Model (ACRM), 20

Baddeley's measure, 77
Bayer colour filter array, 26
beam-splitter prism, 26
Beckmann-Spizzichino model, 17
bidirectional reflectance distribution function (BRDF), 11
bidirectional spectral-reflectance distribution function (BSRDF), 11
blooming, 32
body composite reflectance, 61
body reflectance, 16
body reflection, 12, 14
  Lambert model, 14
  Oren and Nayar model, 16
  Wolff model, 16

cameras
  Bayer colour filter array, 26
  beam-splitter prism, 26
  colour filter array, 26
  complementary colour mosaic, 27
  demosaicking, 27
  F-number, 23
  filter transmittance, 22
  gamma correction, 32
  integration *or* exposure time, 25, 28
  primary colour mosaic, 27
  spatial response, 28
  spectral responsivity, 22
  vignetting, 31
Charge Coupled Device (CCD), 24
  1/f noise *or* flicker noise, 30
  absorption coefficient, 25
  amplifier noise, 30
  blooming, 32
  Charge Transfer Efficiency (CTE), 25
  clipping, 32
  correlated double sampling, 30
  dark current, 28
  dark current non-uniformity (DCNU), 28
  dark current shot noise, 29
  dark *or* thermal electrons, 28
  fixed pattern noise, 28
  Frame Interline Transfer (FIT) CCD, 26
  Frame Transfer (FT) CCD, 26
  Full Frame (FF), 25
  infrared cut-off filter, 32
  Interline Transfer (IL) CCD, 26
  multi-pinned phase technology, 28
  non-spatial *or* random noise, 64
  pattern noise, 28
  photo-response non-uniformity (PRNU), 28
  photoelectrons, 25
  photon *or* shot noise, 29
  pre-kneeing, 32
  quantization noise, 30
  quantum efficiency, 25
  readout noise, 30
  recombination lifetime, 25
  reset noise *or* kTC noise, 30
  shading, 28
  spatial noise, 64
  thermoelectric coolers, 28
Charge Transfer Efficiency (CTE), 25
clipping, 32
colour filter array, 26
complementary colour mosaic, 27
composite reflectance
  ambient, 61
  body, 61
  interface, 61
confusion matrix, 66
contingency matrix, 67

correlated double sampling, 30

dark current, 28
dark current non-uniformity (DCNU), 28
dark current shot noise, 29
dark electrons, 28
demosaicking, 27
Dichromatic Reflection Model (DRM), 19
diffuse albedo, 16
Discrepancy Percentage, 77

error matrix, 67
exposure time, 25, 28

F-number, 23
false negative, 67
false positive, 67
filter transmittance, 22
fixed pattern noise, 28
flicker noise, 30
foreshortened area, 10
Frame Interline Transfer (FIT) CCD, 26
Frame Transfer (FT) CCD, 26
Fresnel term, 18
Full Frame (FF) CCD, 25

gamma correction, 32
generalized confusion matrix, 67

Healey model, 19
hybrid reflection, 19
  Approximate Color Reflectance Model (ACRM), 20
  Dichromatic Reflection Model (DRM), 19

infrared cut-off filter, 32
integrated white condition, 45, 48–53
integration time, 25, 28
inter-reflection, 21
interface composite reflectance, 61
interface reflectance, 13
interface reflection, 12, 13
  Beckmann-Spizzichino model, 17
  Healey model, 19
  Neutral Interface Reflection Model (NIR), 20
  Phong model, 19
  Torrance-Sparrow model, 17
  Unichromatic Reflection Model for metals (URM), 20
Interline Transfer (IL) CCD, 26
irradiance, 10

kTC noise, 30
Kubelka-Munk theory, 20

Lambert Cosine Law, 15
Lambert model, 14
Lambertian surface
  ideal, 14
lighting
  ambient, 20

material
  dielectric, 12, 13
  metal, 12
  optically homogeneous, 12
  optically inhomogeneous, 12
  skin depth, 12
mean error distance, 77
mean square error distance, 77
multi-class type I error, 67
multi-class type II error, 67
multi-pinned phase technology, 28

Neutral Interface Reflection Model (NIR), 20
non-spatial noise, 64

Oren and Nayar model, 16
over-segmentation, 68

pattern noise, 28
percentage of correctly grouped pixels (CG), 69
percentage of over-segmentation (OS), 69
percentage of under-segmentation (US), 69
perceptual colour, 62
performance evaluation
  analytical methods, 65
  edge detection evaluation
    Baddeley's measure, 77
    Discrepancy Percentage, 77
    mean error distance, 77
    mean square error distance, 77
    Pratt's Figure of Merit (FOM), 77
  empirical discrepancy methods, 65
    accuracy, 67
    confusion matrix, 66
    contingency matrix, 67
    error matrix, 67
    false negative, 67
    false positive, 67
    generalized confusion matrix, 67
    multi-class type I error, 67
    multi-class type II error, 67
    Receiver Operating Characteristic (ROC) curve analysis, 68
    type I error, 67
    type I error rate, 67
    type II error, 67
    type II error rate, 67
  empirical goodness methods, 65
  empirical methods, 65
  over-segmentation, 68
  percentage of correctly grouped pixels (CG), 69
  percentage of over-segmentation (OS), 69
  percentage of under-segmentation (US), 69
  under-segmentation, 68
phase angle, 19
Phong model, 19
photoelectrons, 25

photometric angles, 19
  angle of exitance, 19
  angle of incidence, 19
  phase angle, 19
photon noise, 29
Pratt's Figure of Merit (FOM), 77
pre-kneeing, 32
primary colour mosaic, 27

quantization noise, 30
quantum efficiency, 25

radiance, 10
  spectral, 11
radiant exitance, 10
radiant flux, 10
radiant intensity, 10
radiometric calibration, 64
random noise, 64
readout noise, 30
Receiver Operating Characteristic (ROC) curve analysis, 68
recombination lifetime, 25
reflectance
  ambient, 21
  body, 16
  interface, 13
reflectance map, 14
reflection
  body, 12, 14
  diffuse, 12
  inter-reflection, 21
  interface, 12, 13
    ideal specular reflector, 17
  models, 14
    Approximate Color Reflectance Model (ACRM), 20
    Beckmann-Spizzichino, 17
    Dichromatic Reflection Model (DRM), 19
    Fresnel term, 18
    Healey, 19
    hybrid, 19
    ideal Lambertian surface, 14
    integrated white condition, 45, 48–53
    interface, 17
    Lambert, 14
    Lambert Cosine Law, 15
    Neutral Interface Reflection Model (NIR), 20
    Oren and Nayar, 16
    Phong, 19
    Torrance-Sparrow, 17
    Unichromatic Reflection Model for metals (URM), 20
    Wolff, 16
  photometric angles, 19
    angle of exitance, 19
    angle of incidence, 19
    phase angle, 19
  reflectance map, 14
  specular, 12
  surface roughness, 18
refraction
  complex index, 12
Reichman body-scattering model, 20
reset noise, 30

scene noise, 78
segmentation methods
  general
    edge- or contour-based, 35
    pixel- or feature-space-based, 36
    region- or area-based, 35
  physics-based
    cluster-analysis-based, 38
    continuity-of-the-image-surface-based, 38
    photometric-invariant-based, 38
    reflectance-estimation-based, 38
shading, 28
Shape from Motion, 139
Shape from Photometric Stereo, 139
Shape from Shading (SFS), 139, 140
  domain of attraction of an LMSP, 146
  global minimization methods, 145
  global shape reconstruction algorithm, 146
  linear methods, 145
  local methods, 145
  Local Minimum/Maximum Singular Point (LMSP), 146
  local shape reconstruction algorithm, 146
  propagation methods, 145
  singular point, 144
Shape from Stereo, 139
Shape from Texture, 139
Shape from X, 139
  Shape from Motion, 139
  Shape from Photometric Stereo, 139
  Shape from Shading (SFS), 139, 140
  Shape from Stereo, 139
  Shape from Texture, 139
shot noise, 29
singular point, 144
skin depth, 12
solid angle, 9
spatial noise, 64
spatial response, 28
spectral power distribution (SPD), 11
  relative, 11
spectral responsivity, 22
specular reflector
  ideal, 17
surface roughness, 18

thermal electrons, 28
thermoelectric coolers, 28
Torrance-Sparrow model, 17
type I error, 67
type I error rate, 67
type II error, 67

type II error rate, 67

under-segmentation, 68
Unichromatic Reflection Model for metals (URM), 20

vignetting, 31

Wolff model, 16