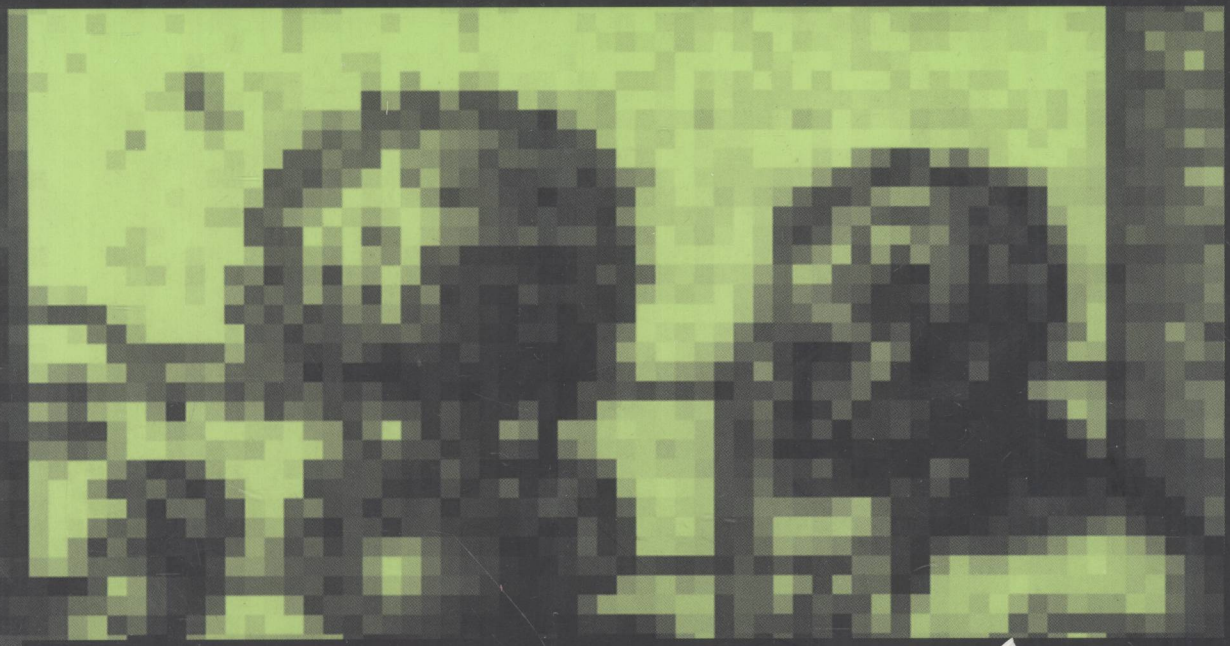

THREE-DIMENSIONAL COMPUTER VISION

A GEOMETRIC VIEWPOINT

OLIVIER FAUGERAS

Olivier Faugeras

Three-Dimensional Computer Vision

A Geometric Viewpoint

The MIT Press
Cambridge, Massachusetts
London, England

This book was formatted in ZzTEX by Paul Anagnostopoulos and typeset by Joe Snowden. The typeface is Lucida Bright and Lucida New Math, created by Charles Bigelow and Kris Holmes specifically for scientific and electronic publishing. The Lucida letterforms have the large x-heights and open interiors that aid legibility in modern printing technology, but also echo some of the rhythms and calligraphic details of lively Renaissance handwriting. Developed in the 1980s and 1990s, the extensive Lucida typeface family includes a wide variety of mathematical and technical symbols designed to harmonize with the text faces.

Library of Congress Cataloging-in-Publication Data

Faugeras, Olivier, 1949-
Three-dimensional computer vision : a geometric viewpoint /
Olivier Faugeras.
p. cm. — (Artificial intelligence)
Include bibliographical references and index.
ISBN 0-262-06158-9
1. Computer vision. I. Title. II. Series: Artificial
intelligence (Cambridge, Mass.)
TA1632.F38 1993
006.3'7—dc20 93-9126
CIP

Three-Dimensional Computer Vision

Series Foreword

Artificial intelligence is the study of intelligence using the ideas and methods of computation. Unfortunately a definition of intelligence seems impossible at the moment because intelligence appears to be an amalgam of so many information-processing and information-representation abilities.

Of course psychology, philosophy, linguistics, and related disciplines offer various perspectives and methodologies for studying intelligence. For the most part, however, the theories proposed in these fields are too incomplete and too vaguely stated to be realized in computational terms. Something more is needed, even though valuable ideas, relationships, and constraints can be gleaned from traditional studies of what are, after all, impressive existence proofs that intelligence is in fact possible.

Artificial intelligence offers a new perspective and a new methodology. Its central goal is to make computers intelligent, both to make them more useful and to understand the principles that make intelligence possible. That intelligent computers will be extremely useful is obvious. The more profound point is that artificial intelligence aims to understand intelligence using the ideas and methods of computation, thus offering a radically new and different basis for theory formation. Most of the people doing work in artificial intelligence believe that these theories will apply to any intelligent information processor, whether biological or solid state.

There are side effects that deserve attention, too. Any program that will successfully model even a small part of intelligence will be inherently massive and complex. Consequently artificial intelligence continually confronts the limits of computer-science technology. The problem

encountered have been hard enough and interesting enough to seduce artificial intelligence people into working on them with enthusiasm. It is natural, then, that there has been a steady flow of ideas from artificial intelligence to computer science, and the flow shows no sign of abating.

The purpose of this series in artificial intelligence is to provide people in many areas, both professionals and students, with timely, detailed information about what is happening on the frontiers in research centers all over the world.

J. Michael Brady
Daniel G. Bobrow
Randall Davis

Figures

2.1 Cross-ratio of four lines: $\{l_1, l_2; l_3, l_4\} = \{P_1, P_2; P_3, P_4\}$. 17

2.2 The angle α between l_1 and l_2 is given by the Laguerre formula: $\alpha = \frac{1}{2i}\log(\{l_1, l_2; i_a, j_a\})$. 22

2.3 Cross-ratio of four planes: $\{\pi_1, \pi_2; \pi_3, \pi_4\} = \{l_1, l_2; l_3, l_4\}$. 26

2.4 The angle between the two planes π_1 and π_2 is a simple function of the cross-ratio $(\pi_1, \pi_2; i_l, j_l)$, where i_l and j_l are the two tangent planes from l to the absolute conic Ω. 30

3.1 Image formation in a pinhole camera. 34

3.2 The pinhole camera model. 35

3.3 The optical axis, focal plane, and retinal plane. 36

3.4 The relationship between image irradiance and scene radiance, which is linear. 37

3.5 The focal plane (x, y) is parallel to the retinal plane (u, v) and at a distance f from it; f is the focal length. 38

3.6 When M gets closer to the focal plane, its perspective projection m in the retinal plane goes to ∞. 39

3.7 The row vectors of matrix $\tilde{\mathbf{P}}$ define the focal plane and the line joining the optical center to the origin of the coordinates in the retinal plane. 40

3.8 Changing coordinate systems in the retinal plane. 42

3.9 Relationship between the real and normalized retinal planes. 44

3.10 How to compute the angle between the optical rays $\langle C, m\rangle$ and $\langle C, n\rangle$ using the image of the absolute conic. 47

3.11 Changing coordinate systems in 3-D space. 49
3.12 The retinal coordinate system may not even be orthogonal. 56
3.13 How the intrinsic parameters vary with the pixel noise. 64
3.14 How the extrinsic parameters vary with the pixel noise. 65
4.1 Edges in an image have different physical sources. 70
4.2 A simple paradigm for edge detection. 72
4.3 The Roberts operator. 75
4.4 Sobel's ($c = 2$) and Prewitt's ($c = 1$) operators. 76
4.5 Edge defects. 77
4.6 The function $D^2G_\sigma(x)$ and its Fourier transform. 81
4.7 The surface $D^2_{45}G(r)$ (see text). 82
4.8 An instance in which the Laplacian's detection is in error. 83
4.9 A picture of an office scene. 84
4.10 The zero-crossings of the Marr-Hildreth operator applied to the image in figure 4.9. 85
4.11 Only the zero-crossings of the image in figure 4.9 for which the magnitude of the gradient is above a certain threshold have been retained. 86
4.12 Edge detection by convolution with $h(x)$. 92
4.13 Difference of boxes. 94
4.14 Problems with the matched filter (see text). 97
4.15 A comparison between the Canny operator and the first derivative of a Gaussian. 101
4.16 The smoothing functions of the Canny, Deriche, and Spacek filters. 109
4.17 A two-dimensional constant-intensity step edge. 110
4.18 In order for $\mathbf{g}$ to be parallel to the normal $\mathbf{n}$ to (c), (C) has to be a special curve on (Σ). 114
4.19 The gradient is normal to the edge, and its magnitude is maximum in that direction. 115
4.20 Checking whether pixel A is a local maximum of the magnitude of the gradient in the direction of the gradient. 115

4.21 The Deriche operator: local maxima for various scales. 116
4.22 The difficulty of choosing the right threshold. 117
4.23 Hysteresis thresholding for the Deriche operator ($\alpha = 1$, $T_2 = 5$, $T_1 = 15$). 118
4.24 The tangent plane T_M to the intensity surface (Σ). 123
5.1 A manifold of dimension 2. 127
5.2 Geometric interpretation of the manifold of two-dimensional directions. 132
5.3 Spherical coordinates. 136
5.4 Stereographic projection. 137
5.5 Representing all directions not parallel to the X_2X_3 plane as a point in the plane of equation $X_1 = 1$. 138
6.1 The 3-D vision problem. 166
6.2 The epipolar geometry. 169
6.3 $\langle C_1, C_2 \rangle$ is parallel to the plane $\mathcal{R}_1$: E_1 is at ∞; the epipolar lines are parallel in the plane $\mathcal{R}_1$ and intersect at E_2 in the plane $\mathcal{R}_2$. 171
6.4 $\langle C_1, C_2 \rangle$ is parallel to the planes $\mathcal{R}_1$ and $\mathcal{R}_2$: E_1 and E_2 are at ∞; the epipolar lines are parallel in both planes $\mathcal{R}_1$ and $\mathcal{R}_2$. 171
6.5 Two cameras and their normalized retinas. 173
6.6 Simple disparity definition. 175
6.7 Relation between depth and disparity. 176
6.8 The epipolar plane configuration. 177
6.9 General definition of disparity. 178
6.10 Forbidden zone attached to M_1. 179
6.11 M and N are not both visible from retina 1. 180
6.12 M and N do not belong to the same object: The ordering constraint does not apply. 180
6.13 The object bulges out of the forbidden zone, and the ordering constraint applies. 181
6.14 Understanding the disparity gradient. 182

6.15 The forbidden zone obtained when $K < 2$ is larger than the usual one. 183

6.16 The forbidden zone obtained when $K > 2$ is smaller than the usual one. 184

6.17 Relationships between a three-dimensional curve (C) and its images in three cameras. 185

6.18 The rectification process is cast as the reprojection of the two retinas $\mathcal{R}_1$ and $\mathcal{R}_2$ onto plane $\mathcal{R}$ containing the two new retinas $\mathcal{R}'_1$ and $\mathcal{R}'_2$. 189

6.19 Principle of the correlation technique. 190

6.20 A stereo pair of images. 191

6.21 The rectified stereo pair. 192

6.22 The agreement between C_{12} and C_{21} (see text). 193

6.23 The three-dimensional reconstruction of figure 6.20. On the left is the smoothed disparity, and on the right is the disparity superimposed as a grid on one of the original images. 194

6.24 The stereo matching process can be considered finding a path in a graph. 199

6.25 The epipolar constraint for two segments AB and CD. 203

6.26 The epipolar constraint for line segments. 203

6.27 Multiple segment correspondences. 204

6.28 Geometry of the stereo problem when looking at a plane. 207

6.29 A stereo pair showing a floor and some obstacles. 210

6.30 The obstacles detected in the scene of figure 6.29. 211

6.31 The principle of disambiguation of matches. 212

6.32 A stereo triplet of an office room. 214

6.33 The line segments found in the images of the previous figure. 215

6.34 The three-dimensional reconstruction of the office room of figure 6.32. 216

6.35 A top view of the reconstructed office scene. 216

6.36 A stereo triplet of a scene with rocks. 217

6.37 The line segments found in the images of the previous figure. 218
6.38 The three-dimensional reconstruction of the rocks scene of figure 6.36. 219
6.39 A top view of the reconstructed rocks scene. 219
6.40 A stereo triplet of a scene with man-made objects, including a box of tea. 220
6.41 The line segments found in the images of the previous figure. 221
6.42 The three-dimensional reconstruction of the tea box scene of figure 6.40. 222
6.43 A top view of the reconstructed tea box scene. 222
6.44 The edge curves found in the images of figure 6.32. 224
6.45 The three-dimensional reconstruction of the office room of figure 6.32. 225
6.46 A top view of the reconstructed office scene. 225
6.47 The edge curves found in the images of figure 6.36. 226
6.48 The three-dimensional reconstruction of the rocks scene of figure 6.36. 227
6.49 A top view of the reconstructed rocks scene. 227
6.50 The edge curves found in the images of figure 6.40. 228
6.51 The three-dimensional reconstruction of the tea box scene of figure 6.40. 229
6.52 A top view of the reconstructed tea box scene. 229
6.53 A photograph of the cameras to be calibrated. 231
6.54 A calibration pattern: calibration points are at the intersection of the horizontal and vertical lines, which can be easily detected by simple algorithms. 232
6.54 An image of the calibration pattern as seen from one of the cameras. 232
6.56 An image of the calibration pattern with a polygonal approximation of the edges superimposed. 233

6.56 An image of the calibration pattern with the detected corners superimposed. 233
6.58 Stereo pair of the calibration pattern. 234
6.58 The epipolar line of a point of interest. 234
6.60 Epipolar pencil. 235
6.61 The square image $abcd$ is projected as a quadrilateral $ABCD$ onto the plane $\mathcal{R}$. 242
7.1 Constraint on the motion parameters derived from point matches. 249
7.2 The plane angles (xCm, xCp) and $(x'C'm', x'C'p')$ are equal. 259
7.3 The images of the absolute conic Ω. 259
7.4 The situation in the image plane at the first time instant. 260
7.5 The situation in the image plane at the second time instant. 262
7.6 The five-points algorithm. 270
7.7 The shortest distance between two lines. 276
7.8 Computing the distance to the epipolar line. 277
7.9 Two motions yielding the same images. 279
7.10 Two frames are not sufficient to constrain motion. 284
7.11 Three frames provide constraints about motion. 285
7.12 A geometric interpretation of equation 7.47. 288
7.13 An example of double degeneracy. 296
7.14 An example of triple degeneracy. 297
8.1 Tracking tokens provides matches. 319
8.2 The first four frames of the hallway scene. 324
8.3 The last four frames of the hallway scene. 325
8.4 The table can rotate about and translate along a vertical axis. 333
8.5 The images taken by the first camera at t_1 and t_{10}. 333
8.6 The superposition of the first frame (in solid lines) and the tenth frame (in dashed lines). 334

8.7 The superposition of the first frame (in solid lines) and the tenth frame (in dashed lines). 335

8.8 The superposition of the third predicted frame (in solid lines) and the third observed frame (in dashed lines). 336

8.9 The superposition of the tenth predicted frame (in solid lines) and the tenth observed frame (in dashed lines). 337

9.1 Optical flow. 344

9.2 The intensity image produced by a Lambertian surface. 345

9.3 Definition of the spatiotemporal surface (Σ). 350

9.4 Projection in the image plane, parallel to the τ-axis, of the curve $S = S_0$ of the surface (Σ): $(c^r_{m_0})$ is the "real" trajectory of m_0. 351

9.5 Projection in the image plane, parallel to the τ-axis, of the curve $s = s_0$ of the surface (Σ): $(c^a_{m_0})$ is the "apparent" trajectory of m_0. 352

9.6 Definition of the two motion fields, the real and the apparent. 353

9.7 Comparison of the two motion fields and the real and apparent trajectories. Here $\mathbf{n}$ is the normal to (c_τ). 353

9.8 The vectors $\mathbf{t}_0$ and $\mathbf{n}_\beta$ span the tangent plane T_P to Σ. 355

9.9 Various vectors of the tangent plane T_P to (Σ). 356

9.10 A geometric interpretation of $\partial_{\mathbf{n}_\beta}$ (see text). 357

9.11 The choice of the origin of arclength on (c_τ). 361

9.12 The vector $\mathbf{n}$ is normal to the plane defined by the optical center C of the camera and the line l. 370

9.13 The perspective projection. 372

9.14 Definition of the local system of coordinates. 380

9.15 Relation between $\mathbf{t}$ and $\mathbf{T}$. 381

9.16 When the 3-D motion is in a plane parallel to the retinal plane, the apparent and real motion fields are identical. 389

9.17 The reconstructed curves corresponding to the two solutions of the motion equations. The left side is the correct solution; the

right side shows the superposition of the correct one and the incorrect one, which is a hyperbola. 391

9.18 The reconstructed curves corresponding to the three solutions of the motion equations. The top left is the correct solution; the bottom and top right show the superposition of the correct one and the incorrect ones. 392

9.19 A plot of the real (left) and apparent (right) motion fields along the observed ellipse. 393

9.20 Matching two curves in stereo. 394

9.21 n possible hypotheses for matching curve (c) in image 1 to a curve in image 2. 398

10.1 A simple shape and its associated quadtree. 405

10.2 The square of figure 10.1 translated by $[0,1]^T$ and its associated quadtree. Note the dissimilarity between 10.1.b and 10.2.b. 406

10.3 An example of the Minkowski sum of two compact sets. 413

10.4 Voronoi diagram of a set of points. 417

10.5 Proof of proposition 10.2. 419

10.6 Proof of proposition 10.3. 420

10.7 The Delaunay triangulation of the points of figure 10.4. 421

10.8 The skeleton of the object X is approximated by the centers of the Delaunay circles. 423

10.9 The hypothesis that ∂X is C^3 is essential to the proof of proposition 10.7. This is an example of a sequence such that $\lim S_i$ does not belong to $Bmax(X)$. 426

10.10 The open segment AB of the skeleton cannot be reached by the centers of the Delaunay spheres. 428

10.11 Proof of proposition 10.9. 430

10.12 End of the proof of proposition 10.9. 430

10.13 The system of polar coordinates centered at P_i. 432

10.14 Since M falls in the discs of triangles AFG, AGB, BGC, and CDG, these four simplexes are deleted. Face AG is common to AFG and AGB. It is deleted, as are GB and GC, and the six

simplexes MAB, MBC, MCD, MDG, MGF, and MFA are added to the list of Delaunay simplexes. 435

10.15 The number of tetrahedra is a linear function of the number of points inserted. 437

10.16 Segment CD is not a Delaunay edge because all circles going through C and D contain at least two other data points. 439

10.17 Adding points M and N to segment CD makes segments CM, MN, and ND Delaunay edges. 439

10.18 PQ is a neighbor of MN in case (a) but not in case (b). 440

10.19 The case where $B_M \cap B_N = \emptyset$. 441

10.20 The case where $B_M \cap B_N \neq \emptyset$. 442

10.21 Initial set of segments to be triangulated. 443

10.22 Unconstrained triangulation of the segments of figure 10.21. 444

10.23 Constrained triangulation of the segments of figure 10.21. 445

10.24 An example of boolean schema. The primary grain is a triangle. The shape and orientation of the triangles as well as their positions in the plane vary randomly. 447

10.25 The shape of $X' \oplus \check{K}$. 448

10.26 The figure shows the original Delaunay triangulation of the points of ∂X and the polygon approximating the surface of the object. 449

10.27 The Labyrinth to be explored by the mobile robot. The Delaunay triangulation of the first points measured by stereo. A model of the environment after removal of the empty triangles. (The robot is represented by a small cross.) 450

10.28 The robot has moved into free space and made more measurements. The Delaunay triangulation of the points measured in the two positions. A new model of the environment. 451

10.29 The robot has progressed farther into its environment. The evolution of its model of free space and obstacles. 452

10.30 One reason for using the constrained Delaunay triangulation (see text). 453

10.31 The stereo triangle PMN intersects the tetrahedron $QRST$, which can then be marked as empty. 454

10.32 Polyhedral approximation of a foundry casting. 455

10.33 Polyhedral approximation of a face. 456

10.34 Polyhedral approximation of an oil bottle. 456

10.35 Polyhedral approximation of a funnel. 457

10.36 Stereo triplet of a sphere (a), and the resulting polyhedral approximation (b). 458

10.37 Stereo triplet of a cylinder (a), and the resulting polyhedral approximation (b). 459

10.38 Stereo triplet of a cone (a), and the resulting polyhedral approximation (b). 460

10.39 A stereo triplet of a part of the laboratory. 461

10.40 The line segments that have been matched in the stereo triplet corresponding to figure 10.39. 462

10.41 A cross-section of the Delaunay triangulation with a plane parallel to the floor before and after the removal of empty tetrahedra for the scene of figure 10.39 (using one viewpoint). 463

10.42 A "synthetic" image created from the polyhedral model of the scene shown in figure 10.39. The result is shown from two different viewpoints. 464

10.43 A stereo triplet of another part of the laboratory. 465

10.44 The line segments that have been matched in the stereo triplet corresponding to figure 10.43. 465

10.45 A cross-section of the Delaunay triangulation with a plane parallel to the floor before and after the removal of empty tetrahedra for the scene of figure 10.43 (using one viewpoint). 466

10.46 The second viewpoint corresponding to figure 10.39. 466

10.47 The line segments that have been matched in the stereo triplet corresponding to figure 10.46. 467

10.47 A cross-section of the Delaunay triangulation with a plane parallel to the floor before and after the removal of empty tetrahedra for the scene of figures 10.39 and 10.46 (using two viewpoints). 467

10.49 The "synthetic" image obtained by combining figures 10.39 and 10.46. 468

10.49 The two extra viewpoints corresponding to figure 10.43. 468

10.51 The line segments that have been matched in the stereo triplet corresponding to figure 10.49. 469

10.52 A cross-section of the Delaunay triangulation with a plane parallel to the floor before and after the removal of empty tetrahedra for the scene of figures 10.43 and 10.49 (using three viewpoints). 470

10.52 The "synthetic" image obtained by combining figures 10.43 and 10.49. 470

10.54 The approximation of the oil bottle of figure 10.34 with planes (a) and quadrics (b). 475

10.55 The approximation of the funnel of figure 10.35 with planes (a) and quadrics (b). 476

11.1 (a) The model and the scene. (b) The Hough accumulator for the rotation angle when we allow all possible matches between model and scene. 486

11.2 (a) The values in the Hough accumulator for the rotation angle when we allow matches between the model and the scene such that the ratio of the lengths of the shortest segment to the longest is $\geq .5$. (b) The result when we allow only matches of segments of equal length. 487

11.3 The two graphs associated with the model and the scene of figure 11.1.a. 489

11.4 The association graph of the two graphs of figure 11.3 490

11.5 The interpretation tree. 496

11.6 An example of the kind of scene that the recognition system can cope with. 498

11.7 Unpredictable artifacts can be present on the objects. 499

11.8 Two reference parts. 499

11.9 Polygonal models of the parts of figure 11.8. 500

11.10 Polygonal models of some of the parts of figure 11.6. 501