3-D Shape Estimation and Image Restoration

Paolo Favaro and Stefano Soatto

3-D Shape Estimation and Image Restoration

Exploiting Defocus and Motion Blur

Springer

Paolo Favaro
Heriot-Watt University, Edinburgh, UK
http://www.eps.hw.ac.uk/~pf21

Stefano Soatto
University of California, Los Angeles, USA
http://www.cs.ucla.edu/~soatto

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2006931781

ISBN-10: 1-84628-176-8 Printed on acid-free paper
ISBN-13: 978-1-84628-176-1
e-ISBN-10: 1-84628-688-3
e-ISBN-13: 978-1-84628-688-9

© Springer-Verlag London Limited 2007

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

The use of registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant laws and regulations and therefore free for general use.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

9 8 7 6 5 4 3 2 1

Springer Science+Business Media
springer.com

To Maria and Giorgio

Paolo Favaro

To Anna and Arturo

Stefano Soatto

Preface

Images contain information about the spatial properties of the scene they depict. When coupled with suitable assumptions, images can be used to infer three-dimensional information. For instance, if the scene contains objects made with homogeneous material, such as marble, variations in image intensity can be associated with variations in shape, and hence the "shading" in the image can be exploited to infer the "shape" of the scene (shape from shading). Similarly, if the scene contains (statistically) regular structures, variations in image intensity can be used to infer shape (shape from textures). Shading, texture, cast shadows, occluding boundaries are all "cues" that can be exploited to infer spatial properties of the scene from a single image, when the underlying assumptions are satisfied. In addition, one can obtain spatial cues from multiple images of the same scene taken with changing conditions. For instance, changes in the image due to a moving light source are used in "photometric stereo," changes in the image due to changes in the position of the cameras are used in "stereo," "structure from motion," and "motion blur." Finally, changes in the image due to changes in the geometry of the camera are used in "shape from defocus." In this book, we will concentrate on the latter two approaches, motion blur and defocus, which are referred to collectively as "accommodation cues." Accommodation cues can be exploited to infer the 3-D structure of the scene as well as its radiance properties, which in turn can be used to generate better quality novel images than the originals.

Among visual cues, defocus has received relatively little attention in the literature. This is due in part to the difficulty in exploiting accommodation cues: the mathematical tools necessary to analyze accommodation cues involve continuous analysis; unlike stereo and motion which can be attacked with simple

linear algebra. Similarly, the design of algorithms to estimate 3-D geometry from accommodation cues is more difficult because one has to solve optimization problems in infinite-dimensional spaces. Most of the resulting algorithms are known to be slow and lack robustness in respect to noise.

Recently, however, it has been shown that by exploiting the mathematical structure of the problem one can reduce it to linear algebra, (as we show in Chapter 4,) yielding very simple algorithms that can be implemented in a few lines of code. Furthermore, links established with recent developments in variational methods allow the design of computationally efficient algorithms. Robustness to noise has significantly improved as a result of designing optimal algorithms.

This book presents a coherent analytical framework for the analysis and design of algorithms to estimate 3-D shape from defocused and blurred images, and to eliminate defocus and blur and thus yield "restored" images. It presents a collection of algorithms that are shown to be optimal with respect to the chosen model and estimation criterion. Such algorithms are reported in MATLAB® notation in the appendix, and their performance is tested experimentally.

The style of the book is tailored to individuals with a background in engineering, science, or mathematics, and is meant to be accessible to first-year graduate students or anyone with a degree that included basic linear algebra and calculus courses. We provide the necessary background in optimization and partial differential equations in a series of appendices.

The research leading to this book was made possible by the generous support of our funding agencies and their program managers. We owe our gratitude in particular to Belinda King, Sharon Heise, and Fariba Fahroo of AFOSR, and Behzad Kamgar-Parsi of ONR. We also wish to thank Jean-Yves Bouguet of Intel, Shree K. Nayar of Columbia University, New York, and also the National Science Foundation.

September 2006

PAOLO FAVARO
STEFANO SOATTO

Organizational chart

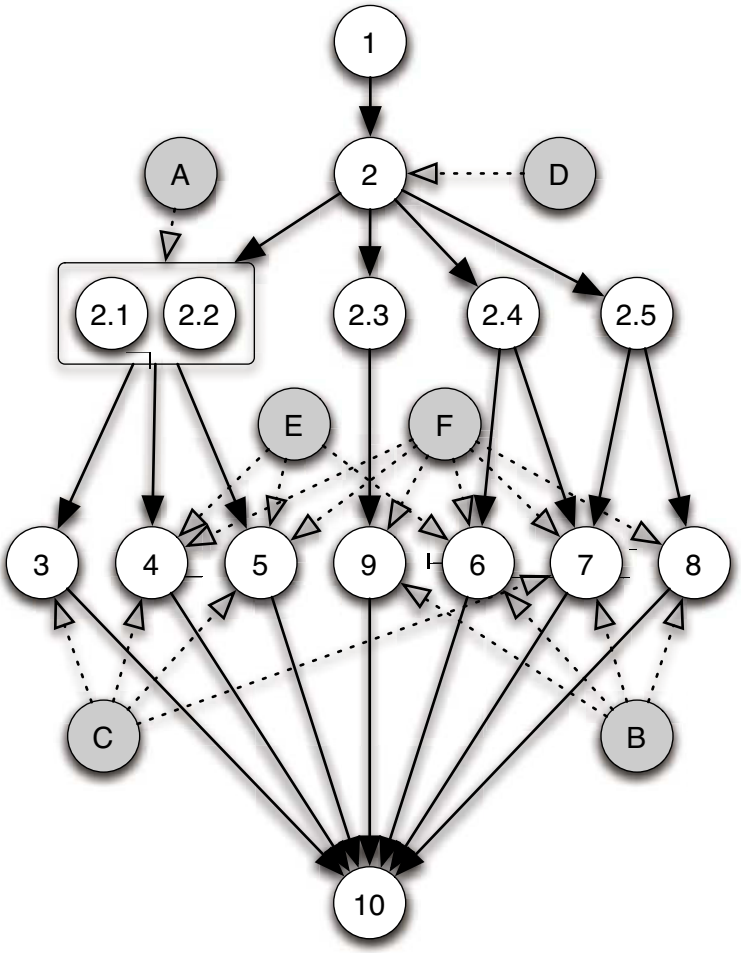

Figure 1. Dependencies among chapters.

Contents

Preface **vii**

1 Introduction **1**

1.1 The sense of vision 1

1.1.1 Stereo 4

1.1.2 Structure from motion 5

1.1.3 Photometric stereo and other techniques based on controlled light 5

1.1.4 Shape from shading 6

1.1.5 Shape from texture 6

1.1.6 Shape from silhouettes 6

1.1.7 Shape from defocus 6

1.1.8 Motion blur 7

1.1.9 On the relative importance and integration of visual cues 7

1.1.10 Visual inference in applications 8

1.2 Preview of coming attractions 9

1.2.1 Estimating 3-D geometry and photometry with a finite aperture 9

1.2.2 Testing the power and limits of models for accommodation cues 10

1.2.3 Formulating the problem as optimal inference 11

1.2.4 Choice of optimization criteria, and the design of optimal algorithms 12

1.2.5 Variational approach to modeling and inference from accommodation cues 12

2 Basic models of image formation **14**
2.1 The simplest imaging model 14
2.1.1 The thin lens 14
2.1.2 Equifocal imaging model 16
2.1.3 Sensor noise and modeling errors 18
2.1.4 Imaging models and linear operators 19
2.2 Imaging occlusion-free objects 20
2.2.1 Image formation nuisances and artifacts 22
2.3 Dealing with occlusions 23
2.4 Modeling defocus as a diffusion process 26
2.4.1 Equifocal imaging as isotropic diffusion 28
2.4.2 Nonequifocal imaging model 29
2.5 Modeling motion blur 30
2.5.1 Motion blur as temporal averaging 30
2.5.2 Modeling defocus and motion blur simultaneously . . 34
2.6 Summary 35

3 Some analysis: When can 3-D shape be reconstructed from blurred images? **37**
3.1 The problem of shape from defocus 38
3.2 Observability of shape 39
3.3 The role of radiance 41
3.3.1 Harmonic components 42
3.3.2 Band-limited radiances and degree of resolution . . . 42
3.4 Joint observability of shape and radiance 46
3.5 Regularization 46
3.6 On the choice of objective function in shape from defocus . . 47
3.7 Summary 49

4 Least-squares shape from defocus **50**
4.1 Least-squares minimization 50
4.2 A solution based on orthogonal projectors 53
4.2.1 Regularization via truncation of singular values . . . 53
4.2.2 Learning the orthogonal projectors from images . . . 55
4.3 Depth-map estimation algorithm 58
4.4 Examples 60
4.4.1 Explicit kernel model 60
4.4.2 Learning the kernel model 61
4.5 Summary 65

5 Enforcing positivity: Shape from defocus and image restoration by minimizing I-divergence **69**
5.1 Information-divergence 70
5.2 Alternating minimization 71
5.3 Implementation 76

5.4 Examples 76
5.4.1 Examples with synthetic images 76
5.4.2 Examples with real images 78
5.5 Summary 79

6 Defocus via diffusion: Modeling and reconstruction 87
6.1 Blurring via diffusion 88
6.2 Relative blur and diffusion 89
6.3 Extension to space-varying relative diffusion 90
6.4 Enforcing forward diffusion 91
6.5 Depth-map estimation algorithm 92
6.5.1 Minimization of the cost functional 94
6.6 On the extension to multiple images 95
6.7 Examples 96
6.7.1 Examples with synthetic images 97
6.7.2 Examples with real images 99
6.8 Summary 99

7 Dealing with motion: Unifying defocus and motion blur 106
7.1 Modeling motion blur and defocus in one go 107
7.2 Well-posedness of the diffusion model 109
7.3 Estimating Radiance, Depth, and Motion 110
7.3.1 Cost Functional Minimization 111
7.4 Examples 113
7.4.1 Synthetic Data 114
7.4.2 Real Images 117
7.5 Summary 118

8 Dealing with multiple moving objects 120
8.1 Handling multiple moving objects 121
8.2 A closer look at camera exposure 124
8.3 Relative motion blur 125
8.3.1 Minimization algorithm 126
8.4 Dealing with changes in motion 127
8.4.1 Matching motion blur along different directions . . . 129
8.4.2 A look back at the original problem 131
8.4.3 Minimization algorithm 132
8.5 Image restoration 135
8.5.1 Minimization algorithm 137
8.6 Examples 138
8.6.1 Synthetic data 138
8.6.2 Real data 141
8.7 Summary 146

9 Dealing with occlusions **147**
9.1 Inferring shape and radiance of occluded surfaces 148
9.2 Detecting occlusions 150
9.3 Implementation of the algorithm 151
9.4 Examples 152
9.4.1 Examples on a synthetic scene 152
9.4.2 Examples on real images 154
9.5 Summary 157

10 Final remarks **159**

A Concepts of radiometry **161**
A.1 Radiance, irradiance, and the pinhole model 161
A.1.1 Foreshortening and solid angle 161
A.1.2 Radiance and irradiance 162
A.1.3 Bidirectional reflectance distribution function 163
A.1.4 Lambertian surfaces 163
A.1.5 Image intensity for a Lambertian surface and a pinhole lens model 164
A.2 Derivation of the imaging model for a thin lens 164

B Basic primer on functional optimization **168**
B.1 Basics of the calculus of variations 169
B.1.1 Functional derivative 170
B.1.2 Euler–Lagrange equations 171
B.2 Detailed computation of the gradients 172
B.2.1 Computation of the gradients in Chapter 6 172
B.2.2 Computation of the gradients in Chapter 7 174
B.2.3 Computation of the gradients in Chapter 8 176
B.2.4 Computation of the gradients in Chapter 9 185

C Proofs **190**
C.1 Proof of Proposition 3.2 190
C.2 Proof of Proposition 3.5 191
C.3 Proof of Proposition 4.1 192
C.4 Proof of Proposition 5.1 194
C.5 Proof of Proposition 7.1 195

D Calibration of defocused images **197**
D.1 Zooming and registration artifacts 197
D.2 Telecentric optics 200

E MATLAB® implementation of some algorithms **202**
E.1 Least-squares solution (Chapter 4) 202

E.2 I-divergence solution (Chapter 5) 212
E.3 Shape from defocus via diffusion (Chapter 6) 221
E.4 Initialization: A fast approximate method 229

F Regularization **232**
F.1 Inverse problems 232
F.2 Ill-posed problems 234
F.3 Regularization 235
F.3.1 Tikhonov regularization 237
F.3.2 Truncated SVD 238

References **239**

Index **247**