Interdisciplinary Applied Mathematics

Volume 26

Problems in engineering, computational science, and the physical and biological sciences are using increasingly sophisticated mathematical techniques. Thus, the bridge between the mathematical sciences and other disciplines is heavily traveled. The correspondingly increased dialog between the disciplines has led to the establishment of the series: *Interdisciplinary Applied Mathematics*.

The purpose of this series is to meet the current and future needs for the interaction between various science and technology areas on the one hand and mathematics on the other. This is done, firstly, by encouraging the ways that mathematics may be applied in traditional areas, as well as point towards new and innovative areas of applications; and, secondly, by encouraging other scientific disciplines to engage in a dialog with mathematicians outlining their problems to both access new methods and suggest innovative developments within mathematics itself.

The series will consist of monographs and high-level texts from researchers working on the interplay between mathematics and other fields of science and technology.

Interdisciplinary Applied Mathematics

Volumes published are listed at the end of this book.

Springer Science+Business Media, LLC

An Invitation to 3-D Vision
From Images to Geometric Models

Yi Ma
Stefano Soatto
Jana Košecká
S. Shankar Sastry

With 170 Illustrations

Springer Science+Business Media, LLC

Yi Ma
Department of Electrical and
Computer Engineering
University of Illinois at
Urbana-Champaign
Urbana, IL 61801
USA
yima@uiuc.edu

Stefano Soatto
Department of Computer Science
University of California, Los Angeles
Los Angeles, CA 90095
USA
soatto@ucla.edu

Jana Košecká
Department of Computer Science
George Mason University
Fairfax, VA 22030
USA
kosecka@cs.gmu.edu

S. Shankar Sastry
Department of Electrical Engineering
and Computer Science
University of California, Berkeley
Berkeley, CA 94720
USA
sastry@eecs.berkeley.edu

Cover Illustration: "Axo-GJ" 1968 (180 × 180 cm) by Victor Vasarely. Copyright Michèle Vasarely.

Mathematics Subject Classification (2000): 51U10, 68U10, 65D18

ISBN 978-1-4419-1846-8 ISBN 978-0-387-21779-6 (eBook)
DOI 10.1007/978-0-387-21779-6
Printed on acid-free paper.

Originally published by Springer-Verlag New York, Inc. in 2004
Softcover reprint of the hardcover 1st edition 2004

9 8 7 6 5 4 3 2 (EB)

springeronline.com

To my mother and my father (Y.M.)

To Giuseppe Torresin, Engineer (S.S.)

To my parents (J.K.)

To my mother (S.S.S.)

Preface

This book is intended to give students at the advanced undergraduate or introductory graduate level, and researchers in computer vision, robotics and computer graphics, a self-contained introduction to the geometry of three-dimensional (3-D) vision. This is the study of the reconstruction of 3-D models of objects from a collection of 2-D images. An essential prerequisite for this book is a course in linear algebra at the advanced undergraduate level. Background knowledge in rigid-body motion, estimation and optimization will certainly improve the reader's appreciation of the material but is not critical since the first few chapters and the appendices provide a review and summary of basic notions and results on these topics.

Our motivation

Research monographs and books on geometric approaches to computer vision have been published recently in two batches: The first was in the mid 1990s with books on the geometry of two views, see e.g. [Faugeras, 1993, Kanatani, 1993b, Maybank, 1993, Weng et al., 1993b]. The second was more recent with books focusing on the geometry of multiple views, see e.g. [Hartley and Zisserman, 2000] and [Faugeras and Luong, 2001] as well as a more comprehensive book on computer vision [Forsyth and Ponce, 2002]. We felt that the time was ripe for synthesizing the material in a unified framework so as to provide a self-contained exposition of this subject, which can be used both for pedagogical purposes and by practitioners interested in this field. Although the approach we take in this book deviates from several other classical approaches, the techniques we use are mainly linear algebra and our book gives a comprehensive view of what is known to date on the geometry of 3-D vision. It also develops homogeneous terminology on a solid analytical foundation to enable what should be a great deal of future research in this young field.

Apart from a self-contained treatment of geometry and algebra associated with computer vision, the book covers relevant aspects of the image formation process, basic image processing, and feature extraction techniques – essentially all that one needs to know in order to build a system that can automatically generate a 3-D model from a set of 2-D images.

Organization of the book

This book is organized as follows: Following a brief introduction, Part I provides background material for the rest of the book. Two fundamental transformations in multiple-view geometry, namely, rigid-body motion and perspective projection, are introduced in Chapters 2 and 3, respectively. Feature extraction and correspondence are discussed in Chapter 4.

Chapters 5, 6, and 7, in Part II, cover the classic theory of two-view geometry based on the so-called epipolar constraint. Theory and algorithms are developed for both discrete and continuous motions, both general and planar scenes, both calibrated and uncalibrated camera models, and both single and multiple moving objects.

Although the epipolar constraint has been very successful in the two-view case, Part III shows that a more proper tool for studying the geometry of multiple views is the so-called *rank condition on the multiple-view matrix* (Chapter 8), which unifies all the constraints among multiple images that are known to date. The theory culminates in Chapter 9 with a unified theorem on a rank condition for arbitrarily mixed point, line, and plane features. It captures *all* possible constraints among multiple images of these geometric primitives, and serves as a key to both geometric analysis and algorithmic development. Chapter 10 uses the rank condition to reexamine and unify the study of single-view and multiple-view geometry given scene knowledge such as symmetry.

Based on the theory and conceptual algorithms developed in the early part of the book, Chapters 11 and 12, in Part IV, demonstrate practical reconstruction algorithms step-by-step, as well as discuss possible extensions of the theory covered in this book. An outline of the logical dependency among chapters is given in Figure 1.

Curriculum options

Drafts of this book and the exercises in it have been used to teach a one-semester course at the University of California at Berkeley, the University of Illinois at Urbana-Champaign, Washington University in St. Louis, the George Mason University and the University of Pennsylvania, and a one-quarter course at the University of California at Los Angeles. There is apparently adequate material for two semesters or three quarters of lectures. Advanced topics suggested in Part IV or chosen by the instructor can be added to the second half of the sec-

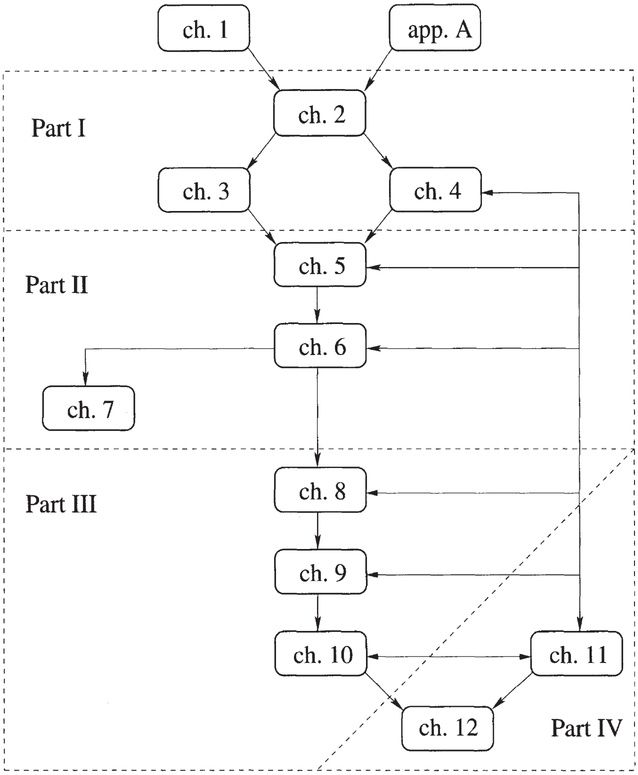

Figure 1. Organization of the book: logical dependency among parts and chapters.

ond semester if a two-semester course is offered. Below are some suggestions for course development based on this book:

1. *A one-semester course:* Appendix A, Chapters 1–6, and part of Chapters 8–10.

2. *A two-quarter course:* Chapters 1–6 for the first quarter, and Chapters 8–10, 12 for the second quarter.

3. *A two-semester course:* Appendix A and Chapters 1–6 for the first semester; Chapters 7–10 and the instructor's choice of some advanced topics from Chapter 12 for the second semester.

4. *A three-quarter sequence:* Chapters 1–6 for the first quarter, Chapters 7–10 for the second quarter, and the instructor's choice of advanced topics and projects from Chapters 11 and 12 for the third quarter.

Chapter 11 plays a special role in this book: Its purpose is to make it easy for the instructor to develop and assign experimental exercises or course projects along with other chapters being taught throughout the course. Relevant code is available at `http://vision.ucla.edu/MASKS`, from which students may get hands-on experience with a minimum version of a working computer vision system. This chapter can also be used by practitioners who are interested in using the algorithms developed in this book, without necessarily delving into the details of the mathematical formulation. Finally, an additional purpose of this chapter is to summarize "the book in one chapter," which can be used in the first lecture as an overview of what is to come.

Exercises are provided at the end of each chapter. They consist of mainly three types:

1. *drill exercises* that help students understand the theory covered in each chapter;
2. *advanced exercises* that guide students to creatively develop a solution to a specialized case that is related to but not necessarily covered by the general theorems in the book;
3. *programming exercises* that help students grasp the algorithms developed in each chapter.

Solutions to selected exercises are available, along with software for examples and algorithms, at `http://vision.ucla.edu/MASKS`.

Yi Ma, Champaign, Illinois
Stefano Soatto, Los Angeles, California
Jana Košecká, Fairfax, Virginia
Shankar Sastry, Berkeley, California
Spring, 2003

Acknowledgments

The idea for writing this book grew during the completion of Yi Ma's doctoral dissertation at Berkeley. The book was written during the course of three years after Yi Ma graduated from Berkeley, when all the authors started teaching the material at their respective institutions. Feedback and input from students was instrumental in improving the quality of the initial manuscript. We are deeply grateful to the student input. Two students whose doctoral research especially helped us are René Vidal at Berkeley and Kun Huang at UIUC. In addition, the research projects of many other students led to the development of new material that became an integral part of this book. We thank especially Wei Hong, Yang Yang at UIUC, Omid Shakernia at Berkeley, Hailin Jin, Paolo Favaro at Washington University in St. Louis, Alessandro Chiuso now at the University of Padova, Wei Zhang at George Mason University, and Marco Zucchelli at the Royal Institute of Technology in Stockholm.

Many colleagues contributed with comments and suggestions on early drafts of the manuscript; in particular, Daniel Cremers of UCLA, Alessandro Duci of the Scuola Normale of Pisa, and attendees of the short course at the UCLA Extension in the Fall of 2002. We owe a special debt of gratitude to Camillo Taylor now at the University of Pennsylvania, Philip McLauchlan now at Imagineer Systems Inc., and Jean-Yves Bouguet now at Intel for their collaboration in developing the real-time vision algorithms presented in Chapter 12. We are grateful to Hany Farid of Dartmouth College for his feedback on the material in Chapters 3 and 4, and Serge Belongie of the University of California at San Diego for his correction to Chapter 5. Kostas Daniilidis offered a one-semester course based on a draft of the book at the University of Pennsylvania in the spring of 2003, and we are grateful to him and his students for valuable comments. We also thank Robert Fossum of the

Mathematics Department of UIUC for proofreading the final manuscript, Thomas Huang of the ECE Department of UIUC for his advice and encouragement all along, Alan Yuille of UCLA, Rama Chellappa and Yiannis Aloimonos of the University of Maryland, and Long Quan of Hong Kong University of Science & Technology for valuable references. We would also like to thank Harry Shum and Zhengyou Zhang for stimulating discussions during our visit to Microsoft Research.

Some of the seeds of this work go back to the early nineties, when Ruggero Frezza, Pietro Perona, and Giorgio Picci started looking at the problem of structure from motion within the context of systems and controls. They taught a precursor of this material in a course at the University of Padova in 1991. Appendix B was inspired by the beautiful lectures of Giorgio Picci. We owe them our sincere gratitude for their vision and for their initial efforts that sparked our interest in the field.

In the vision community, we derived a great deal from the friendly notes of Thomas Huang of UIUC, Yiannis Aloimonos of the University of Maryland, Carlo Tomasi of Duke University, Takeo Kanade of Carnegie Mellon University, and Ruzena Bajcsy now at the University of California at Berkeley.

At the same time, we have also been influenced by the work of Ernst Dickmanns of Bundeswehr University in Munich, Olivier Faugeras of INRIA, and David Mumford of Brown University, who brought the discipline of engineering and the rigor of mathematics into a field that needed both. Their work inspired many to do outstanding work, and we owe them our deepest gratitude.

We are grateful to Jitendra Malik and Alan Weinstein at Berkeley, Pietro Perona at Caltech, Berthold Horn at MIT, P. R. Kumar at UIUC, and Roger Brockett at Harvard for their openness to new points of view and willingness to listen to our initial fledgling attempts. We are also grateful to the group of researchers in the Robotics and Intelligent Machines Laboratory at Berkeley for providing a stimulating atmosphere for the exchange of ideas during our study and visits there, especially Cenk Çavuşoğlu now at Case Western University, Lara Crawford now at Xerox Palo Alto Research Center, João Hespanha now at the University of California at Santa Barbara, John Koo now at Berkeley, John Lygeros now at Cambridge University, George Pappas now at the University of Pennsylvania, Maria Prandini now at Politecnico di Milano, Claire Tomlin now at Stanford University, and Joe Yan now at the University of British Columbia.

Much of the research leading to the material of this book has been generously supported by a number of institutions, funding agencies and their program managers. A great deal of the research was supported by the Army Research Office under the grant DAAH04-96-1-0341 for a center "An Integrated Approach to Intelligent Systems" under the friendly program management of Linda Bushnell and later Hua Wang. We would also like to acknowledge the support of the Office of Naval Research and the program manager Allen Moshfegh for supporting the work on vision-based landing under contract N00014-00-1-062. We also thank the ECE Department, the Coordinated Science Laboratory, the Research Board of UIUC, and the Computer Science Departments at George Mason University and

UCLA for their generous startup funds. We also thank Behzad Kamgar-Parsi of the Office of Naval Research, Belinda King of the Air Force Office of Scientific Research, Jean-Yves Bouguet of Intel, and the Robotics and Computer Vision program of the National Science Foundation.

Finally on a personal note, Yi would like to thank his parents for their remarkable tolerance of his many broken promises for longer and more frequent visits home, and this book is dedicated to their understanding and caring. Stefano would like to thank Anthony Yezzi and Andrea Mennucci for many stimulating discussions. Jana would like to thank her parents for their never ceasing inspiration and encouragement and Frederic Raynal for his understanding, continuing support and love. Shankar would like to thank his mother and Claire Tomlin for their loving support during the preparation of this book.

Contents