



Preface

This book is about mathematics and vision. Researchers and practitioners of machine and biological vision have been working hard over the last forty years or so to understand the laws of image formation, processing and understanding by machines, animals and humans. Although it is clear that this research agenda is far from having been completed, we think that the time has come to provide a more or less complete description of the state of knowledge in one of the subareas of vision, namely the description of the geometric laws that relate different views of a scene.

There are two reasons that makes us believe this; one is theoretical and the other application-motivated. The first reason is that geometry is one of the oldest and most developed parts of mathematics and is at the heart of the process of image formation, object modeling and recognition. Therefore it should not come as a surprise if we state that the framework for studying geometric problems in vision is available and ready for use. The second reason is that in our era of forceful communications through computers, images play a prominent role and will continue to do so in the foreseeable future. There is clearly a need to provide the community of producers and users of images, and in particular of images with a three-dimensional content, with a framework in which their problems can be clearly stated and, one hopes, solved.

This book is thus mostly about geometry because geometry is the natural language to describe 3D shapes and spatial relations. A camera can be thought of as a particular geometric engine, which constructs planar images of the three-dimensional world, through a projection. Although the ancient Greeks already knew several properties of projection, among them the conservation of the cross-ratio, the geometry of the image formation process was first understood by Renaissance painters (see figure 4.1) who made large use of vanishing points and derived

geometric constructions for their practical use. At the time photography was discovered, people studied how to make measurements from perspective views of scenes, and this led to photogrammetry, which has had a wide range of successful applications. During the same century, mathematicians developed *projective geometry*, which was intended to deal with points at infinity and perspective projections. It is the reference framework which will be used all the way through this book, because it deals elegantly with the types of projections that most cameras perform.

Although the natural geometry which we use in most applications is the Euclidean geometry, one of the tenets of this book is that it is simpler and more efficient for vision to consider the Euclidean and affine geometries as special cases of the projective geometry. It is simpler because projective geometry is the geometry of image formation and provides a unified framework for thinking about all geometric problems that are relevant to vision. It is more efficient because the unified framework reduces the need for dealing with special cases and, more importantly, helps the designer of machine vision applications to clearly identify the type of geometric information that is relevant to her/his particular application, e.g. projective, affine or Euclidean, and therefore the type of processing that needs to be applied to the data in order to recover this information from the images. In this sense we are followers of Felix Klein and Herman Weyl, who stated in their Erlangen program of 1872 that the only important geometric properties are those which are invariant to the action of some group of transformations and conversely that every quantity that is invariant to the action of such a group must have an “interesting” geometric interpretation.

Chapter 1 is an introductory chapter which serves to present in a simple way, without much formalization, several of the main ideas of the book.

In order to achieve our goal of enumerating the geometric laws that govern the formation of the images of a scene we have first to collect the relevant mathematical tools. Since those are scattered through the mathematical literature the book has two chapters that develop the geometric and algebraic background that will be needed in the remaining. These chapters might be considered as reference material, and therefore skipped in a first reading. Chapter 2 is an exposition of projective, affine and Euclidean geometry from the Erlangen program viewpoint; that is to say the affine and Euclidean geometries are presented as special cases of the projective geometry. We will find this approach invaluable in almost all the other chapters of the book. The next chapter exposes a complementary viewpoint: whereas chapter 2 is basically geometric, this chapter is mostly algebraic and introduces the Grassman-Cayley algebra of a vector space. The reason this is relevant to vision is that this algebra is to geometry what the Boolean algebra is to logic: it is a tool for computing unions and intersections of linear geometric subspaces. As such it will provide us with another indispensable device for representing and computing with the geometry of multiple cameras or views.

Having laid this groundwork, we can start applying these ideas to systems of

cameras. Chapter 4 is about the simplest of all such systems, one that has only one camera! But besides its pedagogical interest it serves as a testbed for our paradigm: starting from the observation that a camera is a projective engine, we provide a projective description of this engine. This description allows us to speak only about such projective invariant properties as the intersections of planes and lines or projective invariant quantities such as cross-ratios. If we then start wondering about affine invariant properties such as parallelism or certain ratios of lengths, the affine framework enters naturally through the plane at infinity. Finally, if we start pondering about Euclidean invariant properties such as angles and distances, the Euclidean framework becomes a natural thing to use. The power of our stratified approach is that you do not need to throw away everything that you have done so far to work your way through the computation of projective and affine properties; on the contrary, just add a little bit of information, i.e. the image of the absolute conic, and the miracle happens – you can now do Euclidean geometry with your camera.

Having described the simplest of all situations, we next turn in chapter 5 toward a (stratified) study of the systems of two cameras. The key concept in these systems is that of epipolar geometry which is used in all projects dealing with binocular stereo. We show that the epipolar geometry is a projective concept and that it can be described geometrically and algebraically quite simply and efficiently in that framework. No notions of affine or Euclidean geometry are necessary. From the algebraic viewpoint, a single 3×3 matrix, called the Fundamental matrix, summarizes everything you need to know about the epipolar geometry. Keeping in mind that some applications will require more knowledge than just projective, we also analyze how affine and Euclidean information is buried into the Fundamental matrix preparing the ground for their recovery in future chapters.

Because of its conceptual and practical importance chapter 6 is devoted to the methods that allow us to recover the Fundamental matrix from pairs of point correspondences between the views. This will confront us with the problem of parametrizing this matrix in a way that allows practical estimation techniques to be used while guaranteeing that the result of the estimation satisfies the constraint that all Fundamental matrixes must satisfy. It will also take us through the very important issues of deciding which pairs of correspondences are valid pairs and which should be considered as outliers and of characterizing the uncertainty of the estimated matrix, information that is quite important in applications.

Chapter 7 builds on the previous three chapters and achieves two main goals. It first spells out in detail the three levels of representation of the geometry of two views, projective, affine, and Euclidean, together with the amount of information that needs to be recovered either from the images themselves or from some external demon in order to reach a given level. In particular we introduce the idea of the canonical representation of a set of two cameras. The representation is attached to a given level and compactly represents all you have to know about the two cameras

in order to analyze the geometry of the scene. Second, it describes a number of interesting vision tasks that can be achieved at each of the three levels of description.

In chapter 8 we address the problem of three views. Just like in the case of two views where the major concept was that of epipolar geometry and the corresponding Fundamental matrix, the relevant idea in the case of three views is that of trifocal geometry and the corresponding Trifocal tensors. We show that these tensors contain all the information about the geometry of three views. The Grassman-Cayley algebra is really useful here in providing simple and elegant descriptions of the relevant geometry and algebra. Of special importance for the next chapter is the analysis of the constraints that are satisfied by the coefficients of these tensors: they will play a prominent role in the estimation methods that will be presented. The Trifocal tensors are, like the Fundamental matrixes, purely projective entities but also contain affine and Euclidean information. The affine and Euclidean forms of the tensors are presented at the end of the chapter.

If estimating the Fundamental matrix of a pair of views is important it is no surprise that estimating the Trifocal tensors is vital in the case of three views. Chapter 9 is dedicated to this problem, which we solve in pretty much the same way as we solved the corresponding problem for the Fundamental matrix. The complexity is higher for several reasons. First we must use triples of correspondences rather than pairs, and second the constraints that must be satisfied by the coefficients of the Tensors are significantly more complicated than the one satisfied by those of the Fundamental matrixes. These algebraic constraints described in chapter 8 are used to parametrize the tensors so as to guarantee that the results of the estimation procedures will indeed be valid Trifocal tensors.

We went from the analysis of one view to two views, then to three views; is there an end to this? Surprisingly enough, and to the reader's relief, the answer to this question is yes. There are such things as Quadrifocal tensors and so on that describe the correspondences between four views but they are all algebraically dependent upon the Trifocal tensors and the Fundamental matrixes of the sub-triples and sub-pairs of views of the four views. Chapter 10 is therefore a brave leap into the world of N -views geometry for N arbitrary and greater than three. Because of this, the best way to represent the geometry of N cameras is through their projection matrices and the notion of the canonical representation of such matrices introduced in chapter 7 becomes even more useful. As usual we consider three such representations, one for each of the three levels of description. We spend quite some time describing various methods for computing the projective canonical representation from the Fundamental matrixes or the Trifocal tensors because this is the basic representation we start from even in the cases where affine and Euclidean descriptions are required. We indicate how it can be refined by Bundle Adjustment, a technique borrowed from photogrameters and adapted to the projective framework.

Chapter 11 begins with the theoretical analysis of a very practical problem, that

of recovering the Euclidean structure of the environment from a pair of views. In general this is achieved by adding to the environment a calibration object, i.e. an object with known Euclidean properties. This is very cumbersome if not impossible for many applications thereby motivating our interest in a solution that does not require this addition. It turns out that the connection between projective, affine and Euclidean geometry offers a natural set of solutions to the initial problem through the use of the images of the plane at infinity and the absolute conic. We call these techniques “self-calibration” since they do not require the use of any calibration objects other than the previous two mathematical entities. We conclude with examples of applications to the construction of 3D Euclidean models from an arbitrary number of uncalibrated views and to the insertion of synthetic 3D objects in image sequences, video or film.

Most chapters end with two sections, one that summarizes and discusses the main results in the chapter and another that provides more references and some further reading.

A comment on the style of this book. We have deliberately adopted the style of a book in mathematics with definitions, lemmas, propositions and theorems at the risk of losing the interest of some readers. We have also provided most of the proofs or given pointers to references where these proofs could be found. The reason is that we believe that vision has to establish itself as a science and that it will not do so if it does not ground itself in mathematics. A theorem is a theorem and an algorithm that makes use of it may or may not work. But if it does not work the cause will not have to be searched for in the theorem but elsewhere, thereby making the task of the designer of the algorithm, if not easier, at least better defined.

Of course we do not claim that our book solves the vision problem, if such a thing exists. We are very much aware of the fact that vision is much richer and intricate than geometry and that this book covers only a very small part of the material relevant to the field. Nonetheless we believe that our book offers the reader a number of conceptual tools and a number of theoretical results that are likely to find their way into many machine vision algorithms.

Acknowledgments This book took seven years to write while the authors were working in different places in Europe (INRIA Sophia-Antipolis) and the US (MIT, Berkeley and SRI). During the course of the writing, Olivier Faugeras has benefited from many stimulating discussions with colleagues, collaborators and students at INRIA and MIT, in particular Didier Bondyfalat, Sylvain Bougnoux, Gabriela Csurka, Rachid Deriche, Frédéric Devernay, Eric Grimson, Radu Horaud, Stéphane Laveau, Liana Lorigo, Leonard Mcmillan, Eric Miller, Roger Mohr, Bernard Mourrain, Luc Robert, Seth Teller, Thierry Viéville, Cyril Zeller, Zhengyou Zhang and Imad Zoghlami. Special thanks go to Liana Lorigo and Eric Miller for their careful proof reading of the final manuscript. He is also very grateful to Luc Robert and Imad Zoghlami for taking a brave leap into the thriving world of industry and

turning many of the ideas contained in this book into products. Thanks also to Dominique Pouliquen for telling them how to sell these products.

Last but not least he is extremely grateful to his wife, Agnès, and their sons Blaise, Clément, Cyrille and Quentin for their love, patience and support.

Quang-Tuan Luong wrote a first draft of what was to become eventually this book while he was a visiting scientist at the University of California at Berkeley. He would like to acknowledge the support and guidance he received from Jitendra Malik during that time. Discussions with David Forsyth, and the initial proof-reading of Joe Weber were also much appreciated. While colleagues at SRI International were all supportive, Marty Fischler deserves special thanks for showing constant interest for this work, and suggesting several clarifications which made the material more accessible. Carlo Tomasi gave Quang-Tuan Luong the opportunity to teach the material at Stanford. Frank Dellaert provided detailed comments on the introductory chapter.

He would like to extend thanks to old and new members of the Robotvis group, who were as welcoming and helpful during his short stays at INRIA, as when he was a graduate student there. His last thought goes to his parents, who encouraged him to become a scientist, and provided great love and support. In particular, he would like to dedicate this book to his mother Luong Ngoc Thu who inquired about the progress almost weekly, and to his father Luong The Vinh, who would have been pleased with the completion of this work.

The authors have also enjoyed interacting over the years with such colleagues as Richard Hartley, Amnon Shashua, Gunnar Sparr and Andrew Zisserman.

The final word of thanks is for Bernhard Geiger who drew the pictures that come before each chapter and tell us the adventures of Euclide in the world of Computer Vision.