Cameras, Shapes, and Contours: Geometric Models in Computer Vision

Matthew Trager

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'École Normale Supérieure

Cameras, Shapes, and Contours: Geometric Models in Computer Vision

École doctorale n°386

SCIENCES MATHÉMATIQUES DE PARIS-CENTRE

Spécialité INFORMATIQUE

COMPOSITION DU JURY :

M Martial Hebert
Carnegie Mellon University
Directeur de thèse

M Fredrik Kahl
Lund University
Président du jury

M Tomas Pajdla
Czech Technical University
Rapporteur

M Marc Pollefeys
ETH Zürich
Rapporteur

M Jean Ponce
École Normale Supérieure
Directeur de thèse

M Bernd Sturmfels
MPI Leipzig
Membre du jury

Soutenue par **Matthew Trager**
le 10 juillet 2018

Dirigée par **Jean Ponce**
et par **Martial Hebert**

ENS | PSL

Abstract

This thesis studies mathematical models for describing the geometry of imaging processes in computer vision. In a broad sense, our contributions are focused on developing frameworks that are very general and at the same time require minimal assumptions. Our approach is in fact rooted in the language of projective geometry, which provides the most general setting for studying properties of lines and incidences that are at the heart of geometric vision. We also apply some tools from algebraic geometry, since many of the objects that we encounter are described by polynomial equations. For example, the multi-view geometry of n pinhole cameras (as well as other cameras) can be encoded in the "joint image", that is an algebraic variety in $(\mathbb{P}^2)^n$ formed by all point correspondences. The Grassmannian of lines $\mathrm{Gr}(1, \mathbb{P}^3)$ also plays a central role in our study. In particular, surfaces in the Grassmannian (or "line congruences") can be used to represent abstract cameras, that are mappings from points to viewing lines. This description is also convenient for studying in a unified manner the multi-view geometry of families of general imaging systems. In addition to modeling cameras, we also investigate the relationship between 3D shapes and their images. For arbitrary sets projecting onto opaque silhouettes, the image is determined by the set of viewing lines that meet the observed object. This leads to the study of "projective visual hulls" and "multi-view consistency". For smooth surfaces, the "image contour" is determined by the set of viewing lines that are tangent to the surface. From this perspective, the evolution of the contour from a moving viewpoint, and the associated "visual events", can be described by studying sets of lines in $\mathrm{Gr}(1, \mathbb{P}^3)$ that have special incidence properties with the surface.

Acknowledgments

I first wish to thank my advisors Jean Ponce and Martial Hebert. They introduced me to the fascinating world of geometry and vision, and provided invaluable guidance. In particular, I am grateful to Jean for caring so much about my work, and for countless interesting discussions that always made me feel excited about doing research. I thank Martial for his support and wise advice, and for the months I spent at CMU, which I greatly enjoyed. I acknowledge the ERC advanced grant VideoWorld, the Insitut Universitaire de France, the GAYA Inria/CMU associated team for financial support. Many thanks go to my jury members – Tomas Pajdla, Marc Pollefeys, Fredrik Kahl, and Bernd Sturmfels – for agreeing to evaluate my thesis. During these years, I was fortunate to work with great collaborators: Boris Bukh, John Canny, Yasutaka Furukawa, Xavier Goaoc, Alfredo Hubard, Kathlén Kohn, Brian Osserman, Bernd Sturmfels. I am grateful to all of them for teaching me so much; I especially thank Bernd Sturmfels for inviting me to MPI Leipzig, and Yasutaka Furukawa for inviting me to Washington University in St. Louis. A real turning point in my PhD has been the AIM Workshop on algebraic vision in San Jose (2016), and I would like to acknowledge the organizers Rekha Thomas, Max Lieblich, and Sameer Agarwal who invited me. I thank all my colleagues from Willow and Sierra for creating an amazing work environment, and for many great moments at conferences or at "Friday drinks". A special mention goes to my officemates Guilhem and Julia, for making afternoons and evenings in the office so fun and enjoyable (even under the pressure of deadlines). Finally, my biggest thanks go to my parents, Patrizia and Barry, and to Camilla. This thesis would not have been possible without their love and support.

Contents

List of Figures

List of Tables

Introduction

In order to develop algorithms for artificial visual systems, computer vision relies on *models* for representing and manipulating empirical data. In this sense, a model is an abstraction that enables one to study a certain class of physical processes using formal tools. Multi-view reconstruction, for example, can be framed as an inference task, where the set of admissible camera configurations is associated with points in some parameter space, and observations (multi-view correspondences) are used to identify an element within this parameterized set. Models are of course necessary for a theoretical investigation of empirical systems: simplifying the physical reality is a prerequisite for developing theories that can be applied in different contexts.

This thesis is devoted to models for the geometry of vision. The mathematical study of the visual world dates back to ancient times, but the theoretical foundations of computational algorithms in 3D vision were developed mainly in the 80s and 90s. The theory, known as *multi-view geometry* [53, 78], has been essential in the development of many remarkable applications, from 3D reconstruction software to today's technologies for augmented and virtual reality. Our goal in this thesis is to take the classical theory of multi-view geometry further, by developing more general models for *cameras* and *3D shapes*, and by studying these models using tools from *projective* and *algebraic geometry* [73].

The importance of projective geometry as a framework for modeling visual phenomena has been long known. Projective transformations and ambiguities arise naturally in reconstruction tasks, and the use of homogeneous coordinates simplifies many geometric computations (in particular, it linearizes the pinhole camera model). For these reasons, multi-view geometry is usually based on the perspective that affine and euclidean geometry are special cases of projective geometry. This idea dates back to Felix Klein's *Erlangen program*, and is convenient for dealing with a hierarchy of geometric models in a unified setting. In this presentation, however, we will usually disregard affine and euclidean structures, and focus for the most part on actual projective geometry. This is because many basic concepts in vision only require simple contact properties (collinearity, incidence or tangency); hence, a purely projective perspective is simpler and at the same time more general. We will also point out how some models that have traditionally

been studied in a euclidean setting (*e.g.*, *visual hulls* [12] and *visual events* [101]), can actually be defined more naturally using only projective geometry.

Our use of tools from algebraic geometry is due to the fact that many of our models will be based on polynomial representations. For example, it is well known that multi-view point correspondences are characterized by algebraic constraints in image coordinates [54], and it is actually convenient to view these conditions as the defining equations of an an algebraic variety (known as the *joint image* [190], or the *multi-view variety* [4]). An algebraic framework is also useful for modeling non-central imaging systems [173]. We will in fact represent general cameras either geometrically as algebraic surfaces in line space (corresponding to families of viewing rays), or analytically as rational (polynomial) mappings $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$. Furthermore, we will study projections of algebraic surfaces, describing the *visual events* that occur in the image contour curve as the viewpoint changes.

As mentioned above, the topics addressed in this thesis are intended to broaden the scope of the classical theory of geometric vision. For example, while traditional multi-view geometry is largely focused on systems of $2, 3, 4$ cameras and the associated multifocal tensors, we believe that it is important (and interesting) to study configurations involving any number of cameras, investigating concepts such as the "joint image" or "viewing graphs". Moreover, other topics, such as non-central imaging systems or the geometry of shape reconstruction, have arguably not been studied much in a general and systematic fashion, but deserve to be part of a mathematical theory of vision. We will also often insist on the importance of *line geometry* [150] for describing visual processes. Essentially all geometric aspects in vision are in fact determined by how lines (*i.e.*, light rays) intersect objects in space. Although the role of viewing lines is usually hidden in practical structure-from-motion algorithms by the use of image coordinates [133], in many situations it can be convenient to make the geometry explicit. For the study of non-central imaging systems, in particular, researchers have proposed a geometric representation of cameras as mappings from points to viewing rays [145, 136]. This idea will be developed further in the thesis, and we will investigate in detail the geometry of non-linear cameras inside the Grassmannian of lines.

The mathematical models studied in this thesis are quite general, but are still meant to be close to practical applications. Our presentation avoids unnecessary technicalities, and only parts of our discussion will require some background in algebraic geometry. On the other hand, we believe that a minimal amount of rigorous language can be useful, since it allows us to use more powerful tools, and because it helps us identify more clearly what is known and the problems that still need to be addressed. We also hope that precise terminology can facilitate interactions between researchers across different fields. There has been in fact a recent interest in geometric vision from mathematicians, with several journal papers (including some of the works presented in this thesis) and a few workshops on the topic of *algebraic vision*. We hope that this thesis can help bridge the gap between theoretical and practical approaches to vision.

Thesis summary

The following is a brief description of the main contributions of the thesis. The relevant previous literature for each topic is discussed at the beginning of the corresponding chapter.

Chapter 1 is devoted to the multi-view geometry of traditional pinhole cameras. We study the relationship between the joint image and camera configurations, explaining in particular the distinction between sets of multi-view constraints that can be used to characterize point correspondences, and those that are sufficient for determining camera geometry. Moreover, we investigate "solvable" viewing graphs, which describe subsets of fundamental matrices that identify a unique global configuration of cameras.

Chapter 2 describes an abstract model for general imaging systems as two-dimensional families of lines, or line congruences. We focus in particular on congruences of order one, which can be used to define rational mappings from points and lines. We present a complete classification of all such rational geometric cameras, based on a classical result due to Kummer. We also characterize algebraically the set of n-tuples of lines that meet at a point, and we apply this result to describe the multi-view geometry of general systems of cameras.

Chapter 3 extends the geometric camera model introduced in Chapter 2 by describing how lines in a congruence can be associated with image coordinates. After a general discussion, we present an in-depth study of two-slit cameras, that we model analytically as a bilinear map from $\mathbb{P}^3$ to $\mathbb{P}^1 \times \mathbb{P}^1$. We also introduce generalizations of several classical features of pinhole cameras, including calibration matrices and multifocal tensors.

Chapter 4 investigates geometric relationships between sets in $\mathbb{P}^3$ and their projections in different images. In particular, we describe from different perspectives the "geometric consistency" conditions that image sets must satisfy in order to be projections of a single object in space. Conversely, the study of 3D objects that project onto a given set of images leads to the traditional notion of "visual hull". By understanding the 3D geometry of shapes from within the joint-image, we obtain a simple image-based procedure for computing a boundary representation of the visual hull.

Chapter 5 discusses the visual events generated by smooth algebraic curves and surfaces in $\mathbb{P}^3$. We spell out new geometric characterizations of these events, which are based on the study of lines in $\mathrm{Gr}(1, \mathbb{P}^3)$ and planes in $(\mathbb{P}^3)^*$ that meet the given curve or surface in exceptional ways. We also present some effective strategies for recovering the visual events computationally.

The thesis also includes two short appendices that summarize some useful background on projective and line geometry (Appendix A) and on projective algebraic varieties (Appendix B).

Chapter 1

Pinhole Cameras and their Multi-View Geometry

This chapter is devoted to the classical *pinhole camera model*. Our goal is to provide a clear overview of some selected topics in the theory of pinhole cameras, in the language of projective geometry and elementary algebraic geometry. In particular, we hope to give definitive answers to the problem of characterizing the *joint image* formed by image correspondences, while also spelling out its relationship with the space of *camera configurations*.

The material in this chapter is based on the following publications:

- Matthew Trager, Martial Hebert, and Jean Ponce. "The joint image handbook". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2015, pp. 909–917.
- Matthew Trager, Jean Ponce, and Martial Hebert. "Trinocular Geometry Revisited". In: *International Journal of Computer Vision* (2016), pp. 1–19.
- Brian Osserman and Matthew Trager. "Multigraded Cayley-Chow forms". In: *arXiv preprint arXiv:1708.03335* (2017). Submitted to *Advances in Mathematics*.
- Matthew Trager, Brian Osserman, and Jean Ponce. "On the Solvability of Viewing Graphs". In: *European Conference on Computer Vision*. 2018.

1.1 Introduction

Pinhole-based imaging systems have a very long history. The optical phenomenon of the *camera obscura* was accurately described by the 11th-century Arab physicist Ibn al-Haytham (965–1039 CE) and was mentioned already in writings by Aristotle (384-322 BCE) [49]. Throughout the Renaissance and until the 18th century, devices that

formed images by gathering light through a small aperture were used as drawing aids or for entertainment. The photographic camera, developed in the early 19th century, is an adaptation of the box-type camera obscura that was popular at that time. Today's digital cameras use complicated optics to gather and focus light, but can still be roughly modeled using pinhole geometry.

In mathematics, *projective geometry* emerged from the study of perspectivity, initiated by architects Filippo Brunelleschi and Leon Battista Alberti around 1425. The theory flourished in the 17th, 18th, and 19th centuries, with fundamental contributions from Desargues, Poncelet, and Klein, among many others. Problems in projective reconstruction were already studied by these classical geometers. For example, epipolar geometry was known to Hauck in 1883, and image-based 3D reconstruction was described by Finsterwalder in 1889 [171]. These results were largely re-developed by computer vision scientists in the eighties and nineties (even though they were known to photogrammeters [183]). Today, the part of computer vision that studies the theoretical foundations 3D reconstruction algorithms is known as *multi-view geometry*. General overviews on this subject can be found in the textbooks by Faugeras, Luong and Papadopoulo [53], and by Hartley and Zisserman [78], or in the older book by Maybank [124].

Despite this long history, many aspects of the theory of pinhole cameras are not fully settled. The most well-known results in multi-view geometry are concerned with multifocal tensors for 2, 3, or 4 views [120, 164, 75, 74]. Point correspondences for arbitrary numbers of views have also been characterized [54, 190, 84], however results are often scattered in the literature, and they sometimes actually contradict each other [54, 78, 188]. Moreover, basic theoretical facts are not always common knowledge among specialists. For example, many practitioners today would probably be hard pressed to answer simple questions such as how many multilinear relations (and which ones) are necessary to characterize correspondences, or to determine the corresponding camera parameters. It is also not difficult to stumble upon open problems that have not been addressed in the literature (see for example Section 1.4). Partly because of the supply of such problems, there has been a recent interest in computer vision among mathematicians, with the emerging field of *algebraic vision* [4, 3]. As explained in the introduction to the thesis, our work often lies at the interface between vision and mathematics: our goal is to investigate problems in vision using mathematical language, but without losing focus of the underlying practical motivations.

One of our main objects of study is the *joint image*, introduced by Triggs [190] and studied independently by Heyden and Åström [84], and by Aholt, Sturmfels and Thomas [4]. The joint image for n cameras is a subvariety in $(\mathbb{P}^2)^n$ defined as the closure of the set of all n-tuples of point correspondences. Thus, algebraic characterizations of the joint image are equivalent to the constraints that describe multi-view correspondences. In addition, the joint image is a useful conceptual tool, since it can be seen as an (almost) exact replica of 3D-space "distributed" across multiple images. A joint image also uniquely represents a configuration of cameras so, for example, we can interpret the task

of 3D reconstruction as the problem of interpolating a joint image from image data (this is similar to the use of multi-view tensors, but for any number of views). We present in Section 1.3 a detailed description of the joint image, based on our paper [186]. In particular, we give a series of results that explain which sets of multilinear constraints can be used for characterizing point correspondences, or for recovering a camera configuration.

Among the results that we discuss in Section 1.3 is the (previously known) fact that fundamental matrices are sufficient for recovering a camera configuration, assuming that the pinholes of the cameras are not all aligned. On the other hand, it is not necessary to use the fundamental matrices among all pairs of cameras, and this leads to the natural question of describing how many fundamental matrices, and which ones, are actually needed. To study this problem, we use the notion of a *viewing graph*, introduced by Levi and Werman in [113]. This is a graph in which edges represent fundamental matrices among pairs of cameras. We say that a graph is "solvable" when the associated fundamental matrices are (generically) sufficient to determine a global camera configuration. Despite its clear significance, the problem of characterizing which viewing graphs are solvable has received very little attention (mainly in [113, 157]). In Section 1.4 we present results from our paper [187], providing several new criteria that can be used to verify whether a set of fundamental matrices determines a camera configuration.

At the end this chapter we take a closer look at the geometry of three views (Section 1.5). For example, following our paper [188], we describe a special class of trilinearities (previously used by Ponce, Papadopoulo, Teillaud and Triggs in [144]) that have an interpretation in terms of transversals of visual rays, and provide a more geometric approach for characterizing point correspondences among three views. We then turn to trifocal tensors, and show that their relationship with the joint image can be explained using a theory of "multi-graded Chow forms" that we proposed in [131]. In brief, trifocal tensors (and similarly quadrifocal tensors for four views) can be seen as an example of a general strategy for encoding complicated algebraic varieties (in this case, the joint image) using a single polynomial form. Finally, we present some experiments from [188], where we used trilinear constraints to define a "trinocular-epipolar error" that can be minimized to recover camera parameters for three views.

1.1.1 Previous Work

Multiple-view geometry has been studied in computer vision since the seminal work of Longuet-Higgins, who proposed in 1981 the essential matrix for pairs of calibrated cameras [119]. Its uncalibrated counterpart, the fundamental matrix, was introduced by Luong and Faugeras [120]. The trilinear constraints associated with three views of a straight line were discovered by Spetsakis and Aloimonos [169] and by Weng, Huang and Ahuja [192]. The uncalibrated case was tackled by Shashua [164] and by Hartley [75], who coined the term "trifocal tensor". The quadrifocal tensor was introduced by

Triggs [190], and Faugeras and Mourrain gave a simple characterization of all multilinear constraints associated with multiple perspective images of a point [54]. A different formulation for trilinear constraints based on line geometry was introduced in [144].

Closely related to our work is that of Heyden and Åström in [84], who also study multi-view constraints from the point of view of the joint image (which they refer to as the "natural descriptor"). The authors also point out an interesting property of the epipolar constraints for three cameras in general position: these conditions uniquely determine a camera configuration *however* the trilinear conditions do not follow algebraically from the bilinear ones and, in fact, bilinear constraints are not sufficient for characterizing point correspondences in general. We discuss this somewhat paradoxical behavior in a more general setting in Section 1.3 of this chapter.

Aholt, Sturmfels and Thomas describe many algebraic and combinatorial properties of the joint image (in their terminology, the "multi-view variety") [4]. They show for example that traditional bilinear, trilinear and quadrilinear constraints are generators, and form in fact a "universal Gröbner basis", for the polynomial ideal associated with the joint image (see Theorem 1.3.10). They also prove other technical results, such as the fact that the set of all joint images forms a distinguished component in the "Hilbert scheme" parameterizing all varieties with assigned Hilbert function.

The first investigation of viewing graphs and their solvability can be found in [113]. In that work, Levi and Werman characterize all solvable viewing graphs with at most six vertices, and discuss a few larger solvable examples. Although they provide some useful necessary conditions (see our Proposition 1.4.6 and Example 1.4.13), they do not address the problem of solvability in general. In [157], Rudi, Pizzoli and Pirri also consider viewing graphs, studying mainly whether a configuration can be recovered from a set of fundamental matrices using a linear system. They also present some "composition rules" for merging solvable graphs into larger ones. In [186], we provided a sufficient condition for solvability using $2n-3$ fundamental matrices, and pointed to a possible connection with "Laman graphs" and graph rigidity theory. Indeed, Özyesil and Singer [132] show that if one uses *essential* matrices instead of fundamental ones then solvability can be characterized in terms of so-called "parallel-rigidity" for graphs. Their analysis however does not carry over to the more general setting of uncalibrated cameras. Finally, the viewing graph has also been considered in more practical work, particularly to enforce triple-wise consistency among fundamental matrices before estimating camera parameters: this is done for example by Sinha and Pollefeys [166] and by Sweeney, Sattler, Hollerer, Turk and Pollefeys [177].

1.1.2 Main contributions

Our main contributions can be summarized as follows.

- **The joint image.** We make the distinction between ideal-theoretic, set-theoretic,

and "weak" characterizations (our terminology) of the joint image, and use it to explain the difference between multi-view constraints that determine camera geometry, and those that directly describe point correspondences. In particular, we clarify whether bilinear, trilinear or quadrilinear constraints are necessary or sufficient for these tasks, under different assumptions on the configurations of pinholes (Proposition 1.3.19). We also give a simple proof of the important result that a joint image characterizes a configuration of cameras (Theorem 1.3.16).

- **The viewing graph.** We show that the minimum number of fundamental matrices that can be used to recover a configuration of n cameras is always $\lceil (11n-15)/7 \rceil$ (Theorem 1.4.9). We also present several new criteria for deciding whether or not a viewing graph is solvable (Theorems 1.4.12 and 1.4.14), and we describe a simple linear test that can be used to verify whether a viewing graph identifies a finite number of camera configurations (Section 1.4.5).

- **Three-View geometry and Multifocal Tensors.** We apply our analysis on the different types of characterizations of the joint image to the important case of three views. For example, we show that the nine trilinearities encoded in a trifocal tensor are not sufficient to completely ensure correspondence among three views (Proposition 1.5.7), although they can be used to recover the corresponding projection matrices. We also propose a new framework for understanding trifocal and quadrifocal tensors, based on the joint image and the theory of multi-graded Chow forms [131].

Before addressing our main topics, we present in Section 1.2 a general introduction to the pinhole camera model, and to linear projections in general. We review for example the action of "world" and "image" coordinate changes, orbits under euclidean and affine motions, and projections and inverse projections of linear subspaces. These discussions will be useful later in the thesis.

Conventions. Throughout the chapter, we write $\mathbb{P}^n = \mathbb{P}(\mathbb{R}^{n+1})$ for the for the n-dimensional projective space over $\mathbb{R}$. It is worth noting that most of our definitions from Section 1.2 can be given more generally for projective spaces $\mathbb{P}(\mathbb{K}^{n+1})$ over an arbitrary field $\mathbb{K}$. In the next chapters, for example, we will sometimes work over the field of complex numbers $\mathbb{K} = \mathbb{C}$. Since physical three-space is usually identified with $\mathbb{R}^3$, we focus on real spaces here.

We use bold font for vectors and matrices, and normal font for projective objects. For example, a point in $\mathbb{P}^3$ will be written as $p = [\mathbf{p}]$ where $\mathbf{p}$ is a vector in $\mathbb{R}^4$ and p is the equivalence class associated with $\mathbf{p}$. Similarly, a projective transformation represented by a matrix $\mathbf{M}$ will be written as $M = [\mathbf{M}]$.

1.2 Preliminaries

This section summarizes some basic aspects of the theory of pinhole cameras. The material we discuss is well-known, and can be found in textbooks such as [53, 78], although our approach is sometimes different.

The organization of the section is as follows. Section 1.2 is an overview of the pinhole camera model; Section 1.2.2 introduces more formally linear projections in arbitrary spaces; in Section 1.2.3 we describe the group actions of "world" and "image" coordinate transformations; we classify orbits of cameras under euclidean and affine motions in Section 1.2.4; finally, in Section 1.2.5, we use Plücker coordinates to represent projections and inverse projections of linear spaces in arbitrary dimensions.

1.2.1 The pinhole model

The pinhole camera model used in computer vision is a mathematization of the geometry of the *camera obscura*. This is illustrated in Figure 1.1. The "imaging device" consists of a *pinhole* c that collects light rays and a *retinal plane* H where the image is formed. A scene point x is projected onto the image point y, obtained as the intersection of H with unique line through through x and c.

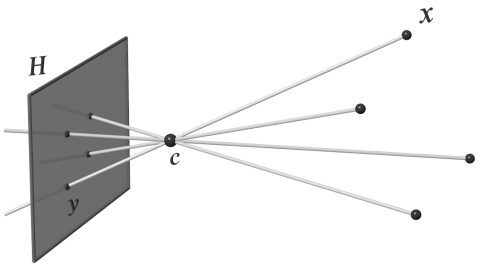


Figure 1.1: The geometry of the camera obscura.

This map is easily described in the language of projective geometry. If we identify the three-dimensional space with $\mathbb{P}^3$, and fix a point $c = \mathbb{P}(K)$ for the pinhole and a plane $H = \mathbb{P}(W)$ not containing c for the retinal plane (so that K, W are vector subspaces in $\mathbb{R}^4$ of dimensions 1 and 3), we consider the unique vector projection $\mathbb{R}^4 = W \oplus K \to W$ onto W with null space K. Then the induced map on projective spaces $\mathbb{P}^3 \setminus \{c\} \to H$ is a mathematical model for the camera obscura shown in Figure 1.1.

A more concrete description can be given in terms of projective coordinates. If we fix a projective reference frame in three-space and a projective reference frame in the retinal plane, we can describe the action of a pinhole camera using a 3×4 matrix of full rank, uniquely determined up to scaling. This is usually known as a *projection matrix*. This analytic representation is very convenient, and widely used in computer vision. However

it is important to emphasize some differences between the projection matrix and the geometric model described above.

- There is no unique projection matrix associated with a fixed "camera obscura" with pinhole c and retinal plane H, since the analytic description depends on the choice of coordinates in the image and in space. In general, it is customary to assume a fixed reference frame in three-space, despite the fact that there is no distinguished coordinate system the physical 3D world. For the retinal plane, on the other hand, there is often a natural choice of coordinates, defined by the pixel grid in a digital camera.

- Conversely, given a projection matrix, it is not possible to recover the retinal plane (while this is possible for the pinhole, which corresponds to the null-space of the matrix). Indeed, the image $\mathbb{P}^2$ of the projection is now an abstract plane that does not correspond to any particular embedded plane in space. Moreover, for any plane in $\mathbb{P}^3$ not containing the pinhole, there exists a unique choice of coordinates so that the corresponding projection is described by the original matrix. The equivalence between all retinal planes follows from the "perspectivity" homography induced by the pinhole.

In light of these ambiguities, it is tempting to exclude the role of retinal plane from the whole imaging process, and to view a pinhole projection simply as a mapping from points into the *bundle of lines* that pass through the pinhole. In this setting, the projection matrix is equivalent to specifying the pinhole together with a reference frame on the corresponding line bundle. This abstract representation of a camera as a mapping from points to lines will be studied in Chapter 2. In the remaining part of this chapter, we focus entirely on the more traditional and "concrete" coordinate-based model that uses projection matrices.

1.2.2 Projections in projective spaces

We next define linear projections in general projective spaces. Although we are mainly interested in projections from $\mathbb{P}^3$ to $\mathbb{P}^2$, it is actually natural to extend many properties of systems of cameras to the case of projections in spaces of arbitrary dimensions. It is not our purpose to develop such a general theory, and in the rest of the chapter we focus mainly on pinhole projections. Nevertheless, it is useful to give definitions in a more general setting, since projections in different dimensions will be sometimes used in later chapters.

By *linear projection* we always mean the projectivization of a surjective linear map on vector spaces. More precisely, if a linear map $\mathbb{R}^{n+1} \to \mathbb{R}^{m+1}$ $(n > m)$ is described

by an $(m+1) \times (n+1)$ matrix $\mathbf{A}$ of full rank, the corresponding linear projection is given by

$$\begin{aligned} \mathbb{P}^n \setminus \mathbb{P}(\mathrm{Ker}\mathbf{A}) &\to \mathbb{P}^m \\ [\mathbf{v}] &\mapsto [\mathbf{A}\mathbf{v}]. \end{aligned} \tag{1.1}$$

The subspace $K = \mathbb{P}(\mathrm{Ker}\mathbf{A})$ inside $\mathbb{P}^n$ is the *center* of the projection, and has projective dimension $n - m - 1$. In the case of cameras ($n = 3$, $m = 2$), the center of projection is simply the pinhole. It is clear that two matrices induce the same linear projection if and only if they are related by a non-zero scalar factor. Just as pinhole cameras, any linear projection can be realized geometrically by fixing a center K of dimension $n - m - 1$ and another subspace H in $\mathbb{P}^n$ of dimension m disjoint from K. These spaces determine a mapping from $\mathbb{P}^3$ to H defined by $p \mapsto (K \vee p) \wedge H$ that can be described analytically by (1.1). See Figure 1.2.

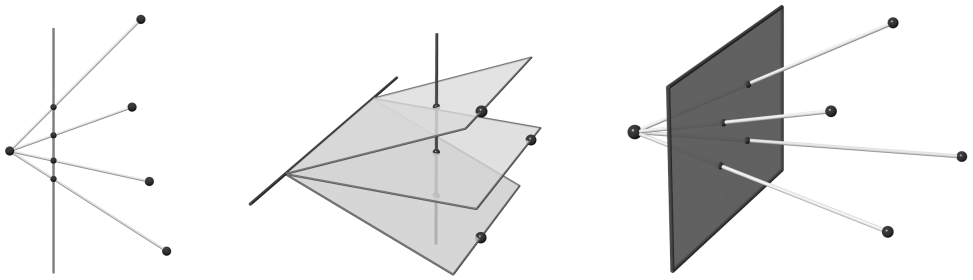


Figure 1.2: Linear projections $\mathbb{P}^2 \dashrightarrow \mathbb{P}^1$ (*left*), $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1$ (*center*), and $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ (*right*). Note that the "retinal space" is a line in the first two cases (shown in dark).

It will often be convenient to write a linear projection as a map $\mathbb{P}^n \dashrightarrow \mathbb{P}^m$. Here the dashed arrow indicates a "rational map", *i.e.*, an algebraic map defined on a dense open set of $\mathbb{P}^n$ (cf. Appendix B). This simplifies notation, since we are not required to exclude the center of projection from the domain. We will also write $P = [\mathbf{A}]$, when the linear projection $P : \mathbb{P}^n \dashrightarrow \mathbb{P}^m$ is induced by the matrix $\mathbf{A}$.

1.2.3 Changes of coordinates

We briefly describe of the action of "changing coordinates" on projection mappings. This analysis will be useful for dealing with camera configurations. A similar study can be carried out for projections in arbitrary projective spaces, but we only consider pinhole cameras here. We write $\mathcal{P}$ for the family of all pinhole cameras, that we represent using projection matrices:

$$\mathcal{P} = \{P = [\mathbf{A}],\ \mathbf{A} \in \mathbb{R}^{3\times 4} \text{ of full rank}\} \subset \mathbb{P}^{11}.$$

There are two natural group actions on $\mathcal{P}$: the group $GL(4,\mathbb{R})$ acts by multiplication on the right (changes of "world" coordinates), while the group $GL(3,\mathbb{R})$ acts by multiplication on the left (changes of "image" coordinates).[1]

Changes of world coordinates. Given pinhole camera $P = [\mathbf{A}]$ where $\mathbf{A}$ is a 3×4 matrix, we consider the action of multiplying $\mathbf{A}$ on the right by matrices in $GL(4,\mathbb{R})$. Clearly, this can be interpreted as a change of coordinates of $\mathbb{P}^3$, but also as applying a projective "camera motion" to P.

We first observe that the action of $GL(4,\mathbb{R})$ on $\mathcal{P}$ is *transitive*. This means that, given any two cameras $P_1 = [\mathbf{A}_1]$ and $P_2 = [\mathbf{A}_2]$, there always exist a change of coordinates (or a camera motion) that maps P_1 to P_2. Indeed, if $c_1 = [\mathbf{c}_1]$ is the pinhole of P_1, we have that any invertible matrix of the form

$$\mathbf{T}_{\alpha,\mathbf{v}} = \alpha \mathbf{A}_\mathbf{1}^\dagger \mathbf{A}_\mathbf{2} + \mathbf{c}_\mathbf{1}\mathbf{v}^\mathbf{T},$$

where $\mathbf{A}_1^\dagger = \mathbf{A}_1^T(\mathbf{A}_1\mathbf{A}_1^T)^{-1}$ (a pseudo-inverse for $\mathbf{A}_1$), $\alpha \in \mathbb{R}$, and $\mathbf{v} \in \mathbb{R}^4$, is such that $[\mathbf{A}_1\mathbf{T}_{\alpha,v}] = [\mathbf{A}_\mathbf{2}]$.

The action of $GL(4,\mathbb{R})$ on $\mathcal{P}$ however is not free. In other words, for any camera $P = [\mathbf{A}]$, there exist projective transformations of $\mathbb{P}^3$ that do not affect P. These transformations form the *stabilizer group*:

$$\mathrm{Stab}_{GL(4,\mathbb{R})}(P) = \{\alpha \mathbf{I}_4 + \mathbf{c}\mathbf{v}^T \mid \alpha \in \mathbb{R} \setminus \{0\}, \mathbf{v} \in \mathbb{R}^4\} \cap GL(4,\mathbb{R}), \tag{1.2}$$

where $c = [\mathbf{c}]$ is the pinhole of P, and $\mathbf{I}_4$ denotes the 4×4 identity matrix. Indeed, all the solutions for $\mathbf{T}$ in $\mathbf{AT} = \alpha\mathbf{A}$ are described by (1.2). As projective transformations, elements in $\mathrm{Stab}_{GL(4,\mathbb{R})}(P)$ "shift" points along viewing lines through c. In fact, it is easy to see that the stabilizer only depends on the pinhole c and not on the actual projection matrix.

We also mention that there is another natural group associated with a camera P with center $c = [\mathbf{c}]$, namely

$$\mathrm{Stab}_{GL(4,\mathbb{R})}(c) = \{\mathbf{T} \mid \mathbf{Tc} = \alpha\mathbf{c}\} \subset GL(4,\mathbb{R}).$$

This corresponds to projective transformations of $\mathbb{P}^3$ that fix the pinhole c. It is clear that $\mathrm{Stab}_{GL(4,\mathbb{R})}(P)$ is a subgroup of $\mathrm{Stab}_{GL(4,\mathbb{R})}(c)$. Moreover, elements in $\mathrm{Stab}_{\mathrm{GL}(4,\mathbb{R})}(c)$ can actually be viewed as changes of *image* coordinates, since they induce projective transformations on the bundle of lines through c. We now address changes of image coordinates directly.

[1] For concreteness, we consider the actions of $GL(4,\mathbb{R})$ and $GL(3,\mathbb{R})$ rather than of $\mathbb{P}GL(4,\mathbb{R})$ and $\mathbb{P}GL(3,\mathbb{R})$. Clearly, matrices that are scalar factors of each other will act the same way.

Changes of image coordinates. Given pinhole camera $P = [\mathbf{A}]$, we consider the action of multiplying $\mathbf{A}$ on the left by matrices in $GL(3,\mathbb{R})$. This action is *not* transitive: if $c = [\mathbf{c}]$ is the pinhole of P, the *orbit* of P under $GL(3,\mathbb{R})$ (*i.e.*, the set of cameras that can be obtained from P by changes of image coordinates) is

$$\mathrm{Orb}_{\mathrm{GL}(3,\mathbb{R})}(P) = \{Q = [\mathbf{B}] \mid \mathbf{Bc} = \mathbf{0}\} \subset \mathcal{P}.$$

This is simply saying that two cameras are related by a change of image coordinates if and only if they have the same pinhole. On the other hand, the *stabilizer group* of this action, only contains trivial projective transformations:

$$\mathrm{Stab}_{GL(3,\mathbb{R})}(P) = \{\alpha\mathbf{I}_3, \mid \alpha \in \mathbb{R} \setminus \{0\}\} \subset GL(3,\mathbb{R}).$$

In more concrete terms, non-trivial projective changes of coordinates in the image will always affect a camera projection.

1.2.4 Geometries in projective space and orbits of cameras

This chapter (and this thesis) focuses mainly on projective camera models. In what follows, however, we give an account of euclidean and affine cameras, viewed as orbits of special subgroups of projective transformations. This serves mostly as reference for when we will generalize these concepts for non-central imaging systems (in Chapter 3).

We first briefly recall the hierarchy of different geometries that can be studied within the projective framework. This viewpoint was set forth by Felix Klein in his famous Erlangen program (1872), and is of great practical importance in computer vision. In this setting, a "geometry" can be seen as the study of some space and its "invariant properties" under the action of a group of transformations. Projective geometry, in particular, studies projective space $\mathbb{P}^n$ up to projective transformations, that we identify with the group of $(n+1)\times(n+1)$ invertible matrices defined up to scale. Among these transformations, we consider subgroups of matrices of the following form (defined up to scale):

$$\begin{aligned}
\mathrm{Iso}(n) &= \left\{\mathbf{T} = \begin{bmatrix}\mathbf{R} & \mathbf{t}\\ \mathbf{0}^T & 1\end{bmatrix} \text{ with } \mathbf{R} \in SO(n,\mathbb{R})\right\},\\
\mathrm{Sim}(n) &= \left\{\mathbf{T} = \begin{bmatrix}k\mathbf{R} & \mathbf{t}\\ \mathbf{0}^T & 1\end{bmatrix} \text{ with } k \in \mathbb{R},\ \mathbf{R} \in SO(n,\mathbb{R})\right\},\\
\mathrm{Aff}(n) &= \left\{\mathbf{T} = \begin{bmatrix}\mathbf{M} & \mathbf{t}\\ \mathbf{0}^T & 1\end{bmatrix} \text{ with } \mathbf{M} \in GL(n,\mathbb{R})\right\},
\end{aligned}$$

so that

$$\mathrm{Iso}(n) \subset \mathrm{Sim}(n) \subset \mathrm{Aff}(n) \subset GL(n+1,\mathbb{R}).$$

Here $SO(n, \mathbb{R})$ denotes the special orthogonal group (matrices with unit determinant). The groups of transformations $\mathrm{Iso}(n)$, $\mathrm{Sim}(n)$ $\mathrm{Aff}(n)$ are respectively associated with *euclidean*, *similarity* and *affine* geometries. Indeed, if we identify $\mathbb{R}^n$ with the chart in $\mathbb{P}^n$ of points of the form $[x_1, \ldots, x_n, 1]$, then these groups coincide with the classical definitions of euclidean, similarity, and affine transformations in $\mathbb{R}^n$ (with respect to the standard euclidean metric). Each these geometries has its associated invariants, and smaller transformation groups correspond to more properties that are preserved.

Example 1.2.1. An important invariant for computer vision is the *absolute quadric*. This is a complex degenerate quadric in $\mathbb{P}^n$ defined by $\Omega = \{x_{n+1} = x_1^2 + \ldots + x_n^2 = 0\}$. This quadric is invariant under all similarity transformations. In fact, a projective transformation is a similarity if and only if it preserves (not necessarily pointwise) the absolute quadric. ♢

We now describe the action of these nested subgroups of transformations on projection mappings. We consider only pinhole camera projections $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$, although a similar analysis could be carried out more generally.

We have already noted that the action of $GL(4, \mathbb{R})$ on the space of cameras can be interpreted either as applying "camera motions", or as applying "change of coordinates" in $\mathbb{P}^3$. In particular, the orbit of a camera P under the action of one of the projective subgroups is the set of cameras that are equivalent to P for the corresponding geometry (in fact, one might say that cameras in the same orbit are geometrically indistinguishable, since a world reference frame is not physically defined). The equivalence of cameras for different geometries leads to a natural "taxonomy" of cameras. Although a complete classification would probably not be very difficult, we describe here only generic orbits that correspond to camera models actually used in computer vision.

Projective geometry. We have seen in Section 1.2.3 that all pinhole cameras are equivalent up to projective transformations, so they form a unique projective orbit. We can thus choose any camera as a representative for the class, for example the camera P associated with the 3×4 matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{1.3}$$

We sometimes refer to this as the *standard pinhole projection*.

Affine geometry. Let $P = [\mathbf{A}]$ be camera with $\mathbf{A} = [\mathbf{Q} \,|\mathbf{b}]$ where $\mathbf{N}$ is a 3×3 matrix. Applying to P an affine transformation described by a matrix of the form

$$\mathbf{T} = \begin{bmatrix} \mathbf{M} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix},$$

we obtain the camera $[\mathbf{QM} \,|\, \mathbf{Qt} + \mathbf{b}]$. This easily implies that cameras of the form $[\mathbf{Q} \,|\mathbf{b}]$, where $\mathbf{Q}$ is an invertible 3×3 matrix, are all equivalent under affine transformations,

so they form a single affine orbit. This is the set of *finite cameras*, *i.e.*, cameras whose pinhole does not lie on the plane at infinity. Note that finite cameras are dense among all pinhole cameras, since for a generic 3×4 matrix, the left 3×3 submatrix will be invertible. The standard pinhole projection (1.3) can be taken as a representative for the affine model of finite cameras.

Among non-finite cameras whose pinhole lies at infinity, a distinguished family is associated with 3×4 matrices whose last row can be scaled to $[0, 0, 0, 1]$. These are known as *affine cameras*, since they can be described as affine projections using non-homogeneous coordinates. It is easy to see that affine cameras are all equivalent under affine transformations. We can choose as representative for this orbit the camera defined by

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{1.4}$$

There exist other affine orbits among non-finite cameras, but their use in computer vision is very limited.

Similarity geometry. The orbit under similarities of the standard pinhole camera (1.3) is the set of cameras associated with matrices of the form $[\mathbf{R} \,|\, \mathbf{t}]$ where $\mathbf{R}$ is an orthogonal matrix. These are sometimes called *normalized* or *calibrated cameras*. More generally, it is easy to characterize similarity orbits among all finite cameras. If $[\mathbf{M} \,|\, \mathbf{b}]$ is a finite camera, then using QR-decomposition of matrices we can uniquely write

$$[\mathbf{M} \,|\, \mathbf{b}] = \mathbf{K}[\mathbf{R} \,|\, \mathbf{t}], \tag{1.5}$$

where $\mathbf{R}$ is an orthogonal matrix and $\mathbf{K}$ is upper triangular with positive diagonal elements. The matrix $\mathbf{K}$ in this decomposition is invariant to similarity transformations (up to scaling, but projection matrices are independent of scale). Moreover, any two matrices whose decomposition yields the same matrix $\mathbf{K}$ up to scale are related by a similarity transformation. In other words, the entries of $\mathbf{K}$ up to scale identify the similarity orbit of a finite camera. The matrix $\mathbf{K}$ is known as the *calibration matrix* and its six entries defined up to scale are the cameras *intrinsic parameters*. These have geometric interpretations (x-scale, y-scale, skew factor, coordinates of the principal point) that describe a projective transformation in $\mathbb{P}^2$ that needs to be applied to standard model (1.3) to recover the similarity orbit of the given camera.

A similar analysis can be carried out for orbits among affine cameras. Any projection matrix of an affine camera can be uniquely decomposed as

$$\begin{bmatrix} \mathbf{M}_{2\times 3} & \mathbf{b} \\ \mathbf{0}^T & 1 \end{bmatrix} = \begin{bmatrix} \alpha_x^s & s^s & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} k\mathbf{r}_1^T & t_1 \\ k\mathbf{r}_2^T & t_2 \\ \mathbf{0}^T & 1 \end{bmatrix}, \tag{1.6}$$

where $\mathbf{r}_1, \mathbf{r}_2$ are orthonormal 3-vectors, and α_x is positive. The entries α_x^s, s^s are invariant to similarity transformations (*not* up to scaling) and in fact they identify the similarity orbit. When $\alpha_x^s = 1$ and $s^s = 0$, the camera belongs to the similarity orbit of (1.4). This type of camera is called a *scaled orthographic projection* (with factor k).

Euclidean geometry. The euclidean orbits among finite cameras *coincide with similarity orbits*. Indeed, applying to a finite camera $[\mathbf{M}|\mathbf{b}]$ a similarity transformation

$$\mathbf{T} = \begin{bmatrix} k\mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix},$$

yields a camera defined by $[k\mathbf{MR} \,|\, \mathbf{Mt} + \mathbf{b}]$. This projection matrix can be rescaled as $[\mathbf{MR} \,|\, \mathbf{Mt}' + \mathbf{b}]$, where $\mathbf{t}'$ is such that $\mathbf{Mt}' + \mathbf{b} = 1/k(\mathbf{Mt} + \mathbf{b})$ (recall that $\mathbf{M}$ has full rank). In particular, $[\mathbf{M}|\mathbf{b}]\mathbf{T}$ is equivalent to $[\mathbf{M}|\mathbf{b}]\tilde{\mathbf{T}}$, where $\tilde{\mathbf{T}}$ is a euclidean transformation:

$$\tilde{\mathbf{T}} = \begin{bmatrix} \mathbf{R} & \mathbf{t}' \\ \mathbf{0}^T & 1 \end{bmatrix}.$$

This is of course a way of explaining the well-known scale ambiguity in pictures taken with perspective pinhole cameras: a scaling a 3D scene by $\mathbf{T}$ has the same effect of applying a rigid camera motion $\tilde{\mathbf{T}}$. The fact that finite cameras are similarity-equivalent if and only if they are euclidean-equivalent means that the calibration matrix $\mathbf{K}$ from (1.5) also identifies euclidean orbits for finite cameras.

Interestingly, the situation is different for affine cameras: in particular, it is easy to see that the affine and euclidean orbits of (1.4) are not the same. A slight variation of the decomposition (1.6) yields

$$\begin{bmatrix} \mathbf{M}_{2\times 3} & \mathbf{b} \\ \mathbf{0}^T & 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & s & 0 \\ 0 & \alpha_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{r}_1^T & t_1 \\ \mathbf{r}_2^T & t_2 \\ \mathbf{0}^T & 1 \end{bmatrix}, \tag{1.7}$$

where $\mathbf{r}_1, \mathbf{r}_2$ are orthonormal 3-vectors, and α_x, α_y are positive. The entries α_x, α_y, s in this decomposition are invariant under euclidean transformations (but not under similarities). When $\alpha_x = \alpha_y = 1$ and $s = 0$, the camera belongs to the euclidean orbit of (1.4), and is called an *orthographic projection*.

1.2.5 Projections and inverse projections of subspaces

Using pinhole cameras, the projection of a general line in $\mathbb{P}^3$ is a line in $\mathbb{P}^2$. Moreover, the pre-image of a line in $\mathbb{P}^2$ is a plane in $\mathbb{P}^3$, and the pre-image of a point in $\mathbb{P}^2$ is a line in $\mathbb{P}^3$. This is a particular example of a general feature of linear projections in arbitrary projective spaces, namely that projections and inverse projections of linear

subspaces are again linear subspaces. All of these associations can be conveniently described using *Plücker coordinates* (see Appendix A). Although we are mostly interested in cameras $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$, it is useful to introduce these mappings first for general projections $\mathbb{P}^n \dashrightarrow \mathbb{P}^m$.

We consider a linear projection $P : \mathbb{P}^n \dashrightarrow \mathbb{P}^m$ with center K. If L is a k-dimensional linear subspace of $\mathbb{P}^n$ disjoint from K (which implies $k \leq m$), then the projection $M = P(L)$ is a k-dimensional linear space in $\mathbb{P}^m$. In particular, writing $\mathrm{Gr}(k, \mathbb{P}^r)$ for the *Grassmannian of k-linear subspaces* in $\mathbb{P}^r$, there is a map

$$\begin{array}{rcl} \mathrm{Gr}(k, \mathbb{P}^n) \setminus Z(k, K) & \to & \mathrm{Gr}(k, \mathbb{P}^m) \\ L & \mapsto & P(L), \end{array} \tag{1.8}$$

where $Z(k, K)$ is the set of k-spaces in $\mathbb{P}^n$ which meet K. We can give a concrete description of (1.8) using Plücker coordinates. Assuming that $P = [\mathbf{A}]$, this description is based on the *compound matrix* $C_{k+1}(\mathbf{A})$, whose elements are the $(k+1) \times (k+1)$ minors of $\mathbf{A}$ ordered lexicographically (this assumes Plücker coordinates are ordered lexicographically).

Proposition 1.2.2. *If the k-dimensional linear space L in $\mathbb{P}^n$ has Plücker coordinates $[\mathbf{v}_L]$ in $\mathbb{P}^{\binom{n+1}{k+1}-1}$, then its image $H = P(L)$ in $\mathbb{P}^m$ for the linear projection $P = [\mathbf{A}]$ has Plücker coordinates $[\mathbf{w}_H]$ in $\mathbb{P}^{\binom{m+1}{k+1}-1}$, where*

$$\mathbf{w}_H = C_{k+1}(\mathbf{A})\mathbf{v}_L.$$

In particular, the map (1.8) can be seen as a linear projection $P_k : \mathbb{P}^{\binom{n+1}{k+1}} \dashrightarrow \mathbb{P}^{\binom{m+1}{k+1}}$ where $P_k = [C_{k+1}(A)]$, restricted to the Grassmannian varieties. The center of P_k is the linear space generated by $Z(k, K)$ inside $\mathbb{P}^{\binom{n+1}{k+1}-1}$.

Proof. Let $\mathbf{V}_L$ be a matrix of size $(n+1) \times (k+1)$ whose columns are coordinate vectors spanning L. Since $H = P(L)$, we have that $\mathbf{W}_H = \mathbf{A}\mathbf{V}_L$ is an $(m+1) \times (k+1)$ matrix whose column vectors span H. By definition, Plücker vectors $\mathbf{v}_L$, $\mathbf{w}_H$ for L and H are given by $C_{k+1}(\mathbf{V}_L)$ and $C_{k+1}(\mathbf{W}_H)$. The first claim now follows from the multiplicativity property of compound matrices: if $\mathbf{M}_1, \mathbf{M}_2$ are arbitrary matrices, then $C_i(\mathbf{M}_1\,\mathbf{M}_2) = C_i(\mathbf{M}_1)C_i(\mathbf{M}_2)$ holds [86]. To argue that $Z(k, K)$ generates the center of P_k, we note that $Z(k, K)$ is contained in the center, and moreover $\dim \mathrm{Span}(Z(k, K)) = \binom{n+1}{k+1} - \binom{m+1}{k+1} - 1$ inside $\mathbb{P}^{\binom{n+1}{k+1}-1}$. □

To describe pre-images of linear spaces, it is convenient introduce a "dual" version of a linear projection, which we refer to as a *linear embedding*. This is a map $J : \mathbb{P}^m \to \mathbb{P}^n$, with $m < n$, that is the projectivization of an injective linear mapping $\mathbb{R}^m \to \mathbb{R}^n$, described by an $(n+1) \times (m+1)$ matrix $\mathbf{B}$ of full rank (and we will write $J = [\mathbf{B}]$). An embedding may be viewed as an identification of (an abstract) $\mathbb{P}^m$ with an m-dimensional

subspace N inside $\mathbb{P}^m$, which is the image of J. In other words, specifying an embedding J is equivalent to fixing a linear space N in $\mathbb{P}^m$ together with a projective reference frame on N.

The duality between projections and embeddings can be seen concretely by noting that a linear projection $P : \mathbb{P}^n \dashrightarrow \mathbb{P}^m$ with $P = [\mathbf{A}]$ has an associated embedding $J : \mathbb{P}^m \to \mathbb{P}^n$ where $J = [\mathbf{A}^T]$. More geometrically, the map J can be interpreted as a map on *dual projective spaces*: the projection P induces in fact a mapping $(\mathbb{P}^m)^* \to (\mathbb{P}^n)^*$ associating every hyperplane H in $\mathbb{P}^m$ with its pre-image $P^{-1}(H)$ inside $\mathbb{P}^n$. It is immediate to verify that using *dual homogeneous coordinates* this association is described by $\mathbf{A}^T$. In terms of this dual representation, the image of J corresponds to $K^\vee$, *i.e.*, the system of all hyperplanes in $\mathbb{P}^n$ containing the center K of the projection P. Note that since K has dimension $n - m - 1$, the dimension of $K^\vee$ is indeed m (cf. Appendix A).

Embeddings are similar to projections in the way they act on linear subspaces. If M is an r-dimensional linear subspace of $\mathbb{P}^m$, its image $J(M)$ under the embedding J is an r-dimensional linear space in $\mathbb{P}^n$. This defines a map

$$\begin{aligned} \mathrm{Gr}(r, \mathbb{P}^m) &\to \mathrm{Gr}(r, \mathbb{P}^n) \\ M &\mapsto J(M), \end{aligned} \tag{1.9}$$

whose image is $Y(r, N)$, the set of of r-linear spaces in $\mathbb{P}^n$ contained in N (the image of J). If $J = [\mathbf{B}]$, the map (1.9) can be described in terms of Plücker coordinates using $C_{r+1}(\mathbf{B})$.

Proposition 1.2.3. *If the r-dimensional linear space M in $\mathbb{P}^m$ has Plücker coordinates $[\mathbf{w}_M]$ in $\mathbb{P}^{\binom{m+1}{r+1}-1}$, then its image $U = J(M)$ in $\mathbb{P}^n$ for the linear embedding $J = [\mathbf{B}]$ has Plücker coordinates $[\mathbf{v}_U]$ in $\mathbb{P}^{\binom{n+1}{r+1}-1}$, where*

$$\mathbf{v}_U = C_{r+1}(\mathbf{B})\mathbf{w_M}.$$

In particular, the map (1.9) can be thought of as a linear embedding $J_r : \mathbb{P}^{\binom{m+1}{r+1}} \to \mathbb{P}^{\binom{n+1}{r+1}}$ where $J_r = [C_{r+1}(\mathbf{B})]$, restricted to the Grassmannian varieties. The image of J_r is the linear space $Y(r, N)$ inside $\mathbb{P}^{\binom{m+1}{r+1}-1}$.

Proof. This is completely analogous to Proposition 1.2.2. □

When an embedding $J = [\mathbf{B}]$ is viewed as the dual of a projection $P = [\mathbf{A}]$ (so $\mathbf{B} = \mathbf{A}^T$), the embedding in Proposition 1.2.3 describes inverse mappings of linear spaces. More precisely, the projection P induces a map $\mathrm{Gr}(m-r-1, \mathbb{P}^m) \to \mathrm{Gr}(n-r-1, \mathbb{P}^n)$ associating an $(m - r - 1)$-dimensional space T in $\mathbb{P}^m$ with its $(n - r - 1)$-dimensional pre-image $P^{-1}(T)$ inside $\mathbb{P}^n$. Proposition 1.2.3 describes this map for *dual Plücker coordinates* (in both $\mathbb{P}^m$ and $\mathbb{P}^n$). The image of J is the set of of $(n - r - 1)$-linear spaces containing the center K of the projection P.

Since $C_{k+1}(\mathbf{A})^T = C_{k+1}(\mathbf{A}^T)$ holds, the analytical expressions of projections and inverse projections of linear spaces are related. However, while $C_{k+1}(\mathbf{A})$ acts on linear spaces of dimension k (in $\mathbb{P}^n$), its transpose $C_{k+1}(\mathbf{A}^T)$ is applied to linear spaces of *codimension* k (in $\mathbb{P}^m$), because of the use of dual coordinates. For example, a general k-dimensional space L in $\mathbb{P}^n$ can be projected using $C_{k+1}(\mathbf{A})$ to a k-dimensional space M in $\mathbb{P}^m$, but the inverse image $S = P^{-1}(M)$ of M is obtained by applying $C_{m-k}(\mathbf{A}^T)$ to the *dual* representation of M. In fact, $S = P^{-1}(P(L))$ has dimension $n - m + k$ (codimension $m - k$), and is spanned by L and the center of projection K.

The case of cameras. We now describe projections and inverse projections of linear spaces more concretely for a camera $P : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ with $P = [\mathbf{A}]$. The first interesting case is the mapping from lines in $\mathbb{P}^3$ to lines in $\mathbb{P}^2$. This is sometimes known as the *line projection* map. According to Proposition 1.2.2, it is described by the 3×6 matrix

$$C_2(\mathbf{A}) = \begin{bmatrix} A_{[12][12]} & A_{[12][13]} & A_{[12][14]} & A_{[12][23]} & A_{[12][24]} & A_{[12][34]} \\ A_{[13][12]} & A_{[13][13]} & A_{[13][14]} & A_{[13][23]} & A_{[13][24]} & A_{[13][34]} \\ A_{[23][12]} & A_{[23][13]} & A_{[23][14]} & A_{[23][23]} & A_{[23][24]} & A_{[23][34]} \end{bmatrix}, \tag{1.10}$$

where $A_{[ij][kl]}$ denotes the 2×2 minor of $\mathbf{A}$ corresponding to the rows i, j and columns k, l. Thus, a line in $\mathbb{P}^3$ with Plücker coordinates $[\mathbf{p}] = [p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}]$ is projected by P onto the line in $\mathbb{P}^2$ with Plücker coordinates $[\mathbf{v}] = [v_{12}, v_{13}, v_{23}]$ where $\mathbf{v} = C_2(\mathbf{A})\mathbf{p}$. Note however that the usual representation of lines in $\mathbb{P}^2$ is in terms of *dual* (Plücker) coordinates: this means that rather than the vector $\mathbf{v}$ it is often more natural to use $\mathbf{v}^* = (v_{23}, -v_{13}, v_{12})$, since the line with Plücker coordinates $[\mathbf{v}]$ is defined by the equation $v_{23}x_1 - v_{13}x_2 + v_{12}x_3 = 0$. In particular, for this choice the line projection map becomes associated with the matrix

$$\begin{bmatrix} A_{[23][12]} & A_{[23][13]} & A_{[23][14]} & A_{[23][23]} & A_{[23][24]} & A_{[23][34]} \\ -A_{[13][12]} & -A_{[13][13]} & -A_{[13][14]} & -A_{[13][23]} & -A_{[13][24]} & -A_{[13][34]} \\ A_{[12][12]} & A_{[12][13]} & A_{[12][14]} & A_{[12][23]} & A_{[12][24]} & A_{[12][34]} \end{bmatrix}. \tag{1.11}$$

We now turn to inverse projection mappings. In the notation of Proposition 1.2.3, the only interesting cases are $r = 0$ and $r = 1$. The case $r = 1$ is simply the dual embedding $(\mathbb{P}^2)^* \to (\mathbb{P}^3)^*$, associating a line in $\mathbb{P}^2$ with its pre-image in $\mathbb{P}^3$, that is a plane containing the pinhole. In normal dual coordinates, it is described by the 4×3 matrix $\mathbf{A}^T$. The case $r = 0$ is a map $\mathbb{P}^2 \to \mathrm{Gr}(1, \mathbb{P}^3)$ associating a point in $\mathbb{P}^2$ with the corresponding "viewing line" through the pinhole. According to Proposition 1.2.3, it is described by the 6×3 matrix $C_2(\mathbf{A}^T)$, which is the transpose of (1.10). This description is however based on *dual* coordinates, in both $\mathbb{P}^2 = \mathrm{Gr}(0, \mathbb{P}^2)$ and $\mathrm{Gr}(1, \mathbb{P}^3)$: this means that the pre-image of a point $[\mathbf{u}] = [u_1, u_2, u_3]$ in $\mathbb{P}^2$ is a line with dual Plücker coordinates $[\mathbf{q}] = [q_{34}, -q_{24}, q_{23}, q_{14}, -q_{13}, q_{12}]$ given by $\mathbf{q} = C_2(\mathbf{A})^T\mathbf{u}^*$, where $\mathbf{u}^* = (u_3, -u_2, u_1)$. To use standard (primal) coordinates in $\mathbb{P}^2$ and $\mathrm{Gr}(1, \mathbb{P}^3)$ one needs to apply the following

6×3 matrix, obtained from $C_2(\mathbf{A})^T$ by appropriately fixing signs and permuting rows and columns:

$$\begin{bmatrix} A_{[23][34]} & -A_{[13][34]} & A_{[12][34]} \\ -A_{[23][24]} & A_{[13][24]} & -A_{[12][24]} \\ A_{[23][23]} & -A_{[13][23]} & A_{[12][23]} \\ A_{[23][14]} & -A_{[13][14]} & A_{[12][14]} \\ -A_{[23][13]} & A_{[13][13]} & -A_{[12][13]} \\ A_{[23][12]} & -A_{[13][12]} & A_{[12][12]} \end{bmatrix}. \tag{1.12}$$

1.3 The Joint Image Variety

In this section we study systems of multiple cameras using the *joint image*. We begin by presenting some basic definitions and geometric properties of the joint image (Sections 1.3.1 and 1.3.2). We discuss in Section 1.3.3 the differences between families of algebraic constraints that can be used to describe point correspondences, making the distinction between "ideal-theoretic", "set-theoretic" and "weak" characterizations of the joint image. In Section 1.3.4 we introduce the space of camera configurations, and prove that a joint image uniquely characterizes a configuration (Theorem 1.3.16). Finally, in Section 1.3.5 we clarify the role of bilinear and trilinear conditions for determining point configurations and camera geometry (Theorem 1.3.19).

1.3.1 Basic definitions

In the following, we consider a family $P_1, \ldots, P_n$ of $n \geq 2$ pinhole projective cameras $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ with distinct pinholes $c_1, \ldots, c_n$.

Definition 1.3.1. An n-tuple of image points $(u_1, \ldots, u_n)$ in $(\mathbb{P}^2)^n$ is a *point correspondence* if there exists x in $\mathbb{P}^3 \setminus \{c_1, \ldots, c_n\}$ such that $P_i(x) = u_i$ for all $i = 1, \ldots, n$.

Definition 1.3.2. The *open joint image* $M^o(P_1, \ldots, P_n)$, is the subset of $(\mathbb{P}^2)^n$ formed by point correspondences.

Remark 1.3.3. The set of all point correspondences has been previously considered, with various names. To our knowledge, it was first introduced by Triggs, who coined the term "joint image" in [190]. Heyden and Åstrom refer to it as "natural descriptor" set, and study some of its properties in [84]. It was then reintroduced in mathematics as the *multi-view variety* by Aholt, Sturmfels and Thomas [4]. We use Triggs' original term the "joint image", which we find to be most descriptive.

It was noted by Heyden and Åström [84] that the joint image is *not* an algebraic set, in other words it cannot be described as the zero-set of a family of polynomial equations.

Definition 1.3.4. The *joint image* (variety) $M(P_1, \ldots, P_n)$ is the Zariski closure of the joint image.

Notation: Although the joint image (resp. open joint image) depends on the cameras $P_1, \ldots, P_n$, we will often denote it simply with M_n (resp. M_n^o) when no confusion can arise.

In the Zariski topology, closed sets coincide with algebraic sets. This means that M_n is the smallest set containing M_n^o that can be described by polynomial equations. The discrepancy between M_n and M_n^o is a consequence of the fact that the 3D point x from Definition 1.3.1 is not allowed to be one of the pinholes $c_1, \ldots, c_n$. This is a necessary requirement because at the pinholes the camera mappings are not all defined. To describe the difference between M_n and M_n^o more concretely, we recall the notion of an *epipole*, that is well known in multi-view geometry. For two cameras P_i and P_j with distinct pinholes c_i and c_j respectively, the epipoles e_{ij}, e_{ji} are image points ($\mathbb{P}^2$) defined by $e_{ij} = P_i(c_j)$ and $e_{ji} = P_j(c_i)$. Note that for n cameras $P_1, \ldots, P_n$ we expect $n-1$ epipoles in each image, however if any three pinholes (say c_i, c_j, c_k) are aligned, then the corresponding epipoles in each image will coincide (*e.g.*, $e_{ij} = e_{ik}$). In particular, if all pinholes are aligned there is only one epipole per image.

Proposition 1.3.5. *Given $n \geq 3$ cameras with non-collinear distinct pinholes, one has $M_n^o = M_n \setminus \mathcal{E}_n$, where*

$$\mathcal{E}_n = \bigcup_{i=1}^{n} \{e_{1i}\} \times \ldots \times \mathbb{P}^2_{(i)} \times \ldots \times \{e_{ni}\}. \tag{1.13}$$

Here $\mathbb{P}^2_{(i)}$ indicates $\mathbb{P}^2$ at position i in the product, and e_{ij} denotes the j-th epipole in the i-th image. However, if $n = 2$, or more generally if the cameras have collinear pinholes, then

$$\mathcal{E}_n = \bigcup_{i=1}^{n} \{e_1\} \times \ldots \times \mathbb{P}^2_{(i)} \times \ldots \times \{e_n\} \setminus \{(e_1, \ldots, e_n)\}, \tag{1.14}$$

where e_i now denotes the unique epipole in image i (so $e_i = P_i(c_j)$ for all $j \neq i$).

A proof of this result is given at the end of the chapter. However, the intuition behind it is quite clear: an n-tuple $(u_1, \ldots, u_n)$ in $(\mathbb{P}^2)^n$ belongs to M_n if and only if the corresponding *viewing lines* $\ell_1, \ldots, \ell_n$ defined by $\ell_i = P_i^{-1}(u_i)$ all meet, and the set $\mathcal{E}_n$ corresponds to the n-tuples for which these lines meet (only) at one of the pinholes c_i.

Example 1.3.6. For two cameras P_1, P_2, any point correspondence (u_1, u_2) in $\mathbb{P}^2 \times \mathbb{P}^2$ with $u_i = [\mathbf{u}_i]$, $i = 1, 2$, satisfies an algebraic relation $\mathbf{u}_1^T \mathbf{F} \mathbf{u}_2 = 0$, where $\mathbf{F}$ is the 3×3 *fundamental matrix* associated with P_1 and P_2 [78]. On the other hand, because the two epipoles e_1, e_2 correspond to the left and right null-spaces of $\mathbf{F}$, the same algebraic relation is also satisfied by all pairs in $\mathbb{P}^2 \times \mathbb{P}^2$ of the form (e_1, u_2) or (u_1, e_2), for arbitrary

u_1 and u_2. Indeed, we have in this case that

$$M_2 = \{\, ([\mathbf{u}_1], [\mathbf{u}_2]) \mid \mathbf{u}_1^T \mathbf{F} \mathbf{u}_2 = 0 \,\} \subset \mathbb{P}^2 \times \mathbb{P}^2, \qquad M_2^o = M_2 \setminus \mathcal{E}_2,$$
$$\text{where} \qquad \mathcal{E}_2 = (\{e_1\} \times \mathbb{P}^2) \cup (\mathbb{P}^2 \times \{e_2\}) \setminus \{(e_1, e_2)\}.$$

This agrees with the result of Proposition (1.3.5). ♢

In practice, we do not lose any actual information by replacing M_n^o by its closure M_n since, as we will see shortly, the epipoles are always "distinguishable" in the joint image. In the following, we will talk about equations that "characterize point correspondences", referring to polynomial constraints that actually describe the joint image variety M_n.

1.3.2 First properties of the joint image

By definition, the joint image M_n is the closure of the image of the "joint-projection" map

$$\begin{array}{ccc} \mathbb{P}^3 \setminus \{c_1, \ldots, c_n\} & \longrightarrow & \mathbb{P}^2 \times \ldots \times \mathbb{P}^2 \\ x & \longmapsto & (P_1(x), \ldots, P_n(x)). \end{array} \tag{1.15}$$

This map is usually injective, the only exception being when all of the pinholes are aligned (in particular, for $n = 2$ cameras): in this case, all points lying on the "baseline" spanned by the pinholes will have the same image. The inverse of (1.15), where it is well-defined, is the *triangulation* map $M_n \dashrightarrow \mathbb{P}^3$. This map returns a space point given corresponding image points. Note that this is a rational map (*i.e.*, it can be described using polynomial expressions) because it amounts to computing the intersection of visual rays. This implies that M_n and $\mathbb{P}^3$ are *birationally equivalent* (cf. Appendix B). Intuitively, this says that M_n is "almost" a copy of $\mathbb{P}^3$, embedded in $\mathbb{P}^2 \times \ldots \times \mathbb{P}^2$. This immediately implies that the joint image is irreducible (it is not the union of proper subvarieties) and has dimension 3.

The following proposition deals with the singularities of M_n. The proof is technical, and deferred to the end of the chapter.

Proposition 1.3.7 (Singularities of the joint image variety)**.** *When the camera pinholes are not collinear, M_n is smooth. When they are collinear (in particular for $n = 2$ views), then M_n has a unique singular point given by the n-tuple of epipoles $(e_1, \ldots, e_n)$.*

This result shows how epipoles represent distinguished points in M_n. Indeed, for any choice of $k \geq 2$ cameras, there exists a natural surjective projection map $M_n \to M_k(P_{i_1}, \ldots, P_{i_k})$, that selects subsets of the image points (note that this is also a birational equivalence). Hence, a pair of epipoles can be seen as the singular points of the projection of M_n to the joint image defined by the corresponding pair of cameras.

1.3.3 Algebraic characterizations of the joint image

The joint image variety is a projective variety in $(\mathbb{P}^2)^n$. This means that it can be characterized by a set of "multi-homogeneous" polynomials in $\mathbb{R}[x_1, y_1, z_1, \ldots, x_n, y_n, z_n]$, where each triple of variables x_i, y_i, z_i $(i = 1, \ldots, n)$ are coordinates associated with an image. A complete description of the ideal of all polynomials that vanish on the joint image is given by Aholt *et al.* in [4] (see Theorem 1.3.10). Their derivation of the generators for this ideal is based on a general strategy for recovering multi-view constraints that was discussed by Heyden and Åstrom in [84, 83], and is also presented in [78, Chapter 17]. We will now review this approach, emphasizing the important differences between the following types of algebraic characterizations:

- An *ideal-theoretic characterization* of M_n provides a set of polynomials that generate the ideal $I(M_n)$ in the ring $\mathbb{R}[x_1, y_1, z_1, \ldots, x_n, y_n, z_n]$. This means that all polynomials that vanish on M_n can be obtained as algebraic combinations of the given generators. A further requirement is that the polynomials form a *Gröbner basis*, which is not true for an arbitrary set of generators.

- A *set-theoretic characterization* of M_n provides a set of polynomials $p_1, \ldots, p_n$ such that

$$M_n = \{(u_1, \ldots, u_n) \in (\mathbb{P}^2)^n \,|\, p_i(\mathbf{u}_1, \ldots, \mathbf{u}_n) = 0,\ i = 1, \ldots, n\} \subset (\mathbb{P}^2)^n. \quad (1.16)$$

  In other words, the variety M_n is cut out by the conditions $\{p_i = 0\}$ for $i = 1, \ldots, n$. By definition, the polynomials $p_1, \ldots, p_n$ will belong to the ideal $I(M_n)$, so a set-theoretic characterization may require fewer conditions than those necessary to generate the whole ideal. Over $\mathbb{C}$, the relationship between set-theoretic and ideal-theoretic descriptions is completely governed by *Hilbert's Nullstallensatz* (for our purposes, in its multi-homogeneous form). See Appendix B.

- An *"indirect"* or *"weak" characterization* of M_n is for us a family of polynomials $p_1, \ldots, p_k$ such that the set

$$M'_n = \{(u_1, \ldots, u_n) \in (\mathbb{P}^2)^n \,|\, p_i(\mathbf{u}_1, \ldots, \mathbf{u}_n) = 0,\ i = 1, \ldots, k\} \subset (\mathbb{P}^2)^n \quad (1.17)$$

  contains M_n, possibly strictly, but that allows us nonetheless to recover M_n in some indirect way. In other words, there must be only "feasible" M_n inside the larger set M'_n defined by (1.17). This type of description is in general the most compact, since fewer conditions are usually necessary to characterize a larger set.

While from a purely algebraic standpoint the description at the level of ideals is most useful, for applications it is natural to focus on set-theoretic characterizations, or even more compact descriptions that encode same information in a more efficient (albeit indirect) way. We will see that these characterizations can be used to recover camera geometry from image data.

We now begin our derivation of multi-view constraints. Let $(u_1, \ldots, u_n)$ in $(\mathbb{P}^2)^n$ be a point correspondence for the cameras $P_1, \ldots, P_n$. After fixing coordinate representatives $u_i = [\mathbf{u}_i]$ and $P_i = [\mathbf{P}_i]$, we define the $3n \times (n+4)$ matrix

$$\mathcal{U}(\mathbf{u}_1, \ldots, \mathbf{u}_n) = \begin{bmatrix} \mathbf{P}_1 & \mathbf{u}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{P}_2 & \mathbf{0} & \mathbf{u}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_n & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{u}_n \end{bmatrix}. \tag{1.18}$$

It is clear that a necessary condition for the n-tuple $(u_1, \ldots, u_n)$ to form a correspondence is that $\mathcal{U}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ be rank deficient. Indeed, if $P_i(x) = u_i$ with $x = [\mathbf{x}]$, then there exists a non-zero vector $[\mathbf{x}; \lambda_1; \ldots; \lambda_n]$ in the null space of the matrix. Moreover, the maximal minors of $\mathcal{U}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ provide a *set-theoretic* characterization of the joint image:

$$M_n = \{(u_1, \ldots, u_n) \in (\mathbb{P}^2)^n \mid \mathcal{U}(\mathbf{u}_1, \ldots, \mathbf{u}_n) \text{ is not full rank}\}. \tag{1.19}$$

In the following it will be useful to say that a polynomial in $\mathbb{R}[x_1, y_1, z_1, \ldots, x_n, y_n, z_n]$ is *k-linear* (for $k \leq n$) when it involves only k triples of variables x_i, y_i, z_i and is linear in each triple that appears. In particular, an n-linear polynomial will be simply referred to as *multilinear*. For $k = 2, 3, 4$ we will also use "bilinear", "trilinear", and "quadrilinear", respectively.

The maximal minors $\mathcal{U}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ provide a characterization of the joint image based on multilinear (n-linear) polynomials. On the other hand, conditions of lower degree can be obtained from the maximal minors of a matrix as in (1.18) for subsets of the original matrix. Indeed, if $(u_1, \ldots, u_n)$ is a correspondence for $P_1, \ldots, P_n$ then any subset $(u_{\alpha_1}, \ldots, u_{\alpha_k})$ will be a correspondence for $P_{\alpha_1}, \ldots, P_{\alpha_k}$.

Definition 1.3.8. We indicate with S_k the set of k-linear polynomials given by the maximal minors of

$$\mathcal{U}_\alpha = \begin{bmatrix} \mathbf{P}_{\alpha_1} & \mathbf{u}_{\alpha_1} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{P}_{\alpha_2} & \mathbf{0} & \mathbf{u}_{\alpha_2} & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{\alpha_k} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{u}_{\alpha_k} \end{bmatrix},$$

with $\alpha = \{\alpha_1, \ldots, \alpha_k\}$ ranging among all subsets of size k of $\{1, \ldots, n\}$.

For $k = 2$, the matrix $\mathcal{U}_\alpha$ from Definition 1.3.8 is square (of size 6×6), so there is only one bilinear constraint $b_{\alpha_1\alpha_2}$ for every pair of cameras $P_{\alpha_1}, P_{\alpha_2}$. The polynomial $b_{\alpha_1\alpha_2}$ in fact the well-known *epipolar constraint*, which is usually identified with the *fundamental matrix* that encodes its coefficients [78].

For any k, the polynomials in S_k will vanish on M_n. However, it is easy to see that we can restrict ourselves to sets with $k \leq 4$.

Proposition 1.3.9. *Every polynomial in S_n is of the form $m \cdot p$ where m is a monomial factor and p belongs to S_k with $k \leq 4$. From this we deduce that the bilinear, trilinear, and quadrilinar polynomials in S_2, S_3, S_4 are sufficient to characterize M_n set-theoretically.*

Proof. The result follows from the fact that a non-vanishing maximal minor of $\mathcal{U}$ requires choosing $n+4$ rows, with at least one row associated with each camera: this distinguishes k cameras with $2 \leq k \leq 4$ for which more than one row is chosen. The monomial factors can be removed from the constraints since each k-linearity is multiplied by sets of monomials that cannot vanish simultaneously. □

One of the main results in [4] shows that in fact the polynomials in S_2, S_3, S_4 describe the joint image in a very strong sense.

Theorem 1.3.10 (Aholt et al. [4])**.** *The polynomials in S_2, S_3, S_4 form a* universal Gröbner basis *for the ideal $I(M_n)$ (this means that they constitute a Gröbner basis of $I(M_n)$ under all possible monomial orderings).*

We will see shortly that this characterization based on bilinear, trilinear and quadrilinear constraints is actually very redundant. In fact, we will argue that the quadrilinear constraints are always completely unnecessary (a well known fact [54]), and, more importantly, much fewer bilinear and trilinear conditions can actually be used (Section 1.4).

We conclude this discussion with a result that provides the dimension of the vector space of all multilinear (n-linear) relations that vanish on M_n. The quantity can also also be deduced from [4, Theorem 3.6] that gives the "multigraded Hilbert function" for the ideal $I(M_n)$.

Proposition 1.3.11. *The multilinear polynomials that vanish on M_n form a vector space of dimension $d_n = 3^n - \binom{n+3}{3} + n$.*

Proof sketch. It is sufficient to compute the dimension of the vector space generated by the initial terms associated to the multilinear constraints. Since the maximal minors of $\mathcal{U}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ defined in (1.18) form a Gröbner basis, the result follows from a counting argument involving the associated initial monomials. □

For example, let us point out that $d_2 = 1$ (the epipolar constraint is the only bilinear relation for two views), $d_3 = 10$ (for three views there are always 10 linearly independent trilinearities), and $d_4 = 50$; these facts can also be verified computationally using Gröbner bases.

Remark 1.3.12. Proposition 1.3.11 is useful to explain the geometry of M_n under the *Segre embedding*. We recall first that the Segre embedding is a map $\mathbb{P}^2 \times \ldots \times \mathbb{P}^2 \hookrightarrow \mathbb{P}^{3^n-1}$ that identifies an n-tuple of projective points with an n-dimensional rank-one tensor defined up to scale (see also Appendix B): for example, we can identify a pair

$(u_1, u_2) = ([\mathbf{u}_1], [\mathbf{u}_2])$ with the rank-one 3×3 matrix $\mathbf{u}_1^T \mathbf{u}_2$, and we can do the same for more factors. The image of $(\mathbb{P}^2)^n$ under the Segre map is the *Segre variety* $\mathrm{Seg}_n(\mathbb{P}^2)$. This is a $2n$-dimensional variety inside $\mathbb{P}^{3^n-1}$, and represents the set of rank-one tensors defined up to scale. We can now view M_n as a projective sub-variety of $\mathrm{Seg}_n(\mathbb{P}^2)$. Since multilinear relations in $(\mathbb{P}^2)^n$ correspond to linear ones in $\mathbb{P}^{3^n-1}$, we see from (1.19) that M_n can be identified with $\mathrm{Seg}_n(\mathbb{P}^2)$ *intersected with a linear space*. The co-dimension of this linear space is the number d_n provided by Proposition 1.3.11.

1.3.4 Camera configurations and the joint image

We interrupt for a moment our study of the joint image to introduce the fundamental notion of camera configuration. In the following, we will say that two sets of n cameras $P_1, \ldots, P_n$ and $Q_1, \ldots, Q_n$ are projectively equivalent if there exists a single projective transformation T such that $P_i = Q_i T$ (so if $T = [\mathbf{T}]$ with $\mathbf{T}$ in $GL(4, \mathbb{R})$, then $\mathbf{P}_i = \alpha_i \mathbf{Q}_i \mathbf{T}$ for non-zero constants α_i).

Definition 1.3.13. The space of *camera configurations* $\mathcal{C}_n$ is the set of n-tuples of cameras up to projective equivalence.

In other words, if we write $\mathcal{P}$ as the space of all pinhole cameras (which we may view as an open subset of $\mathbb{P}^{11}$), then $\mathcal{C}_n = \mathcal{P}^n / GL(4, \mathbb{R})$, where $GL(4, \mathbb{R})$ acts on $\mathcal{P}^n$ by $([\mathbf{P}_1], \ldots, [\mathbf{P}_n])\mathbf{T} = ([\mathbf{P}_1\mathbf{T}], \ldots, [\mathbf{P}_n\mathbf{T}])$. It will also be useful to say that a camera configuration of $P_1, \ldots, P_n$ is *non-degenerate* if the pinholes of $P_1, \ldots, P_n$ are all distinct. The following important lemma refers to the definitions from Section 1.2.3.

Lemma 1.3.14. *Given two cameras P_1, P_2 with distinct pinholes c_1, c_2, we consider the groups $K_{c_1} = \mathrm{Stab}_{GL(4,\mathbb{R})}(P_1)$ and $K_{c_2} = \mathrm{Stab}_{GL(4,\mathbb{R})}(P_2)$ inside $GL(4, \mathbb{R})$ (recall that these only depend on the pinholes c_1, c_2). Then we have that*

$$K_{c_1} \cap K_{c_2} = \{k\mathbf{I}_4 \mid k \in \mathbb{R} \setminus \{0\}\}.$$

In particular, projective transformations act transitively *on non-degenerate camera configurations (or, more generally, whenever at least two pinholes are distinct).*

Proof. Without loss of generality we may assume that $c_1 = [0, 0, 0, 1]$ and $c_2 = [0, 0, 1, 0]$. Given (1.2), the groups K_{c_1} and K_{c_2} can be described as

$$K_{c_1} = \left\{ \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & a & 0 \\ b & c & d & e \end{bmatrix} \in GL(4, \mathbb{R}) \right\}, \quad K_{c_2} = \left\{ \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ b & c & d & e \\ 0 & 0 & 0 & a \end{bmatrix} \in GL(4, \mathbb{R}) \right\}.$$

The fact that $K_{c_1} \cap K_{c_2} = \{k\mathbf{I}_4,\ k \in \mathbb{R}\}$ can now be verified directly. □

A consequence of this lemma is that the space of camera configurations $\mathcal{C}_n$ can be viewed as a a homogeneous space (a manifold) of dimension $11n - 15$. We now discuss the relationship between camera configurations and the joint image.

Proposition 1.3.15. *If $P_1, \ldots, P_n$ and $Q_1, \ldots, Q_n$ are projectively equivalent, then they define the same joint image $M(P_1, \ldots, P_n) = M(Q_1, \ldots, Q_n)$ in $(\mathbb{P}^2)^n$. In particular, we can define a map*

$$\mathcal{J} : \mathcal{C}_n \to \{\text{varieties in } (\mathbb{P}^2)^n\}, \tag{1.20}$$

associating a camera configuration with a joint image.

Proof. This is obvious because the joint image is independent of the choice of coordinates in $\mathbb{P}^3$: if T is a projective transformation of $\mathbb{P}^3$, then $P_1, \ldots, P_n$ and $P_1T, \ldots, P_nT$ have the same set of point correspondences. □

The following important result can be viewed as the theoretical basis for multi-view reconstruction: it states that point correspondences are sufficient for determining a unique camera configuration. This fact may not seem surprising, since it is well known that multifocal tensors can be estimated from point correspondences and they determine a unique configuration. However, it is interesting that this result would not be true if we had considered projections $\mathbb{P}^2 \dashrightarrow \mathbb{P}^1$ rather than cameras $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ (see Theorem 3.3.7 and [77]). We also also mention [90], that gives a proof of this fact in a more a general setting using algebraic geometry.

Theorem 1.3.16. *The map $\mathcal{J}$ from Proposition 1.3.15 is injective. In particular, two families of pinhole cameras have the same joint image if and only if they are in the same configuration.*

In our proof of this result, we make use of the following lemma.

Lemma 1.3.17. *Let $x_1, \ldots, x_6$ and $u_1, \ldots, u_6$ be six points in $\mathbb{P}^3$ and $\mathbb{P}^2$. Assuming that all these points are in general position, then there exists at most one camera P such that $P(x_i) = u_i$ for $i = 1, \ldots, 6$.*

Proof of Lemma. This is the standard set-up for "camera resectioning": every match from 3D to 2D provides two new independent conditions on the camera parameters (the open set $\mathcal{P} \subset \mathbb{P}^{11}$). Indeed, writing $P = [\mathbf{P}]$, $x_i = [\mathbf{x}_i]$, $u_i = [\mathbf{u}_i]$, then we have that $P(x_i) = u_i$ is equivalent to $\text{rank}[\mathbf{P}\mathbf{x}_i | \boldsymbol{u}_i] = 1$, yielding two linear conditions on the coefficients of $\mathbf{P}$. □

Proof of Theorem. We first consider the case of non-degenerate configurations, which is much more important in practice. We will then point out that the result holds for degenerate configurations as well.

The statement of the theorem is well known for $n = 2$ cameras with distinct pinholes, since a 3×3 fundamental matrix can be used to identify correspondences as well as a

camera configuration [78]. Let us now consider two sets of $n \geq 3$ cameras $P_1, \ldots, P_n$ and $Q_1, \ldots, Q_n$ with distinct pinholes such that $M(P_1, \ldots, P_n) = M(Q_1, \ldots, Q_n)$. Since this clearly implies $M(P_1, P_2) = M(Q_1, Q_2)$, we can use the statement for $n = 2$ to say that P_1, P_2 and Q_1, Q_2 are projectively equivalent. Lemma 1.3.14 now guarantees that there is *only one* projective transformation that maps the pair P_1, P_2 to Q_1, Q_2: thus, if we assume that $P_1 = Q_1$ and $P_2 = Q_2$, we need to show that $P_j = Q_j$ for all $j = 3, \ldots, n$. Since we can prove this by considering one pair P_j, Q_j at the time, we will assume $n = 3$.

We now fix six points $x_1, \ldots, x_6$ in $\mathbb{P}^3$ in general position, and write $u_{11}, \ldots, u_{16}$ and $u_{21}, \ldots, u_{26}$ for points in $\mathbb{P}^2$ such that $P_i(x_j) = u_{ij}$ for $i = 1, 2$ and $j = 1, \ldots, 6$. Since $M(P_1, P_2)$ is birationally equivalent to $M_3 = M(P_1, P_2, P_3) = M(P_1, P_2, Q_3)$, for each for each $i = 1, \ldots, 6$, we expect to find only one triple (u_{1j}, u_{2j}, v_j) in M_3. Now we observe that necessarily $P_3(x_j) = Q_3(x_j) = v_j$ for $j = 1, \ldots, 6$. Using Lemma 1.3.17, we conclude that $P_3 = Q_3$.

The idea of the previous argument is that we can use two cameras P_1, P_2 to "reproject" image points from $M(P_1, P_2)$ to $M(P_1, P_2, P_3, \ldots, P_n)$, and this determines all the cameras $P_3, \ldots, P_n$. This strategy applies whenever there are at least two cameras with distinct pinholes. On the other hand, the result also holds when all cameras $P_1, \ldots, P_n$ have the same pinhole: in fact, in this case there is a unique set of projective transformations T_{1i}, $i = 3, \ldots, n$, of $\mathbb{P}^2$ such that $P_1 = T_{1i}P_i$ (see Section 1.2.3). It is easy to see that these transformations uniquely determine the joint image as well as the camera configuration. □

We conclude this section by noting that Theorem 1.3.16 implies that the triangulation map $T : M(P_1, \ldots, P_n) \dashrightarrow \mathbb{P}^3$ defined in Section 1.3.2 is *unique* up to composing with projective transformations in $\mathbb{P}^3$. In particular, there is a unique "projective structure" on $M(P_1, \ldots, P_n)$, and we may talk about projective properties (linear subspaces, linear independence, *etc.*) inside the image space, based on the geometry of the corresponding objects in $\mathbb{P}^3$.

1.3.5 Bilinearities and trilinearities

We now resume our study of the different sets of multilinear constraints that can be used to describe the joint image (and thus, as we have just seen, a camera configuration). In particular, we revisit the characterization of M_n from Theorem 1.3.10 by arguing that, in practice, the constraints from S_2, S_3, S_4 are not all necessary.

We first note that since $\mathbb{P}^2 \times \ldots \times \mathbb{P}^2$ has dimension $2n$ and M_n has dimension 3, one might hope to find a characterization of M_n based on $2n - 3$ constraints. However, typically one cannot represent an algebraic set of codimension m as the intersection of m hypersurfaces (when this is possible, the set is called a "complete intersection").

Interestingly, however, this number of constraints is often sufficient to recover M_n as a distinguished component.

Example 1.3.18. Consider three cameras P_1, P_2, P_3 with non-collinear pinholes. It is easy to verify that the three epipolar constraints b_{12}, b_{13}, b_{23} *do not* cut out the joint image M_3. However, they define a set $\mathcal{W}$ whose irreducible components are given by

$$\mathcal{W} = \{b_{12} = b_{13} = b_{23} = 0\} = M_3 \cup \mathcal{V}_t, \tag{1.21}$$

where $\mathcal{V}_t$ is the product of the three "trifocal lines" spanned by the epipoles [84]. Note that the joint image M_3 appears as an irreducible component, which means that M_3 (and all its defining equations) can be recovered indirectly. In fact, it is well known that in this case the three epipolar constraints (or the fundamental matrices) are sufficient to recover the camera configuration [54, 78]. ♢

Keeping this example in mind, and coming back to a point raised in Section 1.3.3, we will say that a *weak* or *indirect* characterization of the joint image is provided by a set of polynomials that define a set $\mathcal{W}$ in $(\mathbb{P}^2)^n$ such that there is a unique "feasible" joint image variety M_n inside $\mathcal{W}$.[2] In light of Theorem 1.3.16, this is equivalently a set of multi-view constraints that *uniquely determine a camera configuration*. In contrast, a set of polynomials that cut out the joint image in the usual (set-theoretic) sense will be sometimes called a *strong* characterization.

The simplest way for a set of constraints to provide a weak characterization of the joint image is if the associated algebraic set $\mathcal{W}$ contains M_n as an irreducible component of maximal dimension, as in Example 1.3.18. Note that in many practical settings, this kind of weak description can be used in place of a strong one, since it can characterize correspondences away from the spurious components (possibly using many fewer equations). In fact, we believe that some of the previously considered "necessary and sufficient conditions" for multi-view correspondence were unknowingly weak characterizations, that were satisfied by some special extraneous n-tuples of image points. For example, we will point out in Section 1.5.2 that the nine "point-point-point" conditions encoded in a trifocal tensor [78] are not a complete (strong) characterization of the joint image. It is also possible to have weak characterizations of the joint image in which the associated algebraic set $\mathcal{W}$ is higher dimensional: we will discuss many examples in Section 1.4.

In [84], Heyden and Åström point out that for $n \geq 4$, assuming all the pinholes to be in general position, the bilinear constraints are already sufficient to generate the trilinear and quadrilinear ones. We now extend this result, clarifying the role of the different families of constraints for all possible camera configurations. The proof of the following proposition amounts to reducing to the case $n = 4$ and verifying all the relations computationally. Details are given at the end of the chapter.

[2] Here by "feasible joint image" we mean a variety in $(\mathbb{P}^2)^n$ that arises as the joint image of a configuration of pinhole cameras.

Theorem 1.3.19. *Assume n cameras are given.*

1. *Bilinearities and trilinearities (the sets S_2, S_3 from Definition 1.3.8)* always *strongly characterize M_n, independently of the pinhole configuration.*
2. *Bilinear constraints* alone *strongly characterize M_n, if and only if the pinholes are* not all coplanar
3. *Bilinear constraints* alone *weakly characterize M_n if and only if the pinholes are* not all collinear.

	Non coplanar	Coplanar	Collinear
Bil.	Strong	Weak	Not sufficient
Bil.+Tril.	Strong	Strong	Strong

Table 1.1: Summary of the results of Theorem 1.3.19.

We should observe that the first point of Theorem 1.3.19 is well known (see for example [54]), and indeed reconstruction methods are generally only based on epipolar and trifocal constraints. The second point can be deduced from the analysis for four points in general position in [84], although the authors do not point out this general fact. The last point is, to the best of our knowledge, new, at least in such a general formulation: it shows that trilinear relations are essential *only if all pinholes are collinear*, since otherwise the epipolar relations are sufficient to determine the camera configuration and, indirectly, point correspondences. See Table 1.1.

1.4 Epipolar Constraints and Viewing Graphs

In Theorem 1.3.19, we considered *complete* families of bilinear and trilinear constraints. These sets of conditions are redundant in general: for example, it has been observed in [84] that for five cameras in general position, correspondences can be characterized using 9 bilinear constraints instead of the complete set of 10. In [186], we observed the following more general fact (see Figure 1.3).

Proposition 1.4.1. *Consider $n \geq 4$ cameras with pinholes $c_1, \ldots, c_n$. We write F_{ij} for the fundamental matrix associated with cameras i and j.*

(A) If three pinholes (say) c_1, c_2, c_3 are not coplanar with any other c_i for $i \geq 4$, then the fundamental matrices $\{F_{12}, F_{13}, F_{23}\} \cup \{F_{1i}, F_{2i}, F_{3i}\}_{i=4,\ldots,n}$ identify $3n-6$ bilinear constraints that strongly characterizes the joint image.

(B) If two pinholes (say) c_1, c_2 are not collinear with any other c_i for $i \geq 3$, then the fundamental matrices $\{F_{12}\} \cup \{F_{1i}, F_{2i}\}_{i=3,\ldots,n}$ identify $2n-3$ bilinear constraints that weakly characterize the joint image.

Proof. If $(u_1, \ldots, u_n)$ is an n-tuple of image points that satisfy all of the constraints given in (A), then Proposition 1.3.19 guarantees that the visual rays associated to (u_1, u_2, u_3, u_i) converge for all $i = 4, \ldots, n$. Since c_1, c_2, c_3 are necessarily not collinear, this implies that *all* the visual rays intersect, so that $(u_1, \ldots, u_n)$ is in fact a correspondence. Similarly, the set of constraints given in (B) allow one to determine a consistent set of camera parameters: it is enough to note that F_{12}, F_{1i}, F_{2i} can be used to recover the camera configurations for all triples $1, 2, i$, so by fixing cameras P_1, P_2 compatible with F_{12}, we can uniquely recover all of the remaining cameras. □

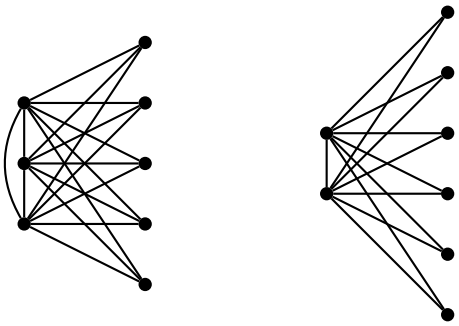


Figure 1.3: Graphs representing epipolar conditions between $n = 8$ cameras: a "strong" characterization based on $3n - 6 = 18$ constraints (*left*), and a "weak" one based on $2n - 3 = 13$ constraints (*right*).

In general, it is convenient to represent sets of fundamental matrices between pairs of views as edges of a graph, as shown in Figure 1.3. This type of graph was called the *viewing graph* by Levi and Werman [113]. In [186], we pointed out that there are generally many viewing graphs with $2n - 3$ edges that represent weak characterizations of the joint image (although we did not use the term "viewing graph", since we were not aware of [113] at the time). For example, if a graph is obtained from 3-cycle by adding vertices of degree two one at the time, then the fundamental matrices associated with its edges uniquely determine a camera configuration. We investigated similar questions in more detail in our follow-up work [187]. There we used the terminology of Levi and Werman, and studied characterizations of "solvable" viewing graphs, which represent sets of fundamental matrices that determine a unique camera configuration.

This section presents the results of [187]. After introducing viewing graphs and the notion of "solvability" in Section 1.4.1, we recall some basic results from [113, 157] in Section 1.4.2. We observe in Section 1.4.3 that any solvable viewing graphs has at least $e(n) = \lceil (11n - 15)/7 \rceil$ edges, and prove that using this minimum number is always possible (Theorem 1.4.9). Note that $e(n) < 2n - 3$ for $n \geq 5$. We then present several new criteria for deciding whether or not a viewing graph is solvable. In particular, we describe a new necessary condition for solvability that is based on the number of edges and vertices of subgraphs (Theorem 1.4.12), as well as a sufficient condition based on

"moves" for adding new edges to a graph (Theorem 1.4.14). In Section 1.4.6 we describe an algebraic formulation for solvability that in principle can always be used to determine whether any viewing graph is solvable. This method is computationally challenging for larger graphs, but we also introduce a much more practical *linear* test, that can be used to verify whether a viewing graph identifies a finite number of camera configurations. Finally, using an implementation of all the proposed methods, we analyze in Section 1.4.7 the solvability of all minimal viewing graphs with at most 9 vertices, and discuss some interesting examples.

1.4.1 The viewing graph

The viewing graph is a graph in which vertices correspond to cameras, and edges represent fundamental matrices between them. More precisely, if $G = (V_G, E_G)$ is an undirected graph with n vertices, and $P_1, \ldots, P_n$ are projective cameras, we write

$$\mathcal{F}_G(P_1, \ldots, P_n) = \{F_{ij} = F(P_i, P_j) \,|\, (i,j) \in E_G\}, \tag{1.22}$$

for the set of fundamental matrices defined by the edges of G. We say that the the set $\mathcal{F}_G(P_1, \ldots, P_n)$ is *solvable* if $\mathcal{F}_G(P_1, \ldots, P_n) = \mathcal{F}_G(P'_1, \ldots, P'_n)$ implies that $(P_1, \ldots, P_n)$ and $(P'_1, \ldots, P'_n)$ are in the same projective configuration. In other words, a set of fundamental matrices is solvable if and only if it uniquely determines a projective configuration of cameras.

Proposition 1.4.2. *The solvability of $\mathcal{F}_G(P_1, \ldots, P_n)$ only depends on the graph G and on the pinholes $c_1, \ldots, c_n$ of $P_1, \ldots, P_n$.*

Proof. The statement expresses the fact that changes of image coordinates are only a "relabeling" of a camera configuration and the associated fundamental matrices. More precisely, if $S_1, \ldots, S_n$ are arbitrary projective transformations of $\mathbb{P}^2$, then $(P_1, \ldots, P_n)$ and $(P'_1, \ldots, P'_n)$ are in the same configuration if and only if the same is true for $(S_1P_1, \ldots, S_nP_n)$ and $(S_1P'_1, \ldots, S_nP'_n)$. This implies that $\mathcal{F}_G(P_1, \ldots, P_n)$ is solvable if and only $\mathcal{F}_G(S_1P_1, \ldots, S_nP_n)$ is. □

Example 1.4.3. If G is a *complete* graph with $n \geq 3$ vertices, then $\mathcal{F}_G(P_1, \ldots, P_n)$ is solvable if and only if the pinholes of the cameras $P_1, \ldots, P_n$ are not all aligned. Indeed, if the pinholes are aligned, then the fundamental matrices between *all* pairs of cameras are not sufficient to completely determine the configuration: replacing any $P_i = [\mathbf{P}_i]$ with $P'_i = [\mathbf{P}_i(\mathbf{I}_4 + \mathbf{c}_j\mathbf{v}^T)]$, where $c_j = [\mathbf{c}_j]$ is the pinhole of another camera and $\mathbf{v}^T$ is arbitrary, yields a new set of cameras which belongs to a different configuration but has the same set of fundamental matrices. Conversely, the complete set of fundamental matrices determines a unique camera configuration whenever there are at least three non-aligned pinholes (cf. Theorem 1.3.19). ♢

In the rest of the section we will only consider generic configurations of cameras/pinholes (so a complete graph will always be solvable). This covers most cases of practical interest, although in the future degenerate configurations (including some collinear or coplanar pinholes) could be studied as well.

Definition 1.4.4. A viewing graph G is said to be *solvable* if $\mathcal{F}_G(P_1, \dots, P_n)$ is solvable for generic cameras $P_1, \dots, P_n$.

In other words, solvable viewing graphs describe sets of fundamental matrices that are generically sufficient to recover a camera configuration. Despite its clear significance, the problem of characterizing which viewing graphs are solvable has not been studied much, and only partial answers are available in the literature (mainly in [113, 157]). It is quite easy to produce examples of graphs that are solvable, but it is much more challenging, given a graph, to determine whether it is solvable or not. The following observation provides another useful formulation of solvability (note that the "if" part requires the genericity assumption, as shown in Example 1.4.3).

Lemma 1.4.5. *A viewing graph G is solvable if and only if, for generic cameras $P_1, \dots, P_n$, the fundamental matrices $\mathcal{F}_G(P_1, \dots, P_n) = \{F(P_i, P_j) \,|\, (i,j) \in E_G\}$ uniquely determine the remaining fundamental matrices $\{F(P_i, P_j) \,|\, (i,j) \notin E_G\}$.*

This viewpoint also suggests the idea that, given any graph G, we can define a "solvable closure" $\overline{G}$, as the graph obtained from G by adding edges corresponding to fundamental matrices that can be deduced generically from $\mathcal{F}_G(P_1, \dots, P_n)$. Hence, a graph is solvable if and only if its closure is a complete graph. We will return to this point in Section 1.4.5.

1.4.2 Simple criteria

We begin by recalling two necessary conditions for solvability that were shown in [113]. These provide simple criteria to show that a viewing graph is *not* solvable.

Proposition 1.4.6. *[113] If a viewing graph with $n > 3$ vertices is solvable, then:*

1. *All vertices have degree at least 2.*
2. *No two adjacent vertices have degree 2.*

We extend this result with the following necessary condition (which implies the first point in the previous statement).

Proposition 1.4.7. *Any solvable graph is 2-connected,* i.e.*, it has the property that after removing any vertex the graph remains connected.*

Proof. Assume that a vertex i disconnects the graph G into two components G_1, G_2, and let $P_1, \dots, P_n$ be a set of n generic cameras, whose pairwise fundamental matrices are

represented by the edges of G. If $c_i = [\mathbf{c}_i]$ is the pinhole of the camera P_i associated with i, then we consider two distinct projective transformations of the form $T_1 = [\mathbf{I}_4 + \alpha_1 \mathbf{c}_i \mathbf{v}_1^T]$ and $T_2 = [\mathbf{I}_4 + \alpha_2 \mathbf{c}_i \mathbf{v}_2^T]$. These transformations fix the camera P_i. If we apply T_1 to all cameras in G_1 and T_2 to all cameras in G_2, while leaving P_i fixed, we obtain a different camera configuration that gives rise to the same set of fundamental matrices as $P_1, \ldots, P_n$ for all edges in G. □

We also recall a result from [157] which will be used in the next section.

Proposition 1.4.8. *[157] If G_1 and G_2 are solvable viewing graphs, then the graph G obtained by identifying two vertices from G_1 and with two from G_2 is solvable.*

Note that if both pairs of vertices in the previous statement are connected by edges in G_1 and G_2, then these two edges will automatically be identified in G.

1.4.3 How many fundamental matrices?

We now ask ourselves what is the minimal number of edges that a graph must have to be solvable (or, equivalently, how many fundamental matrices are required to recover a camera configuration). Since a single epipolar relation provides at most 7 constraints in the $(11n - 15)$-dimensional space camera configuration, we deduce that any solvable graph must have *at least* $e(n) = \lceil (11n-15)/7) \rceil$ edges. This fact was previously observed in [157, Theorem 2]. However, compared to [157], we show here that this bound is tight, *i.e.*, that there always exists a solvable graph with $e = e(n)$ edges. Concretely, this means that, for n generic views, there is always a way of recovering the corresponding camera configuration using $e(n)$ fundamental matrices.

Theorem 1.4.9. *The minimum number of edges of a solvable viewing graph with $n \geq 2$ views is*

$$e(n) = \left\lceil \frac{11n - 15}{7} \right\rceil .$$

Proof. For $n \leq 9$, examples of solvable viewing graphs with $e(n)$ edges are illustrated Figure 1.4. The solvability of these graphs will be shown in Section 1.4.5 (all but one of these also appear in [113]). In particular, let G_0 be a solvable viewing graph with 9 vertices and 12 edges. Using Proposition 1.4.8, we deduce that, starting from a solvable viewing graph G with n vertices and e edges, we can always construct a solvable graph G' with $n + 7$ vertices and $e + 11$ edges. The graph G' is simply obtained by merging G and G_0 as in Proposition 1.4.8, using two pairs of vertices both connected by edges.

Now, for any $n > 9$, we consider the unique integers q, r such that $n = 7q + r$ and $2 \leq r \leq 8$. It is easy to see that

$$e(n) = \left\lceil \frac{11n - 15}{7} \right\rceil = 11q + \left\lceil \frac{11r - 15}{7} \right\rceil .$$

To obtain a solvable viewing graph with n vertices and $e(n)$ edges, we start from a solvable graph with r vertices and $e(r)$ edges, and repeat the gluing construction described above q times. The resulting graph is solvable and has the desired number of vertices and edges. □

Remark 1.4.10. In order to recover projection matrices for n views, it is common to use $2n-3$ fundamental matrices (see for example [83, Sec.4.4]). In fact, as noted at the beginning of this section, a large class of solvable viewing graphs can be defined, starting for example from a 3-cycle, by adding vertices of degree two, one at the time: this always gives a total of $2n-3$ edges. For this type of viewing graphs it is possible to recover projection matrices incrementally, using a pair of fundamental matrices for each camera. It is probably often erroneously believed that $2n-3$ is actually the minimal number of epipolar constraints required: at least, we thought this to be true at first. For us, the confusion was due to the fact that the joint image has dimension three (or codimension $2n-3$) in $(\mathbb{P}^2)^n$. This means means that we expect $2n-3$ conditions to be necessary to cut out generically the set of image correspondences among n views. The existence of solvable viewing graphs with less than $2n-3$ epipolar constraints means that, for these graphs, the associated epipolar constraints cut out a set $\mathcal{W}$ in $(\mathbb{P}^2)^n$ whose dimensions is larger than 3 (and thus does *not* contain the joint image M_n as a maximal dimensional component) but still completely determines M_n. Mathematically, this is an interesting phenomenon that we hope to investigate in the future.

Some values of $e(n)$ are listed in Table 1.2 (here $d(n)$ represents the minimal number of constraints on the fundamental matrices, and will be discussed shortly). Note that $e(n) < 2n-3$ for all $n \geq 5$.

n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$e(n)$	1	3	5	6	8	9	11	12	14	16	17	19	20	22	23
$d(n)$	0	3	6	2	5	1	4	0	3	6	2	5	1	4	0

Table 1.2: The relation between n, $e(n)$, and $d(n)$.

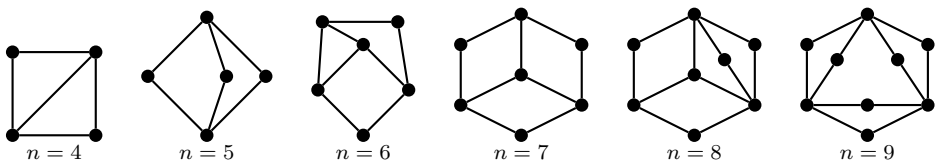


Figure 1.4: Examples of minimal solvable viewing graphs for $n \leq 9$ views (see Section 1.4.5).

1.4.4 Constraints on fundamental matrices

Closely related to the solvability of viewing graphs is the problem of describing *compatibility* of fundamental matrices. Indeed, given a solvable graph G, it is not true in general that any set of fundamental matrices can be assigned to the edges of G, since fundamental matrices must satisfy some feasibility constraints in order to correspond to an actual camera configuration. For example, it is well known that the fundamental matrices F_{12}, F_{23}, F_{31} relating three pairs of cameras with non-aligned pinholes are compatible if and only if

$$\mathbf{e}_{13}^T \mathbf{F}_{12} \mathbf{e}_{23} = \mathbf{e}_{21}^T \mathbf{F}_{23} \mathbf{e}_{31} = \mathbf{e}_{32}^T \mathbf{F}_{31} \mathbf{e}_{12} = 0, \tag{1.23}$$

where $e_{ij} = [\mathbf{e}_{ij}]$ is the epipole in image i relative to the camera j [78, Theorem 15.6]. In most practical situations fundamental matrices are estimated separately, so these constraints need to be taken into account [177]. However, it is sometimes incorrectly stated that compatibility for any set of fundamental matrices only arises from triples and equations of the form (1.23) [157, Theorem 1] [177, Definition 1]. While it is true that for a complete set of $\binom{n}{2}$ fundamental matrices triple-wise compatibility is sufficient to guarantee global compatibility, for smaller sets of fundamental matrices other types of constraints will be necessary. For example, there are many solvable viewing graphs with no three-cycles (*e.g.*, the graph in Figure 1.4 with $n = 5$), however the fundamental matrices cannot be unconstrained if $7e(n) > 11n - 15$, which always true unless $e = 2$ modulo 9 (cf. Table 1.2).

More formally, we can consider the set $\mathcal{X}$ of compatible fundamental matrices between *all pairs* of n views, so that $\mathcal{X} \subset (\mathbb{P}^8)^N$ where $N = \binom{n}{2}$. Since each compatible N-tuple is associated with a camera configuration, we see that $\mathcal{X}$ has dimension $11n - 15$. Given a viewing graph G with n views, we write $\mathcal{X}_G \subset (\mathbb{P}^8)^e$ for the projection of $\mathcal{X}$ onto the factors in $(\mathbb{P}^8)^N$ corresponding to the edges of G. The set $\mathcal{X}_G$ thus represents compatible fundamental matrices for pairs of views associated with the edges of G. The following result follows from dimensionality arguments.

Proposition 1.4.11. *If G is solvable with n vertices, $\mathcal{X}_G$ has dimension $11n - 15$.*

If $\mathcal{X}_G$ has dimension $11n - 15$, then the fundamental matrices assigned to the edges of G must satisfy $d(n, e) = 7e - 11n + 15$ constraints[3]. This also means that the minimum number of constraints on the fundamental matrices associated with a solvable graph is $d(n) = d(n, e(n))$ (see Table 1.2).

We now use Proposition 1.4.11 to deduce a new necessary condition for solvability.

[3]This is the codimension of $\mathcal{X}_G$ in $\mathcal{H}^e$ where $\mathcal{H} \subset \mathbb{P}^8$ is the determinant hypersurface.

Theorem 1.4.12. *Let G be a solvable graph with n vertices and e edges. Then for any subgraph G' of G with n' vertices and e' edges we must have*

$$d(n', e') \leq d(n, e), \tag{1.24}$$

where $d(n, e) = 7e - 11n + 15$. More generally, if $G_1, \ldots, G_k$ are subgraphs of G, each with n_i vertices and e_i edges, with the property that the edge sets $E_{G_i} \subset E_G$ are pairwise disjoint, then we must have

$$\sum_{i=1}^{k} d(n_i, e_i) \leq d(n, e). \tag{1.25}$$

Proof. Using the same notation as above, we note that that $\mathcal{X}_{G'}$ is a projection of $\mathcal{X}_G$ onto e' factors of $(\mathbb{P}^8)^e$: this implies $\dim \mathcal{X}_{G'} + 7(e - e') \geq \dim \mathcal{X}_G$, or $7e' - \dim \mathcal{X}_{G'} \leq 7e - \dim \mathcal{X}_G$. Since $\dim \mathcal{X}_{G'} \leq 11n' - 15$ and $\dim \mathcal{X}_G = 11n - 15$ (because G is solvable), we obtain

$$7e' - 11n' + 15 \leq 7e' - \dim \mathcal{X}_{G'} \leq 7e - \dim \mathcal{X}_G = 7e - 11n + 15.$$

For the second statement, we consider the graph $G' = (\bigcup_i V_{G_i}, \bigcup_i E_{G_i})$. Since the edges of G_i are disjoint, we have

$$\dim \mathcal{X}_{G'} \leq \sum_{i=1}^{k} \dim \mathcal{X}_{G_i} \leq \sum_{i=1}^{k} (11n_i - 15),$$

and $e' = \sum_i^k e_i$. The result follows again from $7e' - \dim \mathcal{X}_{G'} \leq 7e - \dim \mathcal{X}_G$. □

Example 1.4.13. In [113], Levi and Werman observed that all viewing graphs of the form shown in Figure 1.5 are not solvable. This can be easily deduced from Theorem 1.4.12. Indeed, for a graph G of this form, the subgraphs G_1, G_2, G_3, G_4 have disjoint edges, however we have (using the same notation as in the proof of Theorem 1.4.12)

$$\sum_{i=1}^{4} d(n_i, e_i) = d(n, e) - 4 \times 11 + 3 \times 15 > d(n, e).$$

According to Theorem 1.4.12 this means that G is not solvable. ♢

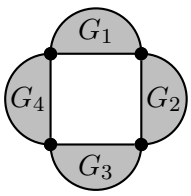


Figure 1.5: A viewing graph of this form (where G_1, G_2, G_3, G_4 represent arbitrary subgraphs) is not solvable

1.4.5 Constructive approach for verifying solvability

Until now we have mainly discussed necessary conditions for solvability, which can be used to show that a given graph is not solvable. We next introduce a general strategy for proving that a graph is solvable. This method is not always guaranteed to work, but in practice it gives sufficient conditions for many of the graphs we tested (cf. Section 1.4.7).

Recall from the beginning of this section that we introduced the "viewing closure" $\overline{G}$ of G as the graph obtained by adding to G all edges corresponding to fundamental matrices that can be deduced from $\mathcal{F}_G(P_1, \ldots, P_n)$. Our approach consists of a series of "moves" which describe valid ways to add new edges to a viewing graph. For this it is convenient to introduce a new type of edge in the graph, which keeps track of the fact that partial information about a fundamental matrix is available. More precisely:

- A *solid (undirected) edge* between vertices i and j means that the fundamental matrix between the views i and j is fixed (as before).
- A *directed dashed edge* (for short, a *dashed arrow*) between vertices i and j means that the i-th epipole in the image j is fixed.

As these definitions suggest, a solid edge also counts as a dashed double- arrow, but the converse is not true. We next introduce three basic "moves" (see Figure 1.6).

I) If there are solid edges defining a four-cycle with one diagonal, draw the other diagonal.

II) If there are dashed arrows $1 \to 2$, $1 \to 3$, and solid edges $2 - 4$ and $3 - 4$, draw a dashed arrow $1 \to 4$.

III) If there are double dashed arrows $1 \leftrightarrow 2$, together with three pairs of dashed arrows $i \to 1, i \to 2$ for $i = 3, 4, 5$, make the arrow between 1 and 2 a solid (undirected) edge.

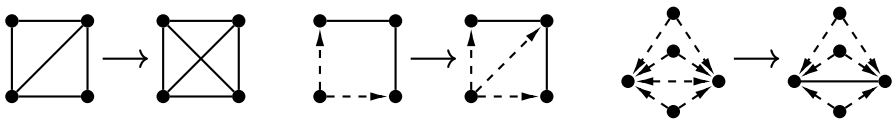


Figure 1.6: Three moves (*left*: I, *center*: II, *right*: III) that can be used to prove solvability.

Theorem 1.4.14. *Let G be a viewing graph. If applying the three moves described above iteratively to G we obtain a complete graph, then G is solvable.*

Proof. For each of the three moves we need to show that the new edges contain information about the unknown fundamental matrices that can actually be deduced from $\mathcal{F}_G(P_1, \ldots, P_n)$.

Move I: The second diagonal of the square is deducible from the other edges because the square with one diagonal is a solvable graph (this is a simple consequence of Proposition 1.4.8).

Move II: Assume that $e_{21} = P_2 c_1$ and $e_{32} = P_3 c_2$ are fixed epipoles in images 2 and 3, and that the fundamental matrices $F_{24} = F(P_2, P_4), F_{34} = F(P_3, P_4)$ are also fixed. If c_1, c_2, c_4 are not aligned, we can use F_{24} to "transfer" the point e_{21}, and obtain a line l_{41} in image 4 that contains epipole e_{41}. Similarly, if c_1, c_3, c_4 are not aligned, we obtain another line m_{41} using the same procedure with F_{34} and e_{31}. If the pinholes c_1, c_2, c_3, c_4 are not all coplanar, the lines l_{41} and m_{41} will be distinct, and their point of intersection will be e_{41}. This implies that we can draw a dashed arrow from 1 to 4.

Move III: Assume that the epipoles e_{21} and e_{12} are fixed, and that the images of three other pinholes c_3, c_4, c_5 are fixed in both images 1 and 2. If the planes c_1, c_2, c_i for $i = 3, 4, 5$ are distinct, then the images of c_3, c_4, c_5 give three correspondences that fix the "epipolar line homography" [78]. This completely determines F_{12}, and we can draw a solid edge between 1 and 2.

□

In practice, the three moves can be applied cyclically until no new edges can be added (it is also easy to argue the order is irrelevant, because we are simply annotating information that is always deducible from the graph). Finally, we note that all three moves are *constructive* and *linear*, meaning they actually provide a practical strategy for computing all fundamental matrices: it is enough to transfer epipoles appropriately, and use them to impose linear conditions on the unknown fundamental matrices.

Example 1.4.15. Using Theorem 1.4.14, we can show that all graphs from Figure 1.4 are solvable. Figure 1.7 illustrates this explicitly for two cases ($n = 6$ and $n = 8$). ♢

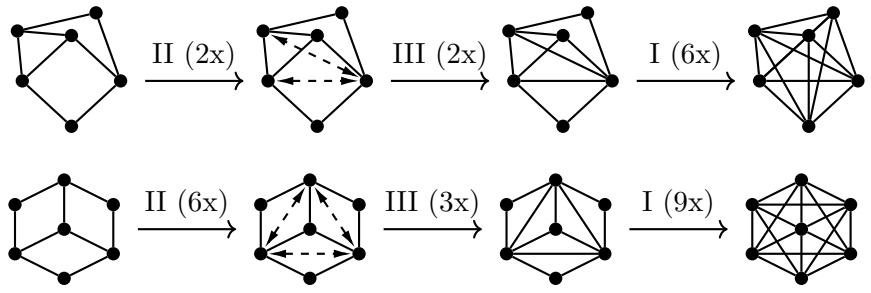


Figure 1.7: Two applications of Theorem 1.4.14 to prove that viewing graphs are solvable.

1.4.6 Algebraic tests for solvability and finite solvability

Given a viewing graph G, it is possible to write down a set of algebraic conditions that will in principle always determine whether G is solvable. One way to do this is by characterizing the set of projective transformations of $\mathbb{P}^3$ that can be applied to all cameras without affecting any of the fundamental matrices represented by the edges of the viewing graph. More precisely, since every pair of vertices connected by an edge represents a projectively rigid pair of cameras, we assign a matrix $\mathbf{g}_\lambda$ in $GL(4,\mathbb{R})$ to each edge λ of the graph (so $\mathbf{g}_\lambda$ describes a projective transformation applied to a pair of cameras). We then impose that matrices on adjacent edges act compatibly on the shared vertex/camera. If the edges λ and λ' share a vertex i, then this compatibility can be written as

$$\mathbf{g}_\lambda \mathbf{g}_{\lambda'}^{-1} = \alpha \mathbf{I}_4 + \mathbf{c}_i \mathbf{v}^T, \tag{1.26}$$

where α is an arbitrary (nonzero) constant and $\mathbf{v}$ is an arbitrary vector: indeed, we are requiring that $\mathbf{g}_\lambda \mathbf{g}_{\lambda'}^{-1}$ belongs to the *stabilizer group*, that was given in (1.2). Thus, if G is a viewing graph with e edges and $c_1, \ldots, c_n$ are a set of pinholes, we consider the set of all compatible assignments of matrices:

$$\mathcal{T}_G(c_1, \ldots, c_n) = \{(\mathbf{g}_\lambda, \lambda \in E_G) \,|\, (1.26) \text{ holds for all adjacent edges in G}\} \subset GL(4,\mathbb{R})^e.$$

If G is solvable, then for general $c_1, \ldots, c_n$ the set $\mathcal{T}_G(c_1, \ldots, c_n)$ will consist of e-tuples of matrices that are all scalar multiples of each other. This in fact means that the only way to act on all cameras without affecting the fixed fundamental matrices is to apply a single projective transformation.

By substituting random pinholes in (1.26), we can use these equations for $\mathcal{T}_G(c_1, \ldots, c_n)$ as an algebraic test for verifying whether a viewing graph is solvable. This approach however is computationally very challenging, since it requires solving a non-linear algebraic system with a large number of variables. On the other hand, if we are only interested in the *dimension* of $\mathcal{T}_G(c_1, \ldots, c_n)$, then we can use a much simpler strategy: noting that $\mathcal{T}_G(c_1, \ldots, c_n)$ may be viewed as an algebraic group (it is a subgroup of $GL(4,\mathbb{R})^e$), it is sufficient to compute the dimension of its *tangent space* at any point, and in particular at the identity (*i.e.*, the product of identity matrices).[4] An explicit representation of the tangent space of $\mathcal{T}_G(c_1, \ldots, c_n)$ is provided by the following result (see the end of the chapter for a proof).

Proposition 1.4.16. *The tangent space of $\mathcal{T}_G(c_1, \ldots, c_n)$ at the identity can be represented as the space of e-tuple of matrices $(\mathbf{h}_\lambda, \lambda \in E_G)$ where each $\mathbf{h}_\lambda$ is in $\mathbb{R}^{4\times 4}$ (not necessary invertible), and with compatibility conditions of the form*

$$\mathbf{h}_\lambda - \mathbf{h}_{\lambda'} = \alpha \mathbf{I}_4 + \mathbf{c}_i \mathbf{v}^T, \tag{1.27}$$

[4] Here we actually need that $\mathcal{T}_G(c_1, \ldots, c_n)$ is smooth: this follows from a technical result, which states that an algebraic group (more properly a "group scheme") over a field of characteristic zero is always smooth [128, Sec.11].

where $\alpha \in \mathbb{R} \setminus \{0\}$ *and* $\mathbf{v} \in \mathbb{R}^4$ *are arbitrary, and* λ *and* λ' *share the vertex* i*.*

When the pinholes have been fixed, the compatibility constraints (1.27) can be expressed as *linear* equations in the entries of the matrices $\mathbf{h}_\lambda$. These equations are obtained by eliminating the variables α and $\mathbf{v}$ from (1.27). The resulting conditions in terms of $\mathbf{h}_\lambda, \mathbf{h}_{\lambda'}, \mathbf{c}_i$ are rather simple, and listed explicitly at the end of the chapter. Using this approach, the dimension of $\mathcal{T}_G(c_1, \ldots, c_n)$ is easy to determine: it is enough to fix the pinholes randomly, and compute the dimension of the induced linear system.

When $\mathcal{T}_G(c_1, \ldots, c_n)$ has dimension $d = 15 + e$ (which accounts for the group of projective transformations, and scale factors for each matrix $\mathbf{g}_\lambda$), we deduce that there are at most a *finite* number of projectively inequivalent ways in which we can act on all the cameras without affecting the fixed fundamental matrices. In other words, the fundamental matrices associated with the edges of G determine at most a finite set of camera configurations (rather than a single configuration, which is our definition for solvability). When this happens, we say that G is *finite solvable*. On the other hand, we were not able to find an example of a finite solvable graph that is provably not solvable, nor to find a proof that "finite solvability" implies "solvability". To our knowledge, whether a set of fundamental matrices can characterize a finite number of configurations, but more than a single one, is a question that has never been addressed.

Open Question: Is it possible for a viewing graph to be finite solvable without being solvable?

Our experiments show that this behavior does not occur for a small number of vertices, but we see no reason why this should be true for larger graphs. This is certainly an important issue that we hope to investigate in the future.

1.4.7 Experiments and examples

We have implemented and tested all of the discussed criteria and methods using the free mathematical software Sage [158], which includes symbolic algebra tools as well as built-in functions for manipulating graphs. We then analyzed solvability for all minimal viewing graphs with $n \leq 9$ vertices and $e(n) = \lceil (11n - 15)/7 \rceil$ edges. The results are summarized in Table 1.3. For every pair $(n, e(n))$, we list the number of all non-isomorphic connected graphs of that size ("connected"), the number of graphs that satisfy the necessary condition from Theorem 1.4.12 ("candidates"), the number of those that satisfy the sufficient condition from Theorem 1.4.14 ("solvable with moves"), and the number of graphs that are finite solvable ("finite solvable"), using the linear method from 1.4.6. We see that Theorems 1.4.12 and 1.4.14 allow us to recover all minimal solvable graphs for $n \leq 7$, since candidate graphs are always solvable with moves. On the other hand, for $n = 8$, and particularly for the unconstrained case $n = 9$, there are some graphs that we could not classify with those methods (although finite solvability

was easy to verify in all cases). For the undecided graphs, we were sometimes able to prove solvability with the general algebraic method from Section 1.4.6, or using other arguments. The following examples present a few interesting cases.

$(n, e(n))$	(3,3)	(4,5)	(5,6)	(6,8)	(7,9)	(8,11)	(9,12)
connected	1	1	5	22	107	814	4495
candidates	1	1	1	4	3	36	28
solvable with moves	1	1	1	4	3	31	5
finite solvable	1	1	1	4	3	36	27

Table 1.3: Solvability of minimal viewing graphs using our methods.

Example 1.4.17. The graph shown in Figure 1.8 is one of the five cases with $n = 8$, $e = 11$ that are "candidates" but are not "solvable with moves". However, we can show that this graph is actually solvable by arguing that the image of the pinhole 1 in the view 7 is fixed, even if this is not a consequence of the moves of Theorem 1.4.14 (this is represented by the gray dashed arrow in the figure). To prove this fact, one needs to keep track of more information, and record also when an epipole is constrained to a line (rather than only when an epipole is fixed, which is the purpose of dashed edges).[5] Indeed, we note that:

1. The image of the pinhole 1 in the view 5 lies on the line l_{51} defined by reprojection of the epipole e_{41} from the view 4 to the view 5.
2. Similarly, the image of the pinhole 1 in the view 6 lies on the line l_{61} defined by the reprojection of the epipole e_{31} from the view 3 to the view 6.
3. Using the previous two observations, we deduce that the image of the pinhole 1 in the view 7 lies on the line l_{71} obtained by transferring l_{51} and l_{61} to the view 7. This line is the projection from 7 of the intersection of the planes spanned by c_1, c_4, c_5 and c_1, c_3, c_6.
4. The image of the pinhole 1 in the view 7 belongs to another line m_{71} that is the reprojection of the epipole e_{81} from the view 8 to the view 7.
5. For generic pinholes the lines m_{71} and l_{71} will be different, so their intersection determines the image of the pinhole 1 in the view 7.

This argument shows that we can draw a dashed arrow from the vertex 1 to the vertex 7. We can now prove that G is solvable using the following sequence of moves (which starts from the situation illustrated at the right of Figure 1.8).

[5] One way to take this information into account is to define a new type of edge together with additional moves. We did not do this in Theorem 1.4.14, because it would have made the statement more complicated, and because this type of edge is never necessary for smaller graphs.

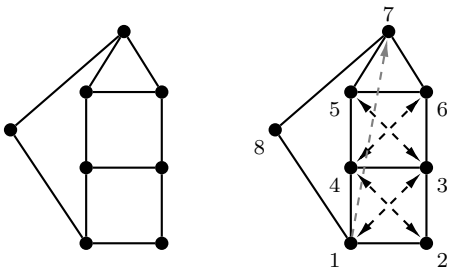


Figure 1.8: The viewing graph from Example 1.4.17.

- Draw double dashed arrows (4,7) and (3,7) (move II).
- Draw a dashed arrow from 1 to 5 (move II for the four-cycle (1,4,5,7)).
- Make (3,5) solid (move III), and then also (4,6), and (4,7) and (3,7) (move I).
- Make (1,7) double dashed (move II for (1,4,7,8)).
- Make (1,3) solid (move III) and complete the graph using move I.

♢

Example 1.4.18. The graph shown in Figure 1.9 is the only viewing graph with $n = 9$ and $e = 12$ that is "candidate" but is not "finite solvable". The fact that it is not finite solvable can also deduced without computations. Indeed, any finite solvable viewing graph of this size cannot impose any constraints on the fundamental matrices associated with its edges (this is because $d(9, 12) = 7 \times 12 - 11 \times 9 + 15 = 0$). However, the image of the pinhole 7 in the view 2 is over-constrained, because we can draw a dashed arrow $7 \to 2$ using move II for two distinct four-cycles ($(7, 1, 2, 5)$ and $(7, 3, 2, 5)$). This implies that the fundamental matrices associated with the edges of the graph cannot be arbitrary. ♢

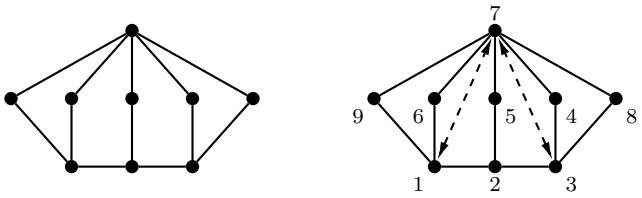


Figure 1.9: The viewing graph from Example 1.4.18.

Example 1.4.19. The graph shown in Figure 1.10 is not "solvable with moves", however we can show that it is solvable: indeed, the general algebraic compatibility equations from Section 1.4.6 are in this case simple enough that they can be solved explicitly (see the end of the chapter for the computations). The fundamental matrices associated with

the edges of this graph are unconstrained, which means that 12 arbitrary fundamental matrices can be used to recover a unique configuration of 9 cameras. ♢

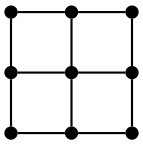


Figure 1.10: The viewing graph from Example 1.4.19.

1.5 Three-View Geometry and Multifocal Tensors

We now return to multi-view constraints, focusing on some aspects that we did not address in our general discussion in Section 1.3. In particular, we describe in Section 1.5.1 weak and strong characterizations of the joint image for $n = 3$ views. We also briefly summarize a more geometric viewpoint on trilinearities, which we used in [188]. In Section 1.5.2, we discuss the connection between multifocal tensors and the joint image. Our approach is based on a more general theory of "multi-graded Chow forms" that we introduced in [131], and that we briefly sketch here. We also make the observation that the nine "point-point-point" trilinearities encoded in the trifocal tensor provide a weak and not a strong characterization of point correspondences (Proposition 1.5.7). We conclude in Section 1.5.3 with some experiments from [188], where we used bilinearities and trilinearities to define a simple "trifocal-epipolar error" that can be minimized to estimate camera parameters for three views.

1.5.1 Trilinearities revisited

According to Theorem 1.3.19, for $n = 3$ views epipolar constraints cannot be used to provide a strong characterization of correspondences. This is also illustrated by Example 1.3.18. The following proposition explains how many constraints are actually necessary. The proof is computational; details are given at the end of the chapter.

Proposition 1.5.1. *Consider $n = 3$ cameras. We write b_{ij} for the bilinear (epipolar) constraint among cameras i and j.*

- *If the pinholes are non-collinear:*
  1. *For* any *trilinearity t that does not vanish on the product of the trifocal lines, the constraints $\{b_{12}, b_{13}, b_{23}, t\}$ provide a strong characterization of the joint image.*

2. *The epipolar constraints* $\{b_{12}, b_{13}, b_{23}\}$ *uniquely determine a camera configuration,* i.e.*, they provide a weak characterization of the joint image.*

- *If the cameras have collinear pinholes:*

  1. *A strong characterization of the joint image is given by* $\{b_{12}, b_{13}, b_{23}, t_1, t_2\}$ *where* t_1 *and* t_2 *are (sufficiently general) trilinear constraints.*

  2. *Two epipolar constraints together with one (sufficiently general) trilinearity* $\{b_{12}, b_{13}, t\}$ *uniquely determine camera geometry,* i.e.*, they give a weak characterization of the joint image.*

It is interesting to note that a similar conclusion can be reached using a purely geometric arguments. In the following, a *transversal* to some family of lines is a line intersecting every element of this family.

Proposition 1.5.2. *A necessary and sufficient condition for three lines to converge is that they be pairwise coplanar, and that they admit a transversal not contained in the planes defined by any two of them.*

Proposition 1.5.2 follows from the observation that if three lines are pairwise coplanar then exactly one of the following is true: 1) the lines are not all coplanar and intersect in one point; 2) they are all coplanar and intersect in one point; 3) they are coplanar, and intersect pairwise in three different points. Since only in cases 1) and 2) do the three lines admit transversals not contained in the planes defined by any two of them, Proposition 1.5.2 follows.

Algebraically, the condition for viewing lines to be pairwise coplanar is expressed (in terms of image coordinates) by the epipolar constraints. The existence of a transversal through a given point, on the other hand, can be imposed using a special class of trilinear polynomials. To show this, let us consider three lines ℓ_1, ℓ_2, ℓ_3 associated with Plücker vectors $\mathbf{p} = (p_{ij})$, $\mathbf{q} = (q_{ij})$, $\mathbf{r} = (r_{ij})$ respectively. For any $i = 1, 2, 3, 4$, we write $\mathbf{p}_{\hat{i}}$ for the subvector of $\mathbf{p}$ that does *not* contain Plücker coordinates with index i (for example, $\mathbf{p}_{\hat{1}} = (p_{23}, p_{24}, p_{34})$) and similarly for $\mathbf{q}$ and $\mathbf{r}$. We now define

$$t_i = \det[\,\mathbf{p}_{\hat{i}} \,|\, \mathbf{q}_{\hat{i}} \,|\, \mathbf{r}_{\hat{i}}\,], \quad i = 1, 2, 3, 4. \tag{1.28}$$

In the following, e_1, e_2, e_3, e_4 indicate the standard fundamental points in $\mathbb{P}^3$.

Lemma 1.5.3. *Given some integer* j *in* $\{1, 2, 3, 4\}$*, a necessary and sufficient condition for* ℓ_1*,* ℓ_2*, and* ℓ_3 *to admit a transversal passing through* e_j *is that* $t_j = 0$*.*

Proof. A necessary and sufficient condition for a line with Plücker coordinates $\mathbf{s} = (s_{ij})$ to pass through e_j is that $\mathbf{s}_{\hat{j}} = 0$. Writing $\mathbf{s}_j$ for the subvector of $\mathbf{s}$ that contains *only* Plücker coordinates with an index i (*i.e.*, the complement of $\mathbf{s}_{\hat{j}}$), then the line through e_j given by $\mathbf{s} = (s_{ij})$ will intersect the lines ℓ_1, ℓ_2, ℓ_3 if and only if $\mathbf{s}_j \cdot \mathbf{p}_{\hat{j}} = 0$, $\mathbf{s}_j \cdot \mathbf{q}_{\hat{j}} = 0$ and $\mathbf{s}_j \cdot \mathbf{r}_{\hat{j}} = 0$. This is equivalent to $t_j = 0$. □

If we consider now a triple of image points u_1, u_2, u_3 for three cameras P_1, P_2, P_3, then the corresponding viewing lines ℓ_1, ℓ_2, ℓ_3 defined by $\ell_i = P_i^{-1}(u_i)$ have Plücker coordinates that are linear expressions in the coordinates of the points (cf. Section 1.2.5). We conclude that the existence of a line transversal to the viewing rays and passing through one of the fundamental points can be expressed as trilinear relation $t_i(u_1, u_2, u_3)$ in image coordinates.

We next explain the relationship between these special trilinearities and the trilinear polynomials in the set S_3 from Definition 1.3.8. Given three cameras P_1, P_2, P_3, we fix a point $p = [\mathbf{p}]$ in $\mathbb{P}^3$, different from the pinholes. If u_1, u_2, u_3 are three image points with $u_i = [\mathbf{u}_i]$, we define the constraint

$$\begin{aligned} t_p(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) &= \det \begin{bmatrix} \mathbf{P}_1 & \mathbf{u}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{P}_2 & \mathbf{0} & \mathbf{u}_2 & \mathbf{0} & \mathbf{P}_2\mathbf{p} & \mathbf{0} \\ \mathbf{P}_3 & \mathbf{0} & \mathbf{0} & \mathbf{u}_3 & \mathbf{0} & \mathbf{P}_3\mathbf{p} \end{bmatrix} \\ &= \det \begin{bmatrix} \mathbf{P}_1 & \mathbf{u}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{P}_1\mathbf{p} \\ \mathbf{P}_2 & \mathbf{0} & \mathbf{u}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{P}_3 & \mathbf{0} & \mathbf{0} & \mathbf{u}_3 & \mathbf{P}_3\mathbf{p} & \mathbf{0} \end{bmatrix} \\ &= \det \begin{bmatrix} \mathbf{P}_1 & \mathbf{u}_1 & \mathbf{0} & \mathbf{0} & \mathbf{P}_1\mathbf{p} & \mathbf{0} \\ \mathbf{P}_2 & \mathbf{0} & \mathbf{u}_2 & \mathbf{0} & \mathbf{0} & \mathbf{P}_2\mathbf{p} \\ \mathbf{P}_3 & \mathbf{0} & \mathbf{0} & \mathbf{u}_3 & \mathbf{0} & \mathbf{0} \end{bmatrix}. \end{aligned} \tag{1.29}$$

The equalities between the different expressions follow from the multilinearity of the determinant. The following result shows that this constraint corresponds to the trilinearities from (1.28), where the role of the basis points e_i is replaced by the point p.

Proposition 1.5.4. *A necessary and sufficient condition for three image points u_1, u_2, u_3 to satisfy $t_p(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) = 0$ as defined in (1.29), is that the associated visual rays admit a transversal passing through p.*

Proof. An element $[\mathbf{q}, \lambda_1, \lambda_2, \lambda_3, \mu_1, \mu_2]$ in the null space of

$$\begin{bmatrix} \mathbf{P}_1 & \mathbf{u}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{P}_2 & \mathbf{0} & \mathbf{u}_2 & \mathbf{0} & \mathbf{P}_2\mathbf{p} & \mathbf{0} \\ \mathbf{P}_3 & \mathbf{0} & \mathbf{0} & \mathbf{u}_3 & \mathbf{0} & \mathbf{P}_3\mathbf{p} \end{bmatrix} \tag{1.30}$$

determines a point $q = [\mathbf{q}]$ in $\mathbb{P}^3$ which projects onto u_1 and on the lines joining u_2 and $p_2 = P_2(p)$ in the second image and u_3 and $p_3 = P_3(p)$ in the third image. This implies that the line $\ell = p \vee q$ is a transversal to the three visual rays. Conversely, it is easy to see that a transversal through p guarantees that (1.30) has a non-trivial null space. □

The formulation in (1.29) shows that $t_p(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) = 0$ is a linear combination of the trilinearities in S_3.

1.5.2 Multifocal tensors

We now turn to multifocal tensors. We recall that the study of three-view and four-view geometry is traditionally based on *trifocal* and *quadrifocal tensors*. These are respectively tensors of size $3 \times 3 \times 3$ and $3 \times 3 \times 3 \times 3$ which:

1. describe matching image features for 3 or 4 views, and
2. uniquely characterize a camera configuration.

For these characteristics, trifocal and quadrifocal tensors are similar to the fundamental matrix for two views. In particular, these tensors can be estimated linearly from image data in order to recover camera parameters. On the other hand, while the fundamental matrix characterizes point-point correspondences in two images, trifocal and quadrifocal tensors describe point-line-line and line-line-line-line matches, respectively. It is also known that analogous tensors do not exist for more than four views. We refer to [78] for more details.

We now propose a new framework for understanding multifocal tensors. This viewpoint helps illustrate the basic properties of these tensors, and emphasizes their relationship with the joint image variety. We will argue that multifocal tensors can be seen as an instance of a generalized version of the *Chow form* for an algebraic variety. We first recall some relevant definitions (see also Appendix B for basics on projective varieties).

Chow forms. Let X be a projective variety of dimension r in $\mathbb{P}^n$. To characterize X, we need in general at least $n - r$ conditions in the coordinates of $\mathbb{P}^n$ (over $\mathbb{C}$, this is strictly necessary). On the other hand, when the $\operatorname{codim} X = n - r$ is greater than one, we can use a *Chow form* to uniquely determine X using only *one* equation for appropriate coordinates. To see this, recall first that that from the geometric definition of dimension, we know that a generic linear space of codimension $\alpha = r + 1$ will *not* intersect X. The set of all linear spaces of codimension α which do meet X has the following properties:

1. It forms a hypersurface Z_X in the Grassmannian $\operatorname{Gr}(n - \alpha, \mathbb{P}^n)$, and can be characterized by a single equation in Plücker coordinates $F_X = 0$ (in addition to the Plücker relations).
2. The degree of F_X is the same as the degree of X
3. Given the F_X, it is possible to to recover the variety X. The idea here is that a point x in $\mathbb{P}^n$ will belong to X if and only if all $(n - \alpha)$-spaces passing through x meet X.

The polynomial form F_X associated with X is called the *Chow form* of X. For proofs, we refer the reader to [64].

As a simple example, we consider the standard twisted cubic in $\mathbb{P}^3$. This is a curve of degree three defined by $\{x_1x_3 - x_2^2 = x_2x_4 - x_3^3 = x_1x_4 - x_2x_3 = 0\}$. The set of lines in $\mathbb{P}^3$ that meet C form a three-fold inside $\mathrm{Gr}(1, \mathbb{P}^3) \subset \mathbb{P}^5$. Its ideal in $\mathbb{Q}[p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}]$ is generated by the Plücker quadric and by

$$p_{23}^3 - p_{13}p_{23}p_{24} + p_{12}p_{24}^2 + p_{13}^2p_{34} - p_{12}p_{14}p_{34} - 2p_{12}p_{23}p_{34}.$$

This polynomial represents Chow form of C.

Multi-graded Chow forms In [131], we proposed a natural generalization of the Chow construction for varieties inside a product of projective spaces. Our motivation was to provide an explanation to multi-view tensors, and to define analogous "multi-view forms" for non-central cameras (see Section 3.4.3). The following is a very brief summary of the main results in the paper.

Let X be a projective variety in $\mathbb{P}^{n_1} \times \ldots \times \mathbb{P}^{n_k}$ of dimension r. Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$ be a vector of indices with $0 \leq \beta_i \leq n_k$ such that $\sum_{i=1}^k \beta_i = r+1$. As before, by dimensionality constraints we expect that a general product of linear spaces $L_1 \times \ldots \times L_k$ where each L_i has dimension β_i will not intersect X. We thus consider the set

$$Z_{X,\boldsymbol{\beta}} = \{(L_1, \ldots, L_k) : X \cap (L_1 \cap \ldots \cap L_k) \neq \emptyset\} \subset \mathrm{Gr}(n - \beta_1, \mathbb{P}^{n_1}) \times \ldots \times \mathrm{Gr}(n - \beta_k, \mathbb{P}^{n_k}). \tag{1.31}$$

As we show in [131], in order for this set to actually provide a generalization of the Chow construction, we must assume some inequality conditions on the dimensions of projections of X. This is essentially because describing the expected dimensions of the intersection of varieties in a product of projective spaces is more complicated than in traditional projective spaces. In the following, the *multidegree* of the variety X in $\mathbb{P}^{n_1} \times \ldots \times \mathbb{P}^{n_k}$ is a polynomial in $\mathbb{Z}[t_1, \ldots, t_k]$ that generalizes the degree of a projective variety.

Theorem 1.5.5. *[131] The set $Z_{X,\boldsymbol{\beta}}$ in (1.31) is a hypersurface determining X if and only if for every non-empty $I \subsetneq \{1, \ldots, k\}$ we have*

$$\sum_{i \in I} \beta_i \leq \dim p_I(X),$$

where $p_I(X)$ is the projection of X on $\Pi_{i \in I} \mathbb{P}^{n_i}$. In this case, $Z_{X,\boldsymbol{\beta}}$ is the zero-set of a single multi-homogeneous polynomial $F_{X,\boldsymbol{\beta}}$ in Plücker coordinates. This defines the multi-graded Chow form *associated with X, for the indices $\boldsymbol{\beta}$.*

Moreover, if X has multidegree

$$\sum_{\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k)} a_{\boldsymbol{\gamma}} t_1^{\gamma_1} \ldots t_k^{\gamma_k}$$

then $F_{X,\boldsymbol{\beta}}$ has multidegree (as a multi-homogeneous polynomial)

$$(a_{\alpha_1+1,\alpha_2,\ldots,\alpha_k},\ldots,a_{\alpha_1,\alpha_2,\ldots,\alpha_k+1})$$

where $\alpha_i = n_i - \beta_i$ (i.e., the multidegree of the multi-graded Chow form is given by a subset of the coefficients of the multidegree polynomial of X).

Multi-graded Chow forms are also a generalization of "Grassmann tensors", introduced by Hartley and Schaffalitzky in [77].

Multifocal tensors and the joint image. Multifocal tensors are the result of applying the multi-graded Chow construction to the joint image variety M_n inside $(\mathbb{P}^2)^n$. This can be seen as follows.

Since $\dim M_n = 3$, we need to choose $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_n)$ such that $0 \leq \beta_i \leq 2$ and $\sum_{i=1}^n \beta_i = 4$. For any such $\boldsymbol{\beta}$, we can consider the set $Z_{M_n,\boldsymbol{\beta}}$ as in (1.31). By definition, this set includes n-tuples $(L_1,\ldots,L_n)$, where each L_i is a linear space of codimension β_i in $\mathbb{P}^2$, such that $L_1 \times \ldots \times L_n$ meets M_n. In other words the space $Z_{M_n,\boldsymbol{\beta}}$ contains products of spaces $(L_1,\ldots,L_n)$ for which there exists a correspondence $(u_1,\ldots,u_n)$ in M_n with u_i in L_i. For example, if $n = 3$ and $\beta = (2,1,1)$, the set $Z_{M_3,\boldsymbol{\beta}}$ contains triples point-line-line which "correspond" in the traditional sense of multi-view geometry.

According to Theorem 1.5.5, the set $Z_{M_n,\boldsymbol{\beta}}$ will:

1. be described by a single equation $F_{M_n,\boldsymbol{\beta}}$ in appropriate coordinates (note that Plücker coordinates for $\mathrm{Gr}(1,\mathbb{P}^2)$ are simply dual projective coordinates).
2. uniquely determine M_n (and thus the camera configuration, according to Theorem 1.3.16),

if and only if $\sum_{i\in I}\beta_i \leq 3$ for all $I \subsetneq \{1,\ldots,k\}$, since we have that $\dim p_I(M_n) = 3$ whenever $|I| \geq 2$. This is equivalent to $\beta_i > 0$ for all i, which together with $\sum_{i=1}^n \beta_i = 4$ clearly requires $n \leq 4$. The only possibilities are:

- For $n = 2$, $\boldsymbol{\beta} = (2,2)$. In this case, the condition $F_{M_2,\boldsymbol{\beta}} = 0$ describes corresponding points in $\mathbb{P}^2 \times \mathbb{P}^2$, *i.e.*, it is simply the epipolar constraint.
- For $n = 3$, $\boldsymbol{\beta} = (2,1,1)$ (or any such permutation). In this case, the condition $F_{M_3,\boldsymbol{\beta}} = 0$ describes "corresponding point-line-line" triples in $\mathbb{P}^2 \times (\mathbb{P}^2)^* \times (\mathbb{P}^2)^*$.
- For $n = 4$, $\boldsymbol{\beta} = (1,1,1,1)$. The condition $F_{M_4,\boldsymbol{\beta}} = 0$ describes "corresponding line-line-line-line" quadruplets in $(\mathbb{P}^2)^* \times (\mathbb{P}^2)^* \times (\mathbb{P}^2)^* \times (\mathbb{P}^2)^*$.

As a last point, we observe that the conditions $F_{M_n,\boldsymbol{\beta}}$ are *multilinear* in the the corresponding coordinates. Indeed, the joint image variety is known to have multidegree

$$t_1^2 t_2^2 \ldots t_n^2 \left(\sum_{1\leq i<j<k\leq n} \frac{1}{t_i t_j t_k} + \sum_{1\leq i,j\leq n} \frac{1}{t_i^2 t_j} \right). \tag{1.32}$$

See Corollary 3.5 in [4]. Since all the coefficients in this polynomial are equal to one, Theorem 1.5.5 says that $F_{M_n,\boldsymbol{\beta}}$ be multilinear. In particular, can identify the forms $F_{M_3,\boldsymbol{\beta}}$ and $F_{M_4,\boldsymbol{\beta}}$ defined above with *tensors*. We have thus recovered all the fundamental properties of multi-view tensors from the viewpoint of multi-graded Chow forms.

Example 1.5.6. We consider the cameras P_1, P_2, P_3 associated with the matrices

$$\mathbf{P}_1 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{P}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{P}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

A simple computation shows that if M_3 is the joint image variety associated with these cameras, the corresponding multi-graded Chow form for $\boldsymbol{\beta} = (2, 1, 1)$ is defined by

$$u_1l_2k_1 + u_2l_2k_1 + u_3l_3k_1 - u_2l_1k_2 + u_2l_2k_2 - u_1l_1k_3 - u_3l_1k_3 - u_2l_2k_3 + u_3l_2k_3 - u_3l_3k_3$$

in $\mathbb{Q}[u_1, u_2, u_3, k_1, k_2, k_3, l_1, l_2, l_3]$, where k, l are variables for dual projective coordinates. This represents a *trifocal tensor* (one of the three) associated with P_1, P_2, P_3. ♢

We conclude this discussion by pointing out that although a trifocal tensor determines the joint image, the nine trilinear conditions that are traditionally encoded in the tensors only provide a *weak* characterization of the joint image. Given a trifocal tensor $\mathcal{T}$ which distinguishes the first view ($\boldsymbol{\beta} = (2, 1, 1)$ in the multi-graded Chow set-up), the so-called "point-point-point" (PPP) trilinear conditions are usually written in tensor (Einstein summation) notation as

$$u_1^i u_2^j u_3^k (\epsilon_{jpr} \epsilon_{kqs} \mathcal{T}_i^{pq}) = 0_{rs}, \tag{1.33}$$

where ϵ_{ijk} the Levi-Civita permutation symbol. See [78, Chapter 15.2.1]. These trilinear constraints correspond to a subset of the maximal minors of the matrix $\mathcal{U}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ defined in (1.18), more precisely to the nine minors obtained by considering all rows associated with the first camera, and two rows from each of the other two. The following result shows that the algebraic set defined by these nine conditions contains the joint image variety strictly. We believe that this fact has not been pointed out in previous literature (although it is closely related to the known degeneracies of transfer based on the trifocal tensor [78, Section 15.3.2]).

Proposition 1.5.7. *For three cameras P_1, P_2, P_3 with non-collinear pinholes, the constraints (1.33) describe a set $\mathcal{W}$ which decomposes as $\mathcal{W} = M_3 \cup \mathcal{S}_{12} \cup \mathcal{S}_{13}$, where*

$$\begin{aligned} \mathcal{S}_{12} &= \{e_{12}\} \times \{e_{21}\} \times \mathbb{P}^2 \subset (\mathbb{P}^2)^3, \\ \mathcal{S}_{13} &= \{e_{13}\} \times \mathbb{P}^2 \times \{e_{31}\} \subset (\mathbb{P}^2)^3, \end{aligned} \tag{1.34}$$

and e_{ij} is the epipole in image i relative to the camera j.

The proof of this result is by direct computation, and a geometric justification is given in Figure 1.11. This fact is also consistent with the theory of multi-graded Chow forms.

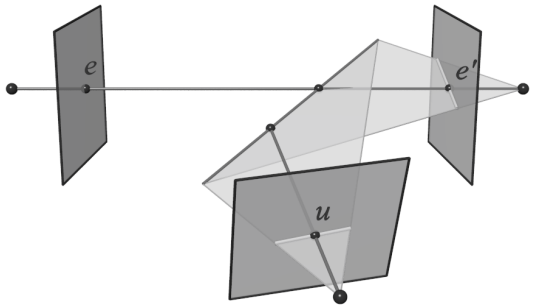


Figure 1.11: Geometric explanation for Proposition 1.5.7: the points e and e' are epipoles; for any choice of $u \in \mathbb{P}^2$ in the second image, all lines through e and e' will give rise to a point-line-line correspondence with u.

In fact, with the general notation introduced above, if we define $S_{X,\beta}$ as the set of points of points $(u_1, \ldots, u_k)$ in $\mathbb{P}^{n_1} \times \ldots \times \mathbb{P}^{n_k}$ such that every $L_1 \times \ldots \times L_k$ containing $(u_1, \ldots, u_k)$ meets X, then $S_{X,\beta}$ determines X, but will in general contain it strictly a distinguished irreducible component. See [131, Proposition 3.6].

1.5.3 Experiments: trinocular geometry

As an application of some of theoretical results from this chapter, we consider the problem of the estimation of camera parameters (structure-from-motion) for three views. In particular, we propose the minimization of a "trinocular-epipolar" error, given by the mean squared distance to the "epipolar" and "trinocular" lines, that we will define shortly. We present some experiments from [188] that show that epipolar constraints are sufficient for recovering camera parameters for non-degenerate configurations, while for collinear pinholes or for images of 3D points lying close to the trifocal plane, trilinear conditions are necessary.

In what follows, we assume that initial values for the projection matrices $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$ of three cameras have been estimated. This is possible using six triples of matching points, for a projective model [31, 151, 148].[6] The traditional approach for bundle-adjustment would proceed by minimizing a "geometric distance error" [78], which measures the mean squared distance of each given triple to the nearest image points which actually correspond, according to the current estimate of the camera parameters. For example, the contribution of a triple of image points $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$, assuming estimated parameters

[6] Three projection matrices provide an over-parameterization of the configuration space of three cameras. In [188], we also presented a minimal parameterization of three view geometry (defined on an open set of all configurations).

$\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$ can be written as:

$$\begin{aligned}\delta(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3 \,|\, \mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3) = \\ \min_{(\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \hat{\mathbf{u}}_3)} \{d(\mathbf{u}_1, \hat{\mathbf{u}}_1)^2 + d(\mathbf{u}_2, \hat{\mathbf{u}}_2)^2 + d(\mathbf{u}_3, \hat{\mathbf{u}}_3)^2\},\end{aligned} \tag{1.35}$$

where $(\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \hat{\mathbf{u}}_3)$ are triples that form a correspondence for $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$ (*i.e.*, belong to the joint image). This error function however has the disadvantage of not being directly computable from image data, since the nearest corresponding triple must be determined by solving a separate optimization problem. Indeed, bundle adjustment requires introducing a set of auxiliary variables that represent the coordinates of the original unknown 3D points. On the other hand, if we consider the "epipolar error" defined by

$$\begin{aligned}\delta_e(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3 \,|\, \mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3) = \\ \sum_{i=1}^{3} d(\mathbf{u}_i, E_{i,i+1}(\mathbf{u}_{i+1}))^2 + d(\mathbf{u}_i, E_{i,i+2}(\mathbf{u}_{i+2}))^2,\end{aligned} \tag{1.36}$$

where $E_{ij}(\mathbf{u}_j)$ represents the epipolar line in the image i corresponding to the point $\mathbf{u}_j$ in image j (defined by $\{\mathbf{u}_j^T \mathbf{F}_{ji} \mathbf{u} = 0\}$, where $\mathbf{F}_{ij}$ is the fundamental matrix), and addition is done modulo 3, then δ_e can easily be computed from image coordinates and camera parameters. In the next paragraph, we will define a "trinocular-epipolar" error that can also be computed very efficiently, and we will compare the effectiveness of these three error functions using both synthetic and real data. The geometric distance (1.35) will be used to evaluate the quality of the solutions recovered using the different approaches.

Non-collinear pinholes. In the case of three cameras with non-collinear pinholes, we define the trinocular-epipolar error as the average squared distance between each image point, two epipolar lines, and one "trinocular line" in each image. For a given triple $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$, the trinocular line in image i is the projection of the unique transversal through the fundamental point $e_4 = [0, 0, 0, 1]$ in $\mathbb{P}^3$ and the viewing lines corresponding to $\mathbf{u}_j, \mathbf{u}_k$ (with distinct i, j, k). Clearly, if the triple $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ were an exact correspondence, each image point would lie on the corresponding trinocular line. Moreover, the trinocular line in each image can be computed easily from the expression of trilinear constraint t_4 defined in Section 1.5.1. The contribution of a triple $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ to the trinocular-epipolar error can be written as

$$\begin{aligned}\delta_t(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3 \,|\, \mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3) = \\ \sum_{i=1}^{3} d(\mathbf{u}_i, E_{i,i+1}(\mathbf{u}_{i+1}))^2 + d(\mathbf{u}_i, E_{i,i+2}(\mathbf{u}_{i+2}))^2 \\ + d(\mathbf{u}_i, T_0^{i+1,i+2}(\mathbf{u}_{i+1}, \mathbf{u}_{i+2}))^2,\end{aligned} \tag{1.37}$$

where addition is always modulo 3, and denotes $T_0^{i+1,i+2}(\mathbf{u}_{i+1}, \mathbf{u}_{i+2})$ is the trinocular line in the i-th image. This error function can be easily computed using only image data. However, we should note that that the trinocular line is actually well-defined if and only if the point e_4 does not project onto (near) any of the image points. In practice, we can guarantee this by applying an appropriate homography of $\mathbb{P}^3$, imposing for example that e_4 it projects at infinity in the three images (which is possible in general). This corresponds to using a different trilinearity t_p, as defined in Section 1.5.1.

We first evaluated this approach on synthetic data: we considered 50 random configurations of triples of camera matrices (simulating realistic extrinsic and intrinsic parameters) and clouds of 20 points (see Figure 1.12), and successively added various amounts of Gaussian noise to the projected points. The averaged results are shown in Figure 1.13. We note that the trinocular-epipolar error gives slightly better results than the epipolar one, but bundle adjustment is the most accurate (as one would expect since it minimizes the geometric distance error (1.35) which is being reported). In this case, incorporating the trilinear constraint brings a slight improvement but does not seem to be essential: indeed, we know that for non-collinear pinholes, and three-dimensional points lying in general position, enforcing the epipolar constraints is sufficient for guaranteeing correspondence. However, if we position the 3D points gradually closer to the trifocal plane, we observe that the epipolar error gives increasingly worse results, ultimately failing in improving the initial reconstruction (Figure 1.14). In fact, when the epipolar lines are close to being coincident, measuring only epipolar distances can cause severe numerical instabilities. On the other hand, using the trinocular-epipolar error or bundle adjustment, the quality of the reconstruction seems to be independent of the distance to the trifocal plane.[7]

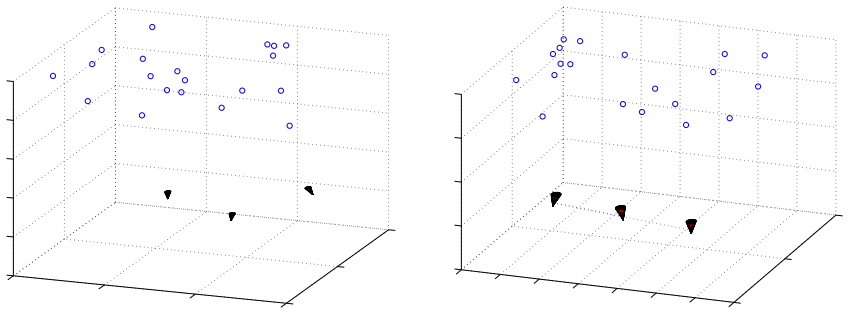


Figure 1.12: Examples of synthetic data: non-collinear pinholes (*left*) and collinear pinholes (*right*). The camera parameters defined random rotations and translations for camera motions, and produced feasible image sizes (around 500 pixels per dimension).

[7]We should point out that we could not place all the points exactly on the trifocal plane, since the minimal reconstruction method we used for initialization requires points to be in general position [31]. Moreover, coplanar points give rise to a critical configuration for three views [78], meaning that projectively non-equivalent triples of cameras can produce the same images.

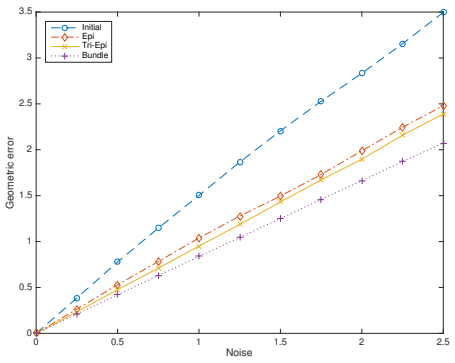


Figure 1.13: Non-collinear pinholes: quantitative results using synthetic datasets (20 points) with different amounts of Gaussian noise, in terms of the geometric distance error. The different lines represent the error of the initial approximation (Initial), and the errors of the solution refined using the trinocular-epipolar (Tri-Epi), epipolar (Epi) and bundle adjustment (Bundle) approach. The results are averages of 50 configurations of points and cameras.

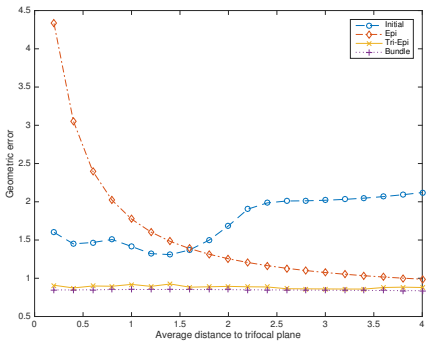


Figure 1.14: Near-degenerate datasets: quantitative results using synthetic datasets (20 points) with points lying near the trifocal plane, in terms of the geometric distance error. This experiment shows the same camera configuration and point sets increasingly close to the trifocal plane (constructed as points on the trifocal plane perturbed by gaussian noise).

We also evaluated the trinocular-epipolar approach on real data, using three images from the "house dataset" (courtesy of B. Boufama and R. Mohr). Figure 1.15 shows 38 correspondences between the images from this dataset, and the corresponding epipolar and trinocular lines, after camera parameters were recovered using the trinocular-epipolar approach. Table 1.4 shows the average distances between the data points and these lines. The mean distance to epipolar lines is on the order of 1pixel, and comparable to that obtained by classical techniques for estimating the fundamental matrix from pairs of images on the same data [57, Ch. 8]. Finally, Table 1.5 compares the results of the

different approaches, in terms of the geometric distance error. In this case, we see that the reconstructions obtained using the trinocular-epipolar and the epipolar approach are both very accurate.

Epi12	*Epi13*	*Tr1*	*Epi23*	*Epi21*	itTr2	*Epi31*	*Epi32*	*Tr3*
0.81	0.81	1.03	0.94	0.78	0.79	0.75	0.86	1.28

Table 1.4: Average distances (in pixels) to epipolar and trinocular lines for the house dataset (Figure 1.15). Here, "Epiij" refers to the distance between points in image i and the corresponding epipolar lines associated with image j, and "Trj" refers to the distance between points in image j and the corresponding trinocular line associated with the other two images.

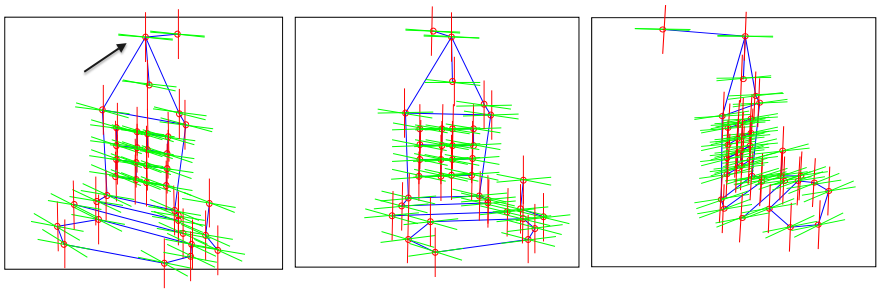


Figure 1.15: Estimated epipolar and trinocular lines. Note that the two families of epipolar lines associated with an image typically contain (near) degenerate pairs (such as the pair shown in the first image): these can be disambiguated using trilinearities.

Tri-Epi	*Epi*	*Bundle adj*
0.73	0.73	0.72

Table 1.5: Quantitative comparison between different approaches for the house dataset (Figure 1.15), in terms of the geometric distance error (measured in pixels).

Collinear pinholes. For three cameras with collinear pinholes, the "trinocular-epipolar" error is defined as the average squared distance of each image point to the three epipolar lines and to two different "trinocular lines" (following Proposition 1.5.1).

$$\begin{aligned}&\delta_t(\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3\,|\,\mathbf{P}_1,\mathbf{P}_2,\mathbf{P}_3) = \\ &\sum_{i=1}^{3} d(\mathbf{u}_i, E_{i,i+1}(\mathbf{u}_{i+1}))^2 + d(\mathbf{u}_i, E_{i,i+2}(\mathbf{u}_{i+2}))^2 \\ &+\; d(\mathbf{u}_i, T_A^{i+1,i+2}(\mathbf{u}_{i+1},\mathbf{u}_{i+2}))^2 + \; d(\mathbf{u}_i, T_B^{i+1,i+2}(\mathbf{u}_{i+1},\mathbf{u}_{i+2}))^2,\end{aligned} \tag{1.38}$$

where $T_A^{i+1,i+2}(\mathbf{u}_{i+1}, \mathbf{u}_{i+2})$ and $T_B^{i+1,i+2}(\mathbf{u}_{i+1}, \mathbf{u}_{i+2})$ are the trinocular lines in the i-th image, defined by the trilinear constraints t_3, t_4 from Section 1.5.1. Once again, the trinocular lines are well defined if and only if e_3 and e_4 do not project onto (near) any of the given image points: in order to satisfy this condition, we select two points which project very far from the image data, and apply a homography of $\mathbb{P}^3$ which maps e_3 and e_4 to these points.

Remark 1.5.8. By restricting ourselves to configurations with collinear pinholes, the degrees of freedom of the camera geometry drop from 18 to 16. This suggests that it is possible to recover initial estimates of the matrices $\mathbf{P}_i$ using fewer correspondences, and indeed we have implemented a method that recovers projection matrices for collinear pinholes given only five triples of correspondences and one correspondence between two images. This minimal method yields up to six real solutions. However, the simpler six-point algorithm from [31] still works in this case.

Figure 1.16 shows the results of our experiments using synthetic data (20 points with various amounts of Gaussian noise added to the projections). We see that in this case the epipolar error fails in recovering the correct parameters, while the trinocular-epipolar error is able to achieve essentially optimal solutions. Indeed, we know that in the collinear case the epipolar constraints are never sufficient to guarantee correspondence.

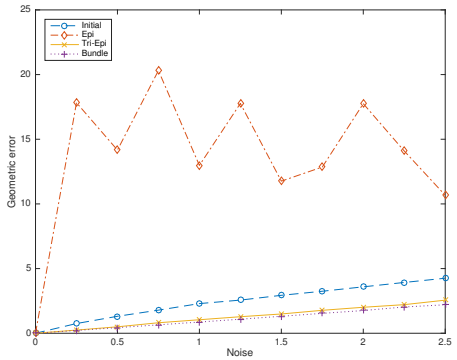


Figure 1.16: Collinear pinholes: Quantitative results using synthetic datasets (20 points) with different amounts of Gaussian noise, in terms of the geometric distance error. The results are averages of 50 configurations of points and cameras.

Finally, we evaluate our approach on real data, using three pictures taken from collinear viewpoints (allowing rotations about the camera's axis) and matching points using SIFT descriptors (Figure 1.17). Table 1.7 shows the average distances between the data points and the epipolar and trinocular lines for the dataset with collinear pinholes and Figure 1.18 displays the estimated epipolar and trinocular lines. Table 1.6 shows our quantitative results, which confirm that for collinear pinholes the trilinear components of the error are actually necessary.

Tri-Epi	*Epi*	*Bundle adj*
0.68	5.33	0.67

Table 1.6: Quantitative comparison between different approaches for the office dataset (Figure 1.17), in terms of the geometric distance error (measured in pixels).

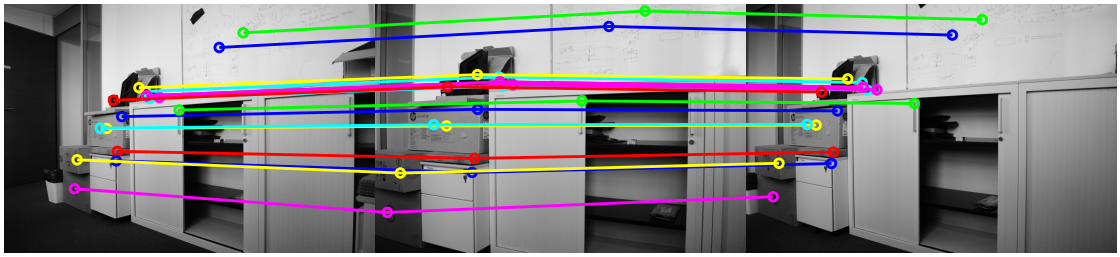


Figure 1.17: Three pictures taken from collinear viewpoints, and point correspondences obtained by matching SIFT descriptors.

Epi12	*Epi13*	*Tr1A*	*Tr1B*	*Epi23*	*Epi21*	*Tr2A*	*Tr2B*	*Epi31*	*Epi32*	*Tr3A*	*Tr3B*
0.60	0.19	0.49	2.95	0.74	0.76	0.77	0.74	0.21	0.66	2.80	0.84

Table 1.7: Average distances (in pixels) to epipolar and trinocular lines for the collinear pinhole dataset (Figure 1.17). Here, "Epiij" refers to the distance between points in image i and the corresponding epipolar lines associated with image j, and "TrjA" and "TrjB" refer to the distances between points in image j and the corresponding two trinocular lines associated with the other two images.

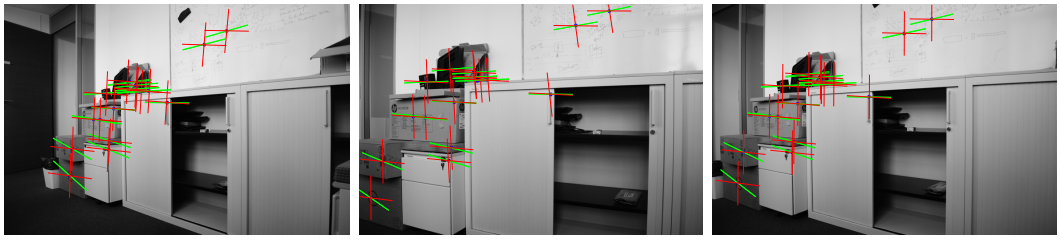


Figure 1.18: Estimated epipolar and trinocular lines. Note that the pairs of epipolar lines essentially coincide, while trinocular lines are always transversal.

1.6 Conclusions

We have presented some topics on the theory of pinhole cameras, focused mainly on the problem of explaining which sets of multi-view constraints can be used 1) to characterize point correspondences and 2) to determine a camera configuration. Indeed, one of the central themes of this chapter is the fact that these two properties are not exactly the same: this is why we defined "weak" and "strong" characterizations of the joint image in Section 1.3. For example, we have seen in Section 1.4 that fewer than $2n-3$ epipolar constraints can be used to determine a camera configuration, even though these will characterize a set in $(\mathbb{P}^2)^n$ that is much larger than the set of all point correspondences. This distinction is on the other hand relatively subtle, since a camera configuration clearly determines all point correspondences, and thus any weak characterization can be used to *indirectly* recover a strong one.

Several questions could deserve further work. In addition to the open problem mentioned at the end Section 1.4.6, we believe that it would be very useful to gain a complete understanding of the algebraic constraints that characterize the *compatibility* of fundamental matrices (*i.e.*, study the algebraic set $\mathcal{X} \subset (\mathbb{P}^8)^N$ considered in Section 1.4.4). This is an issue that has not been considered much in classical multi-view geometry, but is clearly important in practice, since global compatibility constraints are not automatically satisfied using local measurements (*e.g.*, pairwise matches) [177]. More generally, it would be helpful to develop more concrete models for representing the space of camera configurations (or, equivalently, for parameterizing joint-images), beyond traditional multifocal tensors.

Finally, the relationship between point correspondences and camera configurations resembles, in a vague sense, a sort of "duality". This idea can made more precise by appropriately restricting the choice of image coordinates, so that pinholes and scene points play (almost) symmetric roles. This is the basis of *Carlsson-Weinshall duality* [30]. We believe that this duality can be useful to gain further insight on many of the issues addressed in this chapter: we are currently revisiting it from the perspective of the joint images and point configurations [36].

1.7 Proofs

This section collects some proofs. Propositions that appeared previously will be stated here with the same numbering, while auxiliary results will be numbered independently.

1.7.1 The joint image

We first prove Proposition 1.3.5 from Section 1.3, that describes the relationship between the open joint image M_n^o and its closure M_n.

Proposition 1.3.5. *Given $n \geq 3$ cameras with non-collinear distinct pinholes, one has $M_n^o = M_n \setminus \mathcal{E}_n$, where*

$$\mathcal{E}_n = \bigcup_{i=1}^{n} \{e_{1i}\} \times \ldots \times \mathbb{P}^2_{(i)} \times \ldots \times \{e_{ni}\}. \tag{1.13}$$

Here $\mathbb{P}^2_{(i)}$ indicates $\mathbb{P}^2$ at position i in the product, and e_{ij} denotes the j-th epipole in the i-th image. However, if $n = 2$, or more generally if the cameras have collinear pinholes, then

$$\mathcal{E}_n = \bigcup_{i=1}^{n} \{e_1\} \times \ldots \times \mathbb{P}^2_{(i)} \times \ldots \times \{e_n\} \setminus \{(e_1, \ldots, e_n)\}, \tag{1.14}$$

where e_i now denotes the unique epipole in image i (so $e_i = P_i(c_j)$ for all $j \neq i$).

Proof. In Section 1.3.3, we introduced the matrix

$$\mathcal{U}(\mathbf{u}_1, \ldots, \mathbf{u}_n) = \begin{bmatrix} P_1 & \mathbf{u}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ P_2 & \mathbf{0} & \mathbf{u}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_n & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{u}_n \end{bmatrix}, \tag{1.39}$$

and pointed out that joint image variety can be characterized set-theoretically by

$$M_n = \{(\mathbf{u}_1, \ldots, \mathbf{u}_n) \in (\mathbb{P}^2)^n \mid \mathcal{U}(\mathbf{u}_1, \ldots, \mathbf{u}_n) \text{ is not full rank}\}.$$

If the camera pinholes are not all collinear, the null space of $\mathcal{U}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ has dimension at most one, and $(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ belongs to $\mathcal{E}_n = M_n \setminus M_n^o$ if and only if annihilating vectors $[\mathbf{p}, \lambda_1, \ldots, \lambda_n] \in \mathbb{R}^{4+n}$ are such that $\lambda_i = 0$ for some i: in this case $\mathbf{p}$ must be one of the camera centers, and we see that $\mathcal{E}_n$ is given by (1.13). If the cameras are collinear, then the same reasoning applies except when $(\mathbf{u}_1, \ldots, \mathbf{u}_n) = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ (because the null space of $\mathcal{U}(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ has dimension two): in this case, it is easy to realize that $(\mathbf{e}_1, \ldots, \mathbf{e}_n) \in M_n^o$. □

We next give a proof of Proposition 1.3.11. The result is implied by Theorem 3.6 in [4], but the case of multilinear constraints (multidegree $(1, \ldots, 1)$) is not pointed out. Our proof is also adapted from the one given in [4], but we use more direct argument to show Proposition 1.7.3.

Proposition 1.3.11. *The multilinear polynomials that vanish on M_n form a vector space of dimension $d_n = 3^n - \binom{n+3}{3} + n$.*

Remark 1.7.1. In the rest of this section, we will assume that the camera matrices $P_1, \dots, P_n$ satisfy a generality assumption, namely that all 4×4 matrices formed by any four distinct rows from $P_1, \dots, P_n$ have full rank. If the camera pinholes are distinct, this condition can be guaranteed by changing bases appropriately in the different image planes. Since this kind of operation preserves the dimension of spaces generated by multilinearities (the vector spaces are mapped isomorphically onto themselves), this assumption does not affect the general validity of the result.

Proof. Let W_n be the vector space generated by multlinear relations arising as maximal minors of $\mathcal{U}(\mathbf{u}_1, \dots, \mathbf{u}_n)$ defined in (1.18); for simplicity we will refer to these constraints as the "fundamental" multilinearities. We will first compute the dimension of W_n; then we will show that W_n coincides with the space of *all* multilinear constraints which vanish on M_n. Our proof heavily relies on the fact that the fundamental multilinearities form a *Gröbner basis* for lexicographic ordering (and in fact, for all admissible term orderings): this will be shown in Proposition 1.7.3.

To compute the dimension of W_n we will use the following well known property of Gröbner bases (see for example [39]):

> (P) If a set of homogeneous polynomials $f_1, \dots, f_N$ forms a Gröbner basis (for a fixed monomial ordering), then $\dim_{\mathbb{R}}(f_1, \dots, f_N) = \dim_{\mathbb{R}}(In(f_1), \dots, In(f_N))$, where $In(f_i)$ denotes the initial term of f_i.

To use this result, we first introduce a lexicographic ordering for the monomials in the ring $\mathbb{R}[x_1, y_1, z_1, \dots, x_n, y_n, z_n]$, so that $x_1 \succ \cdots \succ x_n \succ y_1 \succ \cdots \succ y_n \succ z_1 \cdots \succ z_n$. In Lemma 1.7.2, we will show that for this ordering the set of leading terms of the fundamental mutlilinearities is given by $B = C_n \setminus A$, where

$$\begin{aligned} C_n = \{x^{\boldsymbol{\alpha}} y^{\boldsymbol{\beta}} z^{\boldsymbol{\gamma}} \mid \\ \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{N}^n, \;\; \boldsymbol{\alpha} + \boldsymbol{\beta} + \boldsymbol{\gamma} = (1, \dots, 1)\}, \end{aligned} \tag{1.40}$$

is the set of all multilinear monomials, and

$$\begin{aligned} A = &\{x^{\boldsymbol{\alpha}} y^{\boldsymbol{\beta}} z^{\boldsymbol{\gamma}} \in M_n \mid \boldsymbol{\alpha} = 0, |\boldsymbol{\beta}| \leq 3\} \\ &\cup \{x^{\boldsymbol{\alpha}} y^{\boldsymbol{\beta}} z^{\boldsymbol{\gamma}} \in M_n \mid |\boldsymbol{\alpha}| = 1, \, |\boldsymbol{\beta}| \leq 1\}, \end{aligned} \tag{1.41}$$

where we used multi-index notation, so that for example $x^{\boldsymbol{\alpha}} = \prod_k x_k^{\boldsymbol{\alpha}(k)}$.

According to property (P), we know that $\dim(W_n) = \dim(\mathrm{Span}(B))$. Since all monomials are linearly independent, we can now compute the dimension of W_n using a simple

counting argument:

$$\begin{aligned}\dim(W_n) = |T| &= |M_n| - |S| = \\ &= 3^n - \left(1 + n + \binom{n}{2} + \binom{n}{3}\right) - n\Big(1 + (n-1)\Big) \\ &= 3^n - \left(1 + n + \binom{n}{2} + \binom{n}{3} + n(n+1)\right) + n \\ &= 3^n - \binom{n+3}{3} + n.\end{aligned} \tag{1.42}$$

Finally, to show that W_n coincides with the space of all multilinear constraints, we recall that the fundamental multilinearities set-theoretically define the joint image variety M_n (they give necessary and sufficient conditions for correspondence). Using the *Nullstellensatz Theorem*, it is enough to prove that they generate a radical ideal[8]. This will also be shown in Proposition 1.7.3. □

Lemma 1.7.2. *Using the ordering $x_1 \succ \cdots \succ x_n \succ y_1 \succ \cdots \succ y_n \succ z_1 \cdots \succ z_n$, and assuming the condition in Remark 1.7.1, the leading terms of the fundamental multilinearities are given by $B = C_n \setminus A$, where C_n and A are defined in (1.40) and (1.41).*

Proof. From (1.40), we see that a monomial $x^{\boldsymbol{\alpha}}y^{\boldsymbol{\beta}}z^{\boldsymbol{\gamma}}$ in C_n can be also identified by the vector $\boldsymbol{\nu} = \boldsymbol{\alpha} + 2\boldsymbol{\beta} + 3\boldsymbol{\gamma} \in \{1,2,3\}^n$; in this case we will write $\mu_{\boldsymbol{\nu}} = x^{\boldsymbol{\alpha}}y^{\boldsymbol{\beta}}z^{\boldsymbol{\gamma}}$. One can verify that $\mu_{\boldsymbol{\nu}_1}$ can be a leading term of a fundamental multilinearity if and only if there exist $\boldsymbol{\nu}_2$, $\boldsymbol{\nu}_3$, $\boldsymbol{\nu}_4$ such that $\mu_{\boldsymbol{\nu}_1} \succ \mu_{\boldsymbol{\nu}_i}$ and $|\boldsymbol{\nu}_1 - \boldsymbol{\nu}_i| \leq 4$ for $i = 2,3,4$. Indeed, a fundamental mutilinearity P is defined by choosing $n+4$ rows of $\mathcal{U}$ from (1.18), and the monomials $\mu_{\boldsymbol{\nu}}$ appearing in P correspond to subsets of n rows (which are expressed by the vectors $\boldsymbol{\nu} \in \{1,2,3\}^n$); note also that the generality assumption guarantees that all possible coefficients are non-zero. It is straightforward to verify that $\mu_{\boldsymbol{\nu}_1}$ does not satisfy the previous property if and only if it belongs to the set A defined in (1.41). □

Proposition 1.7.3. *The fundamental multilinearities form a universal Gröbner basis (*i.e.*, a Gröbner basis relative to any monomial ordering). Moreover, they define a radical ideal.*

Proof. Our proof is motivated by a classical result, which states that in the ring $R = \mathbb{R}[x_{ij}]$, where $i = 1,\ldots,m$, $j = 1,\ldots,n$, and $m \leq n$, the maximal $m \times m$ minors of $X = (x_{i,j})$ form a universal Gröbner basis (generating the *classical determinantal ideal*) [126].

[8] Assuming the ideal J generated by the fundamental multilinearities is radical, we can write $J = I \cap \bigcap_i (x_i, y_i, z_i)$ where $I = I(M_n)$ is the prime ideal describing M_n (defined by bilinear, trilinear, and quadrilinear relations [4]). Any multilinearity that vanishes on M_n lies in I and in (x_i, y_i, z_i) for all i, thus it must belong to J.

In order to obtain a similar statement for the maximal minors of the matrix $\mathcal{U}$, we use the following two results:

1. If Z is a subset of variables in $\mathbb{R}[x_{ij}]$, and X' is the matrix obtained from $X = (x_{i,j})$ by setting the variables in Z to zero, then the nonzero maximal minors of X' form a universal Gröbner basis [18].

2. Let $f_1, \ldots, f_N$ be a set of polynomials forming a universal Gröbner basis in the ring $\mathbb{R}[s_1, \ldots, s_k, t_1, \ldots, t_j]$, and consider the evaluation map

$$\begin{aligned} \phi : \mathbb{R}[s_1, \ldots, s_k, t_1, \ldots, t_j] &\to \mathbb{R}[s_1, \ldots, s_k] \\ g(s_1, \ldots, s_k, t_1, \ldots, t_j) &\mapsto g(s_1, \ldots, s_k, c_1, \ldots, c_j) \end{aligned} \tag{1.43}$$

   for a fixed set $c_1, \ldots, c_j \in \mathbb{R}$. Thinking $f_1, \ldots, f_N$ as polynomials with ring coefficients, *i.e.*, as elements of $R[s_1, \ldots, s_k]$ with $R = \mathbb{R}[t_1, \ldots, t_j]$, we denote with $Lc(f_i) \in R$ the leading coefficient of f_i. If $\phi(Lc(f_1)), \ldots, \phi(Lc(f_N))$ are non-zero, then $\phi(f_1), \ldots, \phi(f_N)$ form a universal Gröbner basis in $\mathbb{R}[s_1, \ldots, s_k]$ [97].

We now consider the matrix $\mathcal{U}'$ as in (1.18), but with indeterminate camera matrices: according to the first result given above, the maximal minors form a universal Gröbner basis. The fundamental multilinearities are obtained from these minors by substituting the actual entries of the camera matrices. If the leading coefficients of the minors are not mapped to zero under this specialization, then we can use the second result and obtain the claim. However, it is easy to realize that the leading coefficients are given by the determinants of the matrices obtained by selecting four rows from the camera matrices; assuming the generality condition from the Remark, these do not vanish for the actual camera entries. This implies that the fundamental multilinearities form a universal Gröbner basis.

Finally, we now see that the monomials in $B = C_n \setminus A$ (defined in (1.40) and (1.41)) generate the *initial ideal* for the ideal generated by fundamental multilinearities. The fact that this monomial ideal is squarefree, allows us to conclude that the fundamental multilinearities generate a radical ideal; see [82, Corollary 2.2]. □

We next show that the joint image variety is closely related to the *blow up*, a fundamental construction in algebraic geometry [79]. For our purposes we can give the following definition for the blow up:

Definition 1.7.4. Consider a smooth algebraic variety X of dimension n, and a subvariety $Z \subseteq X$ of dimension d. Locally, we may write X as $Z \times W$ where W has dimension $n - d$ and is transversal to Z. Let $\lambda : X \setminus Z \to \mathbb{P}^{n-d-1}$ be the map associating with $x = (z, w) \in Z \times W$ the line ℓ_x passing through w and z. The *blow up* of X in Z is given by closure of the graph of λ:

$$\tilde{X}_Z = \overline{\{(x, \ell_x), x \in X \setminus Z\}} \subseteq X \times \mathbb{P}^{n-d-1}. \tag{1.44}$$

This is a smooth variety, with a natural map $\pi : \tilde{X}_Z \to X$, known as the *blow up map*. The *exceptional locus* is the inverse image $\tilde{Z} = \pi^{-1}(Z) = Z \times \mathbb{P}^{n-d-1}$ of the center Z. It is is the set of points where π fails to be a local isomorphism: in fact, π always contracts the second factor $\mathbb{P}^{n-d-1}$ of $\tilde{Z}$ to a point. See Figure 2 in the main part of paper.

Basically, the blow up replaces a subvariety $Z \subset X$ with all the directions in X pointing out of Z. Blowups are an extremely important tool for the *resolution of singularities*, that is, the operation of constructing suitable smooth models for varieties with singularities. We refer to [80] for a nice introduction.

Proposition 1.7.5. *If $n \geq 3$ cameras do not have collinear pinholes, the joint image variety M_n is isomorphic to the blow up of $X = \mathbb{P}^3$ at $Z = \{c_1, \ldots, c_n\}$, where c_i, $i = 1, \ldots, n$ are the camera pinholes.*

Proof. Iterating the construction given in Definition 1.7.4, we see that the blow up of $\mathbb{P}^3$ at $Z = \{c_1, \ldots, c_n\}$ is given by

$$\tilde{X}_Z = \overline{\{(p, \ell_p^1, \ldots, \ell_p^n),\ p \in \mathbb{P}^3 \setminus Z\}} \subseteq \mathbb{P}^3 \times (\mathbb{P}^2)^n, \tag{1.45}$$

where ℓ_p^i denotes the line through p and c_i. In particular, the projection of $\tilde{X}_Z$ onto $(\mathbb{P}^2)^n$ is exactly the joint image variety M_n, so we have a natural map

$$f : \tilde{X}_Z \to M_n. \tag{1.46}$$

In order to prove that f is an isomorphism we have to show that: 1) f is injective 2) the differential Tf of f is injective.

The fact that f is injective for $(p, \ell_p^1, \ldots, \ell_p^n)$ with $p \in \mathbb{P}^3 \setminus Z$ follows from the observation that the "joint projection map" $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2 \times \ldots \times \mathbb{P}^2$ is injective (where it is defined) if one assumes non- collinear pinholes. Moreover, if $p = c_i$ for some i, then the exceptional set $\pi^{-1}(c_i)$ is mapped isomorphically onto $\{(e_{1i}, \ldots, \ell, \ldots e_{ni}),\ \ell \in \mathbb{P}^2\} \subseteq M_n$, where e_{ji} is the epipole in the j-th image for the center c_i (note that the image of the exceptional locus $f(\pi^{-1}(Z))$ is exactly the set $\mathcal{E}_n$ given by Proposition 1.3.5).

To prove that the differential Tf is injective at points $(p, \ell_p^1, \ldots, \ell_p^n)$ with $p \in \mathbb{P}^3 \setminus Z$, it is enough to observe that all lines in $\mathbb{P}^3$ are mapped injectively by the joint projection map (since f is locally an affine map, if the differential were not injective, some line would have to be contracted). Similarly, if $p = c_i$, then all lines contained in $\pi^{-1}(c_i)$ are mapped isomorphically on some image $\mathbb{P}^2$, and the same holds for all lines passing through c_i in $\mathbb{P}^3$. □

Proposition 1.3.7 (Singularities of the joint image variety)**.** *When the camera pinholes are not collinear, M_n is smooth. When they are collinear (in particular for $n = 2$ views), then M_n has a unique singular point given by the n-tuple of epipoles $(e_1, \ldots, e_n)$.*

Proof. If the camera pinholes are not collinear, then Proposition 1.7.5 immediately implies that the joint image variety is smooth. If the pinholes are collinear, then the projection map $f : \tilde{X}_Z \to M_n$ considered in the previous proof is still well defined, and is a local isomorphism except at points that lie on the baseline $\ell \subseteq \mathbb{P}^3$ containing the centers (more precisely, on its *strict transform* in $\tilde{X}_Z$, given by $\overline{\pi^{-1}(\ell \setminus Z)}$). In fact, f contracts this set onto $(e_1, \ldots, e_n)$. In general, however, birational morphisms between smooth varieties can only contract sets of codimension 1 [118, Theorem 2.2], while the baseline has codimension 2: we conclude that $(e_1, \ldots, e_n)$ must be the only singular point of M_n. □

We now turn to algebraic characterizations of the joint image, and prove Theorem 1.3.19.

Theorem 1.3.19. *Assume n cameras are given.*

1. *Bilinearities and trilinearities (the sets S_2, S_3 from Definition 1.3.8)* always *strongly characterize M_n, independently of the pinhole configuration.*

2. *Bilinear constraints* alone *strongly characterize M_n, if and only if the pinholes are* not all coplanar

3. *Bilinear constraints* alone *weakly characterize M_n if and only if the pinholes are* not all collinear.

Proof. Let us first assume that $n = 4$. If $P_1, \ldots, P_4$ are cameras with pinholes $c_1, \ldots, c_n$ in *general position* (*i.e.*, non coplanar), then up to homographies of $\mathbb{P}^3$ (that do not affect the joint image) and homographies of image planes (that simply result in linear changes of variables that map k-linearities isomorphically onto themselves) we may assume that:

$$\mathbf{P}_1 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{P}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \\ \mathbf{P}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad \mathbf{P}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{1.47}$$

In this case, one can verify using Gröbner bases that the six epipolar relations are sufficient to generate, up to irrelevant components, the ideal associated to the joint image variety M_4 (see also [84]).

If $P_1, \ldots, P_4$ are cameras with pinholes $c_1, \ldots, c_4$ in *general coplanar position* (*i.e.*, copla-

nar with no subset of three that are collinear), then we may similarly assume that:

$$\begin{aligned}\mathbf{P}_1 &= \begin{bmatrix}1 & 0 & 0 & 1\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix}, \quad \mathbf{P}_2 = \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 1\\ 0 & 0 & 1 & 0\end{bmatrix},\\ \mathbf{P}_3 &= \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 1\end{bmatrix}, \quad \mathbf{P}_4 = \begin{bmatrix}1 & 0 & 0 & 1/3\\ 0 & 1 & 0 & 1/3\\ 0 & 0 & 1 & 1/3\end{bmatrix}.\end{aligned} \tag{1.48}$$

This time, using Gröbner basis we see that the epipolar conditions describe an algebraic set $\mathcal{W}$ that is strictly larger than M_4. In fact, $\mathcal{W}$ decomposes as

$$\mathcal{W} = M_4 \cup \mathcal{V}_t \tag{1.49}$$

where $\mathcal{V}_t = \{x_i + y_i + z_i = 0 \,|\, i = 1, \ldots, 4\}$ is the product of the "trifocal lines" (the projection of the plane containing the pinholes). We see that the epipolar constraints are only a *weak* characterization of the joint image: camera geometry can be recovered (for example by considering subsets of three cameras, see [78]) although correspondences are not directly characterized. On the other hand, one can verify that including trilinear relations is sufficient to exclude the spurious solutions and yields a strong characterization of M_4

If $P_1, \ldots, P_4$ are cameras with pinholes $c_1, \ldots, \mathbf{c}_n$ that are *coplanar, and with a subset of three that are collinear*, then we can consider

$$\begin{aligned}\mathbf{P}_1 &= \begin{bmatrix}1 & 0 & 0 & 1\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix}, \quad \mathbf{P}_2 = \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 1\\ 0 & 0 & 1 & 0\end{bmatrix},\\ \mathbf{P}_3 &= \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 1\end{bmatrix}, \quad \mathbf{P}_4 = \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 1/2\\ 0 & 0 & 1 & 1/2\end{bmatrix}.\end{aligned} \tag{1.50}$$

Once again, we see that that the bilinear and trilinear constraints completely characterize M_4, while only bilinear conditions describe a set $\mathcal{W}$ that also decomposes as

$$\mathcal{W} = M_4 \cup \mathcal{V}_t \tag{1.51}$$

where $\mathcal{V}_t = \{x_i + y_i + z_i = 0 \,|\, i = 1, \ldots, 4\}$ is the product of the "trifocal lines".

Finally, if all $P_1, \ldots, P_4$ have collinear pinholes, we cannot simply change reference frame in $\mathbb{P}^3$ and assume the cameras to be fixed, since four collinear points are not always projectively equivalent (the projective invariant is given by the cross-ratio). However,

we can use a single parameter to describe all such camera configurations:

$$\mathbf{P}_1 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{P}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \\ \mathbf{P}_3 = \begin{bmatrix} 1 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{P}_4 = \begin{bmatrix} 1 & 0 & 0 & t \\ 0 & 1 & 0 & 1-t \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{1.52}$$

In this case, computing the primary decomposition we see that bilinear and trilinear constraints have spurious components only for $t = 0, 1/2, 1$ which correspond to non-valid values for t (we assume the camera pinholes are distinct). Bilinear constraints, on the other hand, characterize (for feasible values of of t) a *single irreducible component* $\mathcal{W}$, that strictly contains contains the joint image. With our choice of camera matrices, this component is described by

$$\begin{aligned} \mathcal{W} = \{ & -z_3x_4 - z_3y_4 + x_3z_4 + y_3z_4 = 0, \\ & -z_2x_4 - z_2y_4 + x_2z_4 + y_2z_4 = 0, \\ & -z_2x_3 - z_2y_3 + x_2z_3 + y_2z_3 = 0, \\ & -z_1x_4 - z_1y_4 + x_1z_4 + y_1z_4 = 0, \\ & -z_1x_3 - z_1y_3 + x_1z_3 + y_1z_3 = 0, \\ & -z_1x_2 - z_1y_2 + x_1z_2 + y_1z_2 = 0\}. \end{aligned} \tag{1.53}$$

Note that these expressions *do not depend on the parameter* t, which shows that camera matrices cannot be determined by the epipolar conditions. We conclude that in this case the bilinearities do not give a weak characterization of the joint image.

This analysis proves Proposition 1.3.19 for $n = 4$, since we have exhausted all possible configurations of four pinholes in $\mathbb{P}^3$. The case $n > 4$ now follows easily. Indeed, bilinear and trilinear are always strongly sufficient because we have shown this to be true for all sets of four cameras (and this is enough thanks to Proposition 1.3.9). If the pinholes are non-coplanar, we may assume that cameras $(1, 2, 3)$ have non-collinear pinholes and span a plane that does not contain any other pinhole, so applying the previous argument for all quadruplets of cameras $(1, 2, 3, i)$ we can conclude that bilinearities give a strong characterization of the joint image. If the pinholes are only non-collinear, we can assume that pinholes $(1, 2)$ span a line that doesn't contain other pinholes, and similarly use this to conclude that bilinearities give a weak characterization of the joint image. Finally, whenever the pinholes are all aligned, camera geometry cannot be uniquely determined from bilinearities: for example, the expressions in (1.53) show that, using appropriate image coordinates, *all* pairs of cameras with pinholes on a given line yield the same epipolar constraint. More generally, it is easy to realize that if $P_1, \ldots, P_n$ have collinear pinholes, then any set of camera matrices $P'_1, \ldots, P'_n$, where P'_i is obtained from P_i by adding to its columns multiples of the coordinate vector for the epipole, will yield the same set of epipolar constraints. □

1.7.2 Viewing graphs

We next prove Proposition 1.4.16 from Section 1.4.6. We first need the following result, that shows that the conditions (1.27) are in fact linear constraints on $\mathbf{h}_\lambda$ and $\mathbf{h}_{\lambda'}$.

Proposition 1.7.6. *If $\mathbf{c} = (c_0, c_1, c_2, c_3)$ is a non-zero vector, then a matrix $\mathbf{M} = (m_{ij})_{i,j=0,\dots,3}$ can be written in the form $\mathbf{M} = \alpha\mathbf{I}_4 + \mathbf{c}\mathbf{v}^T$ for some arbitrary α and $\mathbf{v}$ if and only if the following linear expressions vanish:*

$$\begin{gathered}
m_{31}c_2 - m_{21}c_3 \\
m_{30}c_2 - m_{20}c_3 \\
m_{32}c_1 - m_{12}c_3 \\
m_{30}c_1 - m_{10}c_3 \\
m_{23}c_1 - m_{13}c_2 \\
m_{20}c_1 - m_{10}c_2 \\
m_{32}c_0 - m_{02}c_3 \\
m_{31}c_0 - m_{01}c_3 \\
m_{23}c_0 - m_{03}c_2 \\
m_{21}c_0 - m_{01}c_2 \\
m_{13}c_0 - m_{03}c_1 \\
m_{12}c_0 - m_{02}c_1 \\
m_{22}c_1 - m_{33}c_1 - m_{12}c_2 + m_{13}c_3 \\
m_{21}c_1 - m_{11}c_2 + m_{33}c_2 - m_{23}c_3 \\
m_{30}c_0 - m_{32}c_2 - m_{00}c_3 + m_{22}c_3 \\
m_{22}c_0 - m_{33}c_0 - m_{02}c_2 + m_{03}c_3 \\
m_{20}c_0 - m_{00}c_2 + m_{33}c_2 - m_{23}c_3 \\
m_{31}c_1 - m_{32}c_2 - m_{11}c_3 + m_{22}c_3 \\
m_{11}c_0 - m_{33}c_0 - m_{01}c_1 + m_{03}c_3 \\
m_{10}c_0 - m_{00}c_1 + m_{33}c_1 - m_{13}c_3.
\end{gathered} \tag{1.54}$$

Proof. The result is easily shown using a computer algebra system. Inside the ring $\mathbb{Q}[m_{00}, \dots, m_{33}, c_0, \dots, c_3, v_0, \dots, v_3, \alpha]$, we consider the ideal I obtained by eliminating the variables v_0, v_1, v_2, v_3 and α from the coordinates of $\mathbf{M} - \alpha\mathbf{I}_4 + \mathbf{c}\mathbf{v}^T$. We can then verify that (1.54) generate an ideal that decomposes into two prime components: one of these is irrelevant for us (it describes the vanishing of $\mathbf{c}$) and the other one is I. □

Proposition 1.4.16. *The tangent space of $\mathcal{T}_G(c_1, \dots, c_n)$ at the identity can be represented as the space of e-tuple of matrices $(\mathbf{h}_\lambda, \lambda \in E_G)$ where each $\mathbf{h}_\lambda$ is in $\mathbb{R}^{4\times 4}$ (not necessary invertible), and with compatibility conditions of the form*

$$\mathbf{h}_\lambda - \mathbf{h}_{\lambda'} = \alpha\mathbf{I}_4 + \mathbf{c}_i\mathbf{v}^T, \tag{1.27}$$

where $\alpha \in \mathbb{R} \setminus \{0\}$ and $\mathbf{v} \in \mathbb{R}^4$ are arbitrary, and λ and λ' share the vertex i.

Proof. According to Proposition 1.7.6, a matrix $\mathbf{M}$ can be written in the form $\mathbf{M} = \alpha\mathbf{I}_4 + \mathbf{c}\mathbf{v}^T$ for some $\alpha \in \mathbb{R}\setminus\{0\}$ and $\mathbf{v} \in \mathbb{R}^4$ if and only if it satisfies a set linear equations that depend on $\mathbf{c}$. Let us write $L_{\mathbf{c}}(\mathbf{M}) = \mathbf{0}$ for these linear conditions (so $L_{\mathbf{c}}$ is a linear map). Note that necessarily $L_{\mathbf{c}}(\mathbf{I}_4) = \mathbf{0}$.

A constraint of the form (1.26) can now be expressed as $F(\mathbf{g}_\lambda, \mathbf{g}_{\lambda'}) = L_{\mathbf{c}}(\mathbf{g}_\lambda \mathbf{g}_{\lambda'}^{-1}) = \mathbf{0}$. Writing the first order expansion or $F(\mathbf{g}_\lambda, \mathbf{g}_{\lambda'}) = \mathbf{0}$ at $(\mathbf{I}_4, \mathbf{I}_4)$ we obtain ($\approx$ denotes equality up to higher order terms)

$$F(\mathbf{I}_4 + \mathbf{h}_\lambda, \mathbf{I}_4 + \mathbf{h}_{\lambda'}) \approx L_{\mathbf{c}}((\mathbf{I}_4 + \mathbf{h}_\lambda)(\mathbf{I}_4 - \mathbf{h}_{\lambda'})) \approx L_{\mathbf{c}}(\mathbf{h}_\lambda - \mathbf{h}_{\lambda'}). \tag{1.55}$$

This shows that $L_{\mathbf{c}}(\mathbf{h}_\lambda - \mathbf{h}_{\lambda'}) = \mathbf{0}$ is the tangent space at the identity of each constraint (1.26). By Proposition 1.7.6 we have that $L_{\mathbf{c}}(\mathbf{h}_\lambda - \mathbf{h}_{\lambda'}) = \mathbf{0}$ is equivalent to (1.27), and this concludes the proof. □

By substituting $\mathbf{M} = \mathbf{h}_\lambda - \mathbf{h}_{\lambda'}$ inside (1.54), we obtain explicit equations that can be used to determine finite solvability (as explained in Section 1.4.6).

We now discuss the computations required to analyze the viewing graph in Example 1.4.19 from Section 1.4.7. The following simpler example serves as a useful preparation.

Example 1.7.7. Let G be the four-cycle shown in Figure 1.19. This graph is clearly not solvable, because a solvable graph with four vertices must have at least $e(4) = 5$ edges. It is however useful to understand this example algebraically. Following the general approach described in Section 1.4.6, we assign the identity $\mathbf{I}_4$ to the edge $(1, 2)$, and unknown matrices $\mathbf{g}_{(2,3)}$, $\mathbf{g}_{(3,4)}$, $\mathbf{g}_{(1,4)}$ to the remaining edges. The compatibility equations yield

$$\begin{aligned} \mathbf{g}_{(1,4)} &= \alpha_1\mathbf{I}_4 + \mathbf{c}_1\mathbf{v}_1^T \\ \mathbf{g}_{(2,3)} &= \alpha_2\mathbf{I}_4 + \mathbf{c}_2\mathbf{v}_2^T \\ \mathbf{g}_{(3,4)}\mathbf{g}_{(2,3)}^{-1} &= \alpha_3\mathbf{I}_4 + \mathbf{c}_3\mathbf{v}_3^T \\ \mathbf{g}_{(3,4)}\mathbf{g}_{(1,4)}^{-1} &= \alpha_4\mathbf{I}_4 + \mathbf{c}_4\mathbf{v}_4^T \end{aligned} \tag{1.56}$$

which imply that

$$\mathbf{g}_{(3,4)} = (\alpha_3\mathbf{I}_4 + \mathbf{c}_3\mathbf{v}_3^T)(\alpha_2\mathbf{I}_4 + \mathbf{c}_2\mathbf{v}_2^T) = (\alpha_4\mathbf{I}_4 + \mathbf{c}_4\mathbf{v}_4^T)(\alpha_1\mathbf{I}_4 + \mathbf{c}_1\mathbf{v}_1^T). \tag{1.57}$$

Expanding (1.57) we obtain

$$(\alpha_3\alpha_2 - \alpha_4\alpha_1)\mathbf{I}_4 = \mathbf{c}_1\mathbf{w}_1^T + \mathbf{c}_2\mathbf{w}_2^T + \mathbf{c}_3\mathbf{w}_3^T + \mathbf{c}_4\mathbf{w}_4^T, \tag{1.58}$$

where

$$\mathbf{w}_1 = \alpha_4\mathbf{v}_1, \quad \mathbf{w}_2 = -\alpha_3\mathbf{v}_2, \quad \mathbf{w}_3 = -(\alpha_2\mathbf{I}_4 + \mathbf{v}_2\mathbf{c}_2^T)\mathbf{v}_3, \quad \mathbf{w}_4 = (\alpha_1\mathbf{I}_4 + \mathbf{v}_1\mathbf{c}_1^T)\mathbf{v}_4. \tag{1.59}$$

From (1.58) we see that the vectors $\mathbf{w}_i$ must be scalar multiples of the rows of the matrix $\mathbf{C}^{-1}$ where $\mathbf{C}$ has columns $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4$. This easily implies that for fixed general coefficients $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, there is a unique solution for $\mathbf{v}_1, \ldots, \mathbf{v}_4$ in (1.56). Moreover, since the matrices $\mathbf{g}_{(i,j)}$ represent projective transformations, we can rescale these equations so that for example $\alpha_1 = \alpha_2 = \alpha_3 = 1$. This shows that there is one degree of projective freedom corresponding to the choice of α_4. This projective freedom can be explained in other ways: for example, our analysis implies that $\mathbf{g}_{(1,4)}$ can be any matrix of the form $\mathbf{g}_{(1,4)} = \alpha_1 \mathbf{I}_4 + \mathbf{c}_1 \mathbf{v}_1^T$ where $\mathbf{v}_1 \cdot \mathbf{c}_2 = \mathbf{v}_1 \cdot \mathbf{c}_3 = \mathbf{v}_1 \cdot \mathbf{c}_4 = 0$. Not counting the scale factor, this gives one degree of freedom, and fixing $\mathbf{g}_{(1,4)}$ of this type determines all other matrices up to scale. ♢

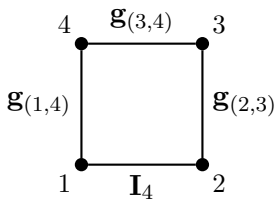


Figure 1.19: The viewing graph from Example 1.7.7.

Example 1.7.8. Let G be the graph with 9 vertices shown in Figure 1.20. According to Example 1.4.19 in Section 1.4.7, this graph is solvable. We can argue this fact using the analysis of the four-cycle from the previous example. Indeed, if we assign the matrices $\mathbf{I}_4, \mathbf{g}_{(4,5)}, \mathbf{g}_{(5,6)}, \mathbf{g}_{(5,8)}$ as shown in the figure, then up to recaling we have that $\mathbf{g}_{(4,5)} = \mathbf{I}_4 + \mathbf{c}_5 \mathbf{v}_4^T$ and $\mathbf{g}_{(5,6)}^{-1} = \mathbf{I}_4 + \mathbf{c}_5 \mathbf{v}_6^T$, where $\mathbf{v}_4$ and $\mathbf{v}_6$ are each determined up to a scalar multiple (more precisely, $\mathbf{v}_4$ must satisfy $\mathbf{c}_1 \cdot \mathbf{v}_4 = \mathbf{c}_2 \cdot \mathbf{v}_4 = \mathbf{c}_4 \cdot \mathbf{v}_4 = 0$ while $\mathbf{v}_6$ must satisfy $\mathbf{c}_2 \cdot \mathbf{v}_6 = \mathbf{c}_3 \cdot \mathbf{v}_6 = \mathbf{c}_6 \cdot \mathbf{v}_6 = 0$).[9] Moreover, we can write

$$\begin{aligned} \mathbf{g}_{(4,5)} \mathbf{g}_{(5,8)}^{-1} &= \mathbf{I}_4 + \mathbf{c}_5 \mathbf{v}_8^T \\ \mathbf{g}_{(5,8)} \mathbf{g}_{(5,6)}^{-1} &= \alpha \mathbf{I}_4 + \mathbf{c}_5 \mathbf{v}_8'^T \end{aligned} \tag{1.60}$$

where $\mathbf{c}_4 \cdot \mathbf{v}_8 = \mathbf{c}_7 \cdot \mathbf{v}_8 = \mathbf{c}_8 \cdot \mathbf{v}_8 = 0$ and $\mathbf{c}_6 \cdot \mathbf{v}_8' = \mathbf{c}_8 \cdot \mathbf{v}_8' = \mathbf{c}_9 \cdot \mathbf{v}_8' = 0$. Multiplying these two expressions together we obtain

$$\mathbf{g}_{(4,5)} \mathbf{g}_{(5,6)}^{-1} = (\mathbf{I}_4 + \mathbf{c}_5 \mathbf{v}_8^T)(\alpha \mathbf{I}_4 + \mathbf{c}_5 \mathbf{v}_8'^T) = (\mathbf{I}_4 + \mathbf{c}_5 \mathbf{v}_4^T)(\mathbf{I}_4 + \mathbf{c}_5 \mathbf{v}_6^T), \tag{1.61}$$

which yields

$$\alpha \mathbf{I}_4 + \mathbf{c}_5 (\mathbf{v}_8' + (\alpha + \mathbf{c}_5^T \mathbf{v}_8') \mathbf{v}_8)^T = \mathbf{I}_4 + \mathbf{c}_5 (\mathbf{v}_6 + (1 + \mathbf{c}_5^T \mathbf{v}_6) \mathbf{v}_4). \tag{1.62}$$

This relation can only be satisfied if $\alpha = 1$ and

$$\mathbf{v}_8' + (1 + \mathbf{c}_5^T \mathbf{v}_8') \mathbf{v}_8 = \mathbf{v}_6 + (1 + \mathbf{c}_5^T \mathbf{v}_6) \mathbf{v}_4. \tag{1.63}$$

[9] Note that the inverse of a matrix of the form $\mathbf{I} + \mathbf{c}\mathbf{v}_1^T$ is given by $\mathbf{I} + \mathbf{c}\mathbf{v}_2^T$ where $\mathbf{v}_2 = -\frac{1}{1 + \mathbf{c}^T \mathbf{v}_1} \mathbf{v}_1$. In particular $\mathbf{v}_2$ is a scalar multiple of $\mathbf{v}_1$.

Because the orthogonality conditions that defined $\mathbf{v}_4, \mathbf{v}_6, \mathbf{v}_8, \mathbf{v}'_8$ are all independent (since the pinholes are generic), (1.63) can be satisfied only if these vectors are all zero. From this we deduce that all the matrices $\mathbf{g}_{(i,j)}$ must be the identity, and the graph G is solvable. ♢

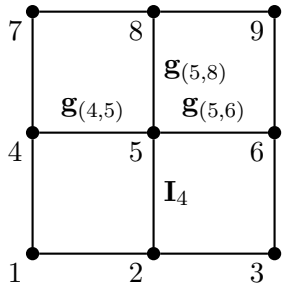


Figure 1.20: The viewing graph from Example 1.7.8.

1.7.3 Three-view geometry

We focus on the case of $n = 3$ views. We first point out that any three cameras with non-collinear pinholes can be transformed by homographies of $\mathbb{P}^3$ and of the image planes into

$$\mathbf{P}_1 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{P}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \mathbf{P}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}. \tag{1.64}$$

Similarly, three cameras with collinear pinholes can be transformed into

$$\mathbf{P}_1 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{P}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \mathbf{P}_3 = \begin{bmatrix} 1 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{1.65}$$

These camera matrices will be used for our explicit computations.

We now consider the trilinear constraints encoded in the trifocal tensor. We recall that these conditions can be deduced from the maximal minors of the matrix

$$\mathcal{U}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) = \begin{bmatrix} \mathbf{P}_1 & \mathbf{u}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{P}_2 & \mathbf{0} & \mathbf{u}_2 & \mathbf{0} \\ \mathbf{P}_3 & \mathbf{0} & \mathbf{0} & \mathbf{u}_3 \end{bmatrix}. \tag{1.66}$$

More precisely, the nine trilinearities in the tensor that distinguishes the first view are given by

$$t_{ij},\ i \in \{4,5,6\},\ j \in \{7,8,9\} \tag{1.67}$$

where t_{ij} is obtained by considering all rows of $\mathcal{U}$ excluding i and j. These are the same as (1.33).

Proposition 1.5.7. *For three cameras P_1, P_2, P_3 with non-collinear pinholes, the constraints (1.33) describe a set $\mathcal{W}$ which decomposes as $\mathcal{W} = M_3 \cup \mathcal{S}_{12} \cup \mathcal{S}_{13}$, where*

$$\begin{aligned} \mathcal{S}_{12} &= \{e_{12}\} \times \{e_{21}\} \times \mathbb{P}^2 \subset (\mathbb{P}^2)^3, \\ \mathcal{S}_{13} &= \{e_{13}\} \times \mathbb{P}^2 \times \{e_{31}\} \subset (\mathbb{P}^2)^3, \end{aligned} \tag{1.34}$$

and e_{ij} is the epipole in image i relative to the camera j.

Proof. Considering the camera matrices (1.64), one can verify using primary decomposition that the nine relations describe a space $\mathcal{W}$ that decomposes as $\mathcal{W} = M_3 \cup \mathcal{S}_{12} \cup \mathcal{S}_{13}$, where

$$\begin{aligned} \mathcal{S}_{12} = \{&x_1 + y_1 = 0; \\ &x_2 + y_2 = 0; \\ &z_1 = 0; \\ &z_2 = 0\} \\ \mathcal{S}_2 = \{&x_1 + y_1 = 0; \\ &x_3 + y_3 = 0; \\ &z_1 = 0; \\ &z_3 = 0\}. \end{aligned} \tag{1.68}$$

For our choice of cameras, these sets are exactly the ones described by (1.67), which in turn are a characterization that is independent of the choice of reference frames. □

By using the collinear cameras of (1.65) one can also show the following

Proposition 1.7.9. *If P_1, P_2, P_3 have collinear pinholes, the constraints (1.67) strongly characterize the joint image M_3.*

Let us now clarify a subtle point relative to this last result: from Proposition 1.3.11, we know that 9 trilinearities can never generate all of the trilinear relations which vanish on M_3 (since this is always a vector space of dimension 10). However, this does not exclude the possibility for these to *set-theoretically* define M_3: in other words, the zero-set of some relations can still describe M_3, even though they do not algebraically generate the whole space of the trilinearities. This is the case for the nine trilinearities (1.67) for cameras with collinear pinholes. For example, one can verify that for the matrices (1.65) the trilinearity

$$t_{26} = -z_1x_2x_3 - 2z_1y_2x_3 + z_1x_2y_3 + x_1x_2z_3 + x_1y_2z_3 \tag{1.69}$$

does not belong to the vector space generated by (1.67), although its zero-set clearly contains that of these nine trilinearities. To motivate this fact, one can show that $(t_{26})^2$

does belong to the ideal generated by (1.67) (*i.e.*, $(t_{26})^2$ can be written as an algebraic combination of the nine trilinearities) even if t_{26} doesn't: this clearly implies that adding the constraint t_{26} to (1.67) does not impose any additional conditions on the zero-set (since $t_{26} = 0$ if and only if $t_{26}^2 = 0$).

Proposition 1.5.1. *Consider $n = 3$ cameras. We write b_{ij} for the bilinear (epipolar) constraint among cameras i and j.*

- *If the pinholes are non-collinear:*
  1. *For* any *trilinearity t that does not vanish on the product of the trifocal lines, the constraints $\{b_{12}, b_{13}, b_{23}, t\}$ provide a strong characterization of the joint image.*
  2. *The epipolar constraints $\{b_{12}, b_{13}, b_{23}\}$ uniquely determine a camera configuration,* i.e.*, they provide a weak characterization of the joint image.*
- *If the cameras have collinear pinholes:*
  1. *A strong characterization of the joint image is given by $\{b_{12}, b_{13}, b_{23}, t_1, t_2\}$ where t_1 and t_2 are (sufficiently general) trilinear constraints.*
  2. *Two epipolar constraints together with one (sufficiently general) trilinearity $\{b_{12}, b_{13}, t\}$ uniquely determine camera geometry,* i.e.*, they give a weak characterization of the joint image.*

Proof. Let us first assume that P_1, P_2, P_3 have non-collinear pinholes.

1. The epipolar constraints describe the same projective locus as nine trilinearities (namely, $t_{12}, t_{13}, t_{23}, t_{45}, t_{46}$ *etc.*), that are obtained by multiplying bilinearities with a variable associated to the excluded image. In the case of non-collinear pinholes, one can verify that these constraints span a vector space of dimension 9. Any trilinearity that does not vanish on the product of the trifocal lines cannot belong to this space, so the three bilinearities together with such a trilinearity must describe the same zero-set as the whole space of trilinear constraints, *i.e.*, exactly M_3. See also [188].
2. This is well known: see [78, 84], or the discussion in Example 1.3.18.

We now assume that P_1, P_2, P_3 have collinear pinholes.

1. Consider constraints $\{b_{12}, b_{13}, b_{23}, t_{ij}\}$, where t_{ij} is chosen from (1.67). In *generic conditions* these describe a set $\mathcal{W}$ that decomposes as $\mathcal{W} = M_3 \cup \mathcal{V}_1 \cup \mathcal{V}_2$, where $\mathcal{V}_1$ and $\mathcal{V}_2$ are each products of three corresponding (epipolar) lines in each image, respectively passing through the $(i-2)$-th coordinate points in the second image and through the $(j-5)$-th coordinate point in the third (*e.g.*, if we consider T_{48} then $\mathcal{V}_1$ is the product of the three epipolar lines containing $[1, 0, 0] \in \mathbb{P}^2$ in the second image and $\mathcal{V}_2$ is the product of the three epipolar lines containing

$[0, 1, 0] \in \mathbb{P}^2$ in the third). Moreover, in general, if another trilinearity T_{kl} among (1.67) is such that $k \neq i$ and $l \neq j$, then $\{b_{12}, b_{13}, b_{23}, t_{ij}, t_{kl}\}$ are strongly sufficient, since the spurious sets associated with each trilinearity are excluded by the other constraint. The "generic" assumption that was used in this argument is that the triples of epipolar lines associated with the fundamental points are all distinct: in other words, we assumed that 1) no fundamental point in each image corresponds to any other fundamental point in another 2) no pair of fundamental points in a given image lies on the same epipolar line[10]. However, even if these conditions are not satisfied, there will *always* be at least two trilinearities among (1.67) that can be used together with the epipolar constraints to characterize M_3: this follows from the fact that in each image one can always choose two fundamental points that do not lie on the same epipolar line.

2. Let t be any trilinearity whose zero-set does not contain that of $\{b_{12}, b_{13}\}$: then we may be sure that $\mathcal{W} = \{b_{12} = b_{13} = t = 0\}$ is a set of dimension 3, which must contain M_3 as a component of maximal dimension. By removing spurious components we can recover a strong characterization of the joint image, and thus camera geometry can be determined.

□

[10] We point out that the matrices (1.65) *do not* satisfy these assumptions.

Chapter 2

Geometric Cameras

In the previous chapter, we modeled a pinhole camera as a linear map $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$, described by a 3×4 matrix up to scale. We now adopt a more geometric viewpoint, and represent a camera as a mapping from $\mathbb{P}^3$ to a two-dimensional family of lines, or a *line congruence*. This model ignores image planes and measurements, and focuses on the intrinsic geometry of the imaging process. This setting also allows us to study very general cameras in a unified manner: our model serves in fact as a natural abstraction of many existing devices, including pushbroom, panoramic and catadioptric cameras.

The material in this chapter is based on the following publication:

– Jean Ponce, Bernd Sturmfels, and Matthew Trager. "Congruences and concurrent lines in multi-view geometry". In: *Advances in Applied Mathematics* 88 (2017), pp. 62–91.

2.1 Introduction

Despite the practical usefulness of image coordinates for representing pixels in photographs, the geometry of vision is really a matter of lines in space. As pointed out for example by Grossberg and Nayar [71], the intensity value of a pixel in a photograph does not depend on the physical location of photosensitive elements on the retinal plane, but only on the "primary" 3D viewing ray associated with that pixel. In particular, any imaging system can be seen abstractly as a family of viewing rays, and this viewpoint is especially useful for studying *non-central cameras* [11, 135, 136, 145]. The set of all viewing rays associated with a camera forms in fact a *line congruence*, *i.e.*, a surface inside $\mathrm{Gr}(1, \mathbb{P}^3)$ (because every pixel in the two-dimensional image is associated with one viewing line). The representation of cameras as congruences has been used explicitly in [145, 11]. On the other hand, these works have focused mainly on "linear congruences", *i.e.*,

congruences that are characterized by linear equations in Plücker coordinates. In this chapter, we study more general congruences that can be used to model cameras.

Our definition of a *geometric camera* is a map $\mathbb{P}^3 \dashrightarrow \mathrm{Gr}(1,\mathbb{P}^3)$ from 3-space into the Grassmannian of lines, with the property that the (closure of) the pre-image of a point (line) in the Grassmannian is that same line in $\mathbb{P}^3$. We will also always assume that the coordinates of the map from points to lines are algebraic functions. As mentioned above, a geometric camera is always associated with a congruence of lines [92], that is simply the image of the camera in $\mathrm{Gr}(1,\mathbb{P}^3)$. An important subclass of these cameras are those associated with congruences of *order one*: this means that a generic point in $\mathbb{P}^3$ is contained in exactly one line of the congruence. This type of congruence defines a *rational* geometric camera, where the map from points to image lines is given by rational functions. Congruences of order one have been classified in 1866 by Kummer [105]. Using this classification, we can give a concrete representation of *all* possible geometric cameras that are described by rational functions. Some of these models represent existing devices (*e.g.*, 'two-slit" [55, 135], "pushbroom" [72], "oblique" [136] cameras), but many others do not (yet) have physical realizations.

Our framework is also useful for studying the multi-view geometry of general camera models in a unified manner. Taking pictures with n rational cameras for congruences $C_1,\ldots,C_n$ defines a rational map

$$\phi : \mathbb{P}^3 \dashrightarrow C_1 \times C_2 \times \cdots \times C_n \subset (\mathrm{Gr}(1,\mathbb{P}^3))^n \subset (\mathbb{P}^5)^n. \tag{2.1}$$

Here the congruence C_i plays the role of the i-th image plane $\mathbb{P}^2$ in classical multi-view geometry. The closure of the image of ϕ is a projective variety $M(C_1,C_2,\ldots,C_n)$ in $(\mathbb{P}^5)^n$, that we call the *multi-image variety*. This is analogous to the joint image variety in $(\mathbb{P}^2)^n$ that was defined for pinhole cameras in Section 1.3. In order to characterize $M(C_1,C_2,\ldots,C_n)$ algebraically, we study the set V_n in $(\mathrm{Gr}(1,\mathbb{P}^3))^n$ of all n-tuples of lines that are concurrent at a point. Under suitable genericity assumptions, the multi-image variety equals the intersection

$$(C_1 \times C_2 \times \cdots \times C_n) \cap V_n \qquad \text{in} \quad \mathrm{Gr}(1,\mathbb{P}^3)^n \subset (\mathbb{P}^5)^n. \tag{2.2}$$

In particular, for any set of congruences $C_1,\ldots,C_n$, we can recover "multi-view constraints" that describe concurrent visual rays by simply adding the equations defining the $C_1,\ldots,C_n$ to those defining V_n.

2.1.1 Previous work

Congruences of lines were first studied by classical geometers of the nineteenth century [105]. They have however remained a popular topic in more modern mathematical works [44, 45, 103]. Many of these papers are interested in differential properties of algebraic varieties (*e.g.*, "focal loci" [7]).

The relationship between classical line congruences and non-central cameras was pointed out by Ponce [145], who noted that many existing devices, including two-slit [55, 135], pushbroom [72] and oblique [136] cameras, could be modeled using a *linear* congruence. Batog, Goaoc, and Ponce took this further in [11], where they classified all cameras associated with linear congruences, describing them as maps of the form $x \mapsto x \vee Ax$ where A is an appropriate linear transformation of $\mathbb{P}^3$ (they show that A must have minimal polynomial of degree 2). This representation was previously introduced by Pajdla for the case of oblique cameras [136].

The idea of using line geometry and Plücker coordinates for deriving multi-view constraints is not new and, indeed, it is quite common to recover the epipolar relation between cameras by imposing the concurrence of two visual rays (see, *e.g.*, [143]). In [188] we used a similar approach to study the case of $n = 3$ lines. On the other hand, to our knowledge a complete algebraic characterization of the concurrence conditions for general n-tuples of lines has not been proposed or used. In this sense, an interesting related work is by Sturm [172], who deduced multi-view constraints for general camera systems and expressed them using Plücker coordinates of viewing lines. His derivation is very similar to the usual method for obtaining multi-view constraints for pinhole cameras (as presented in Section 1.3). However, Sturm's approach actually provides conditions on the (euclidean) *camera motions*, rather than on the cameras or congruences themselves. Thus, his equations are not well suited for describing the multi-view geometry of a system of cameras (indeed, the cameras are never actually modeled explicitly), although they can be used to recover camera poses from correspondences.

More works on non-central imaging models will be discussed in the next chapter.

2.1.2 Main contributions

Our main contributions can be summarized as follows.

- We study general "geometric cameras" that are associated with congruences of lines in $\mathrm{Gr}(1, \mathbb{P}^3)$. We use Kummer's classification of congruences of order one [105] to characterize all rational geometric cameras (Section 2.3.2).
- We study the *concurrent lines variety* V_n, that is the $(2n+3)$-dimensional variety of n-tuples of concurrent lines in $\mathbb{P}^3$. We provide a complete description of the ideal generators associated with V_n (Theorem 2.4.1).
- We introduce the *multi-image variety* $M(C_1, \ldots, C_n)$ for n rational cameras (or congruences) $C_1, \ldots, C_n$. This is analogous to the joint image variety (see Section 1.3), and indeed if each C_i is a pinhole camera then the two varieties are isomorphic (Proposition 2.4.5). Assuming that cameras are in general position, an algebraic description of $M(C_1, \ldots, C_n)$ is easily obtained from the generators of the ideal of V_n (Theorem 2.4.3).

- We study some geometric cameras of order greater than one, for which point-to-line map is algebraic but not rational. In particular, we describe the algebraic geometry of "panoramic" and "catadioptric" cameras (Section 2.5).

Before our main discussion, we briefly review in Section 2.2 some fundamental notions on the geometry of line congruences.

Conventions: Throughout the chapter, we write $\mathbb{P}^n = \mathbb{P}(\mathbb{C}^{n+1})$ for the n-dimensional projective space over the *complex* numbers. This setting is most convenient for dealing with nonlinear algebraic varieties and the corresponding ideals (see Appendix B). On the other hand, our varieties will always be defined by polynomials that have coefficients in the field $\mathbb{R}$ of real numbers, and we will be mostly interested in the real locus of these varieties.

Compared to the previous chapter, the distinction between points in $\mathbb{P}^n$ and their vector representatives will not be as important, so to avoid unnecessary clutter will use regular font to indicate both. For example, if $x = [\mathbf{x}]$ is a point in $\mathbb{P}^3$ and $A = [\mathbf{A}]$ represents a projective transformation, then we may write Ax in place of $\mathbf{Ax}$. This should not cause any confusion.

2.2 Preliminaries

This section reviews some basics on line geometry and congruences that will be central for our discussion. For more details on Plücker coordinates and the Grassmannian $\mathrm{Gr}(1, \mathbb{P}^3)$ we refer the reader to [150] or to the Appendix A.

2.2.1 The Grassmannian of lines

The set of all lines of $\mathbb{P}^3$ forms a four-dimensional variety, the *Grassmannian* of lines, denoted by $\mathrm{Gr}(1, \mathbb{P}^3)$. The line passing through two distinct points x and y in $\mathbb{P}^3$ can be represented using *Plücker coordinates* $p = [p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}]$ where $p_{ij} = x_i y_j - x_j y_i$. This defines a point in $\mathbb{P}^5$ that is independent of the choice of x and y along the line. Plücker coordinates of lines satisfy the quadratic *Plücker relation*

$$p_{12}p_{34} - p_{13}p_{24} + p_{14}p_{23} = 0. \tag{2.3}$$

Hence, $\mathrm{Gr}(1, \mathbb{P}^3)$ can be identified with a quadric hypersurface in $\mathbb{P}^5$. We can also represent a line in $\mathbb{P}^3$ by its *dual Plücker coordinates* $[q_{12}, q_{13}, q_{14}, q_{23}, q_{24}, q_{34}]$. These are the 2×2-minors of a 2×4-matrix whose kernel is the line. Primal and dual Plücker coordinates are related by $q_{ij} = \sigma_{(ijkl)} p_{kl}$, where i, j, k, l are distinct indices, and $\sigma_{(ijkl)}$ is the sign of the permutation $(ijkl)$.

To express incidences of lines with points and planes, it is convenient to write the Plücker coordinates of a line $p = [p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}]$ and its dual as the entries of two skew-symmetric 4×4-matrices:

$$P = \begin{bmatrix} 0 & p_{34} & -p_{24} & p_{23} \\ -p_{34} & 0 & p_{14} & -p_{13} \\ p_{24} & -p_{14} & 0 & p_{12} \\ -p_{23} & p_{13} & -p_{12} & 0 \end{bmatrix} \quad \text{and} \quad P^* = \begin{bmatrix} 0 & p_{12} & p_{13} & p_{14} \\ -p_{12} & 0 & p_{23} & p_{24} \\ -p_{13} & -p_{23} & 0 & p_{34} \\ -p_{14} & -p_{24} & -p_{34} & 0 \end{bmatrix}. \tag{2.4}$$

With these definitions, the conditions $\operatorname{rank}(P) = 2$, $\operatorname{rank}(P^*) = 2$, and $\operatorname{trace}(PP^*) = 0$ are all equivalent to the Plücker relation (2.3). Moreover, the join of a line p and a point x in $\mathbb{P}^3$ is the plane with dual coordinates Px. Similarly, the meet of p with a plane u in $(\mathbb{P}^3)^*$ is the point with coordinates Pu. Two lines p and q are concurrent if and only if $\operatorname{trace}(PQ^*) = \operatorname{trace}(QP^*) = 0$.

2.2.2 Congruences of lines

A two-dimensional family of lines in $\mathbb{P}^3$ is a surface in $\operatorname{Gr}(1, \mathbb{P}^3)$. This is classically known as a *line congruence* [44, 45]. Throughout this chapter (and this thesis) we will only consider *algebraic* congruences, that can be defined by polynomial equations in Plücker coordinates. Any algebraic congruence C is associated with a *bidegree*, that is a pair of nonnegative integers (α, β). This is analogous to the notion degree for a projective variety. The integer α is known as the *order*, and is the number of lines in C that pass through a general point of $\mathbb{P}^3$. The integer β is the *class* and represents the number of lines in C that lie in a general plane of $\mathbb{P}^3$. The study of congruences was an active area of research in the second half of the 19-th century. Many results from that period can be found in the book by Jessop [92] on *line complexes*, the classical term for threefolds in $\operatorname{Gr}(1, \mathbb{P}^3)$.

Example 2.2.1. *(1,0) and (0,1)-Congruences.* A congruence C has bidegree $(1, 0)$ if and only if C is the set of lines through a point x in $\mathbb{P}^3$ (C is an *α-plane* in $\operatorname{Gr}(1, \mathbb{P}^3)$). Dually, a congruence C has bidegree $(0, 1)$ if and only if it is the set of lines on a plane H in $\mathbb{P}^3$ (C is an *β-plane* in $\operatorname{Gr}(1, \mathbb{P}^3)$). ♢

Given an (α, β)-congruence C, a point $x \in \mathbb{P}^3$ is a *focal point* if x does not belong to α distinct lines of C. This may happen if x belongs to fewer than α distinct lines, or if x belongs to an infinite number of lines. In the latter case, x is a *fundamental point*. The variety $\mathcal{F}(C)$ of focal points is the *focal locus*, while the variety $\mathcal{G}(C)$ of fundamental points is the *fundamental locus*. Clearly, $\mathcal{G}(C)$ is contained in $\mathcal{F}(C)$. Moreover, the focal locus $\mathcal{F}(C)$ is typically a surface in $\mathbb{P}^3$. It is known (cf. [45, Proposition 2]) that $\mathcal{F}(C)$ has lower dimension if and only if C has order at most one, in which case $\mathcal{F}(C) = \mathcal{G}(C)$.

Several natural congruences are derived from geometric objects in $\mathbb{P}^3$. Given a surface X in $\mathbb{P}^3$, the set of all lines that are tangent to X at two points forms the

bitangent congruence $\mathrm{Bit}(X)$, while the set of lines that are tangent X with order of contact three form the *inflectional congruence* $\mathrm{Infl}(X)$. For a curve Y in $\mathbb{P}^3$, the set of lines that intersect Y in two points form the *secant congruence* $\mathrm{Sec}(Y)$. We will return to these objects in Chapter 5 when we will study "aspects" of curves and surfaces in $\mathbb{P}^3$.

Remark 2.2.2. The focal locus $\mathcal{F}(C)$ of a congruence C can be recovered computationally as follows. Let I be the ideal in $\mathbb{C}[p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}]$ that defines C. Of course, $p_{14}p_{23} - p_{13}p_{24} + p_{12}p_{34} \in I$. The set of lines in C that pass through a point x in $\mathbb{P}^3$ is given by the ideal

$$I + \langle Px \rangle, \tag{2.5}$$

where P is the 4×4 Plücker matrix with indeterminate coefficients from (2.4). For a generic x in $\mathbb{P}^3$, the ideal (2.5) has exactly α complex zeros in $\mathbb{P}^5$. To compute the focal locus, we treat the coordinates of x as parameters, and we add to (2.5) the 5×5 minors of the Jacobian matrix of (2.5) with respect to the Plücker coordinates. This gives an ideal in $\mathbb{C}[p_{12}, \ldots, p_{34}, x_1, \ldots, x_4]$. By saturating and eliminating $p_{12}, \ldots, p_{34}$, we obtain the ideal in $\mathbb{C}[x_1, x_2, x_3, x_4]$ that defines the focal locus $\mathcal{F}(C)$ in $\mathbb{P}^3$.

2.3 Rational Cameras

A *geometric camera* is a map that associates points in $\mathbb{P}^3$ with "viewing lines" in $\mathrm{Gr}(1, \mathbb{P}^3)$. A natural requirement for this map is that (the closure of) the pre-image of a generic element in $\mathrm{Gr}(1, \mathbb{P}^3)$ is that same line in $\mathbb{P}^3$. We will also always assume that the coordinates of this mapping are *algebraic functions*, *i.e.*, functions that can be defined as roots of polynomial equations. This guarantees that the Zariski closure of the image of a geometric camera inside $\mathrm{Gr}(1, \mathbb{P}^3)$ is a line congruence defined by algebraic equations.

In this section, we focus on geometric cameras that are associated with line congruences of *order one*. We call these *rational cameras*, since the corresponding map $\mathbb{P}^3 \dashrightarrow \mathrm{Gr}(1, \mathbb{P}^3)$ is a rational map, *i.e.*, its coordinates are homogeneous polynomial functions. Rational cameras are also completely determined by the corresponding congruence. Indeed, if C is a congruence of bidegree $(1, \beta)$, we can define a map

$$\mathbb{P}^3 \dashrightarrow C \subset \mathrm{Gr}(1, 3) \tag{2.6}$$

that associates with a generic point x in $\mathbb{P}^3$ the *unique* line in C that passes through x. This map is defined everywhere except at the *focal locus* of C which, as was noted above, has dimension either one or zero, and coincides with the fundamental locus. We will often write $C(x)$ for the image of x under the map (2.6).

We begin our discussion by analyzing in Section 2.3.1 the two simplest examples of rational cameras, namely pinhole cameras and two-slit cameras. These are "linear"

congruences, obtained by intersecting the Plücker quadric with linear subspaces of $\mathbb{P}^5$. In Section 2.3.2, we provide a complete description of all possible rational cameras, based on a classical theorem of Kummer that classified all congruences of order one. In addition to general formulae, we present explicit examples that illustrate instances of these cameras.

2.3.1 Pinhole and two-slit cameras

We next describe geometric cameras associated with *pinhole* and *two-slit* cameras. These correspond to congruences of bidegree $(1,0)$ and $(1,1)$, respectively. We will consider rational cameras in full generality in the next subsection.

Pinhole cameras. If C is a $(1,0)$-congruence, then C is an α-plane for some point c in $\mathbb{P}^3$, and (2.6) represents a traditional pinhole camera. The image of a point x is the line with Plücker coordinates

$$C(x) \;=\; x \vee c \;=\; \begin{bmatrix} c_1x_2 - c_2x_1 \\ c_1x_3 - c_3x_1 \\ c_1x_4 - c_4x_1 \\ c_2x_3 - c_3x_2 \\ c_2x_4 - c_4x_2 \\ c_3x_4 - c_4x_3 \end{bmatrix} \;\in\; \mathrm{Gr}(1,\mathbb{P}^3). \tag{2.7}$$

Note that there is a complete symmetry between the center c and the projected point x, and if we write C_c and C_x for the α-planes of lines through c and x respectively, then $\{x \vee c\} = C_c \cap C_x$.

Two slit cameras. A congruence C defined by two general linear forms in Plücker coordinates is a $(1,1)$-congruence. To see this, we note that C is the intersection of $\mathrm{Gr}(1,\mathbb{P}^3)$ with a linear space L in $\mathbb{P}^5$ of dimension 3, and that the dual $L^\vee$ of this space is a line in $(\mathbb{P}^5)^*$ that intersects the (dual) Grassmannian in two points p^* and q^* (over the field $\mathbb{C}$ of complex numbers). These two points correspond to lines p and q in $\mathbb{P}^3$, and C is characterized geometrically as the family of common transversals to p and q. Each point of $\mathbb{P}^3$ outside these two lines lies on a unique such transversal; similarly, a generic plane in $\mathbb{P}^3$ contains exactly one transversal to p and q. The rational camera associated with C is a *two-slit camera*, and p and q are the two *slits*. These lines are also the focal locus for C.

As mentioned in Section 2.2, the plane containing a general point x in $\mathbb{P}^3$ and the line p is the point in $(\mathbb{P}^3)^*$ with coordinates $x \vee p = Px$. Likewise, $x \vee q = Qx$ is the plane spanned by the point x and the line q. Here P and Q are the skew-symmetric 4×4-matrices that

represent p and q. Intersecting these two planes gives the line in the congruence that contains x. In symbols,

$$C(x) \;=\; (x \vee p) \wedge (x \vee q) \;=\; Pxx^TQ - Qxx^TP. \tag{2.8}$$

The coordinates of the Plücker vector $C(x)$ are quadratic in the coordinates of x, and they are bilinear in (p, q). For instance, the first coordinate of $C(x)$, indexed by 12, is equal to

$$\begin{aligned}(q_{24}p_{34} - q_{34}p_{24})x_1x_3 + (q_{34}p_{23} - q_{23}p_{34})x_1x_4 + (q_{34}p_{14} - q_{14}p_{34})x_2x_3 + (q_{13}p_{34} - q_{34}p_{13})x_2x_4 \\ +(q_{14}p_{24} - q_{24}p_{14})x_3^2 + (q_{23}p_{14} - q_{14}p_{23} - q_{13}p_{24} + q_{24}p_{13})x_3x_4 + (q_{13}p_{23} - q_{23}p_{13})x_4^2.\end{aligned}$$

In summary, the picture of x taken with the two-slit camera C is the line given by (2.8).

Example 2.3.1 (Pushbroom cameras)**.** A *pushbroom camera* [72] is a device consisting of a linear array of sensors mounted on a platform that can move along a line perpendicular to the sensors (see Figure 2.1). As the platform moves, the camera scans a family of viewing planes. This type of optical system is commonly used in aerial and satellite cameras as well as CT systems.

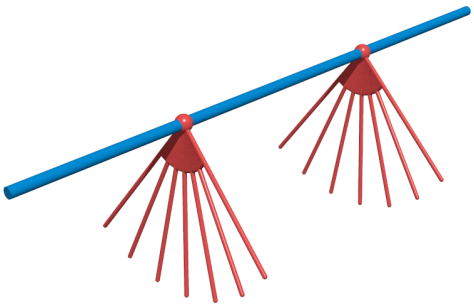


Figure 2.1: The geometry of a pushbroom camera.

It was observed in [145] that pushbroom cameras are two-slit cameras where one of the two slits lies on the plane at infinity. If we identify euclidean 3-space with the affine chart $U_1 = \{x_1 \neq 0\}$ then q can be any line of the form $q = [0, 0, 0, q_{23}, q_{24}, q_{34}]$. A standard choice is the line at infinity that is orthogonal to p, with respect to the usual scalar product on $U_1 \simeq \mathbb{R}^3$. That line has the Plücker coordinates $q = [0, 0, 0, p_{14}, -p_{13}, p_{12}]$. For this choice of q, the polynomial formula (2.8) for the image line $C(x)$ specializes to

$$\begin{bmatrix} -p_{13}p_{23}x_1^2 - p_{14}p_{24}x_1^2 + p_{13}^2x_1x_2 + p_{14}^2x_1x_2 - p_{12}p_{13}x_1x_3 - p_{12}p_{14}x_1x_4 \\ p_{12}p_{23}x_1^2 - p_{14}p_{34}x_1^2 - p_{12}p_{13}x_1x_2 + p_{12}^2x_1x_3 + p_{14}^2x_1x_3 - p_{13}p_{14}x_1x_4 \\ p_{12}p_{24}x_1^2 + p_{13}p_{34}x_1^2 - p_{12}p_{14}x_1x_2 - p_{13}p_{14}x_1x_3 + p_{12}^2x_1x_4 + p_{13}^2x_1x_4 \\ p_{12}p_{23}x_1x_2 - p_{14}p_{34}x_1x_2 - p_{12}p_{13}x_2^2 + p_{13}p_{23}x_1x_3 + p_{14}p_{24}x_1x_3 + p_{12}^2x_2x_3 - p_{13}^2x_2x_3 + p_{12}p_{13}x_3^2 - p_{13}p_{14}x_2x_4 + p_{12}p_{14}x_3x_4 \\ p_{12}p_{24}x_1x_2 + p_{13}p_{34}x_1x_2 - p_{12}p_{14}x_2^2 - p_{13}p_{14}x_2x_3 + p_{13}p_{23}x_1x_4 + p_{14}p_{24}x_1x_4 + p_{12}^2x_2x_4 - p_{14}^2x_2x_4 + p_{12}p_{13}x_3x_4 + p_{12}p_{14}x_4^2 \\ p_{12}p_{14}x_2x_3 - p_{12}p_{24}x_1x_3 - p_{13}p_{34}x_1x_3 + p_{13}p_{14}x_3^2 + p_{12}p_{23}x_1x_4 - p_{14}p_{34}x_1x_4 - p_{12}p_{13}x_2x_4 - p_{13}^2x_3x_4 + p_{14}^2x_3x_4 - p_{13}p_{14}x_4^2 \end{bmatrix}.$$

This Plücker vector represents the picture of the point x taken by the pushbroom camera. ♢

Remark 2.3.2. If C is $(1,1)$-congruence defined over the real numbers, then we can distinguish three possibilities for the focal locus: the two lines p and q in $\mathcal{F}(C)$ may be real and distinct, real and coincide (when the line $L^\vee$ intersects $\mathrm{Gr}(1,\mathbb{P}^3)$ in a double point), or they may form a complex conjugate pair of lines. In the first case, the $(1,1)$-congruence C is called *hyperbolic*. In the second case, C is said to be *parabolic*, and consists of a one-parameter family of flat pencils of lines centered on the line $p = q$. In the last case, the focal locus $\mathcal{F}_C$ has no real points, and the $(1,1)$-congruence C is said to be *elliptic*. The real geometry of linear cameras is discussed in more detail in [11].

2.3.2 Classification of rational cameras

In this section we discuss the classification $(1,\beta)$-congruences C for any β, and we derive some explicit formulas for the rational maps $x \mapsto C(x)$. Congruences of order one were first classified in 1866 by Kummer [105]. His result was then refined and extended by various authors in the 20th century. The following version was derived by De Poi in [44]. We refer to his article for more information.

Theorem 2.3.3. *Let C be a $(1,\beta)$-congruence with focal locus $\mathcal{F}(C)$. Then one of the following four situations is the case (see Figure 2.2):*

1. *$\mathcal{F}(C)$ is a point c, and C is the α-plane of lines through c. Here $\beta = 0$.*
2. *$\mathcal{F}(C)$ is a twisted cubic in $\mathbb{P}^3$, and C consists of its secant lines. Here $\beta = 3$.*
3. *$\mathcal{F}(C)$ is the union of a rational curve X of degree β and a line L that intersects X in $\beta - 1$ points. The congruence C is the family of lines that intersects both L and X. Here we allow for degenerate cases: the points in $X \cap L$ are counted with multiplicity.*
4. *$\mathcal{F}(C)$ is (a non-reduced) line L. The congruence C is described by a morphism ϕ of degree $\beta > 0$ from $L^\vee$ to L, where $L^\vee$ denotes the planes containing L: a line is in C if it belongs to a pencil of lines lying in a plane Π in $L^\vee$ and passing through $\phi(\Pi)$.*

We next describe the rational cameras (2.6) for each of these families of congruences.

Type 1: $\mathcal{F}(C)$ is a point. This is the pinhole camera described in Section 2.3.1.

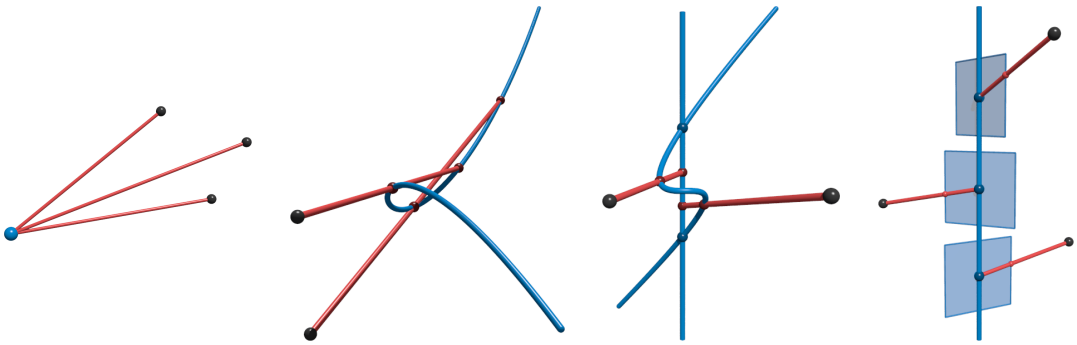


Figure 2.2: The four types of rational cameras associated with congruences of order one.

Type 2: $\mathcal{F}(C)$ is a twisted cubic. After a change of coordinates, the twisted cubic in $\mathbb{P}^3$ is the image of the map $[s,t] \mapsto [s^3, s^2t, st^2, t^3]$. The corresponding rational camera is

$$C(x) \;=\; \begin{bmatrix} (x_1x_3 - x_2^2)^2 \\ (x_1x_3 - x_2^2)(x_1x_4 - x_2x_3) \\ x_1x_3^3 + x_2^3x_4 - 3x_1x_3x_2x_4 + x_1^2x_4^2 \\ (x_2x_4 - x_3^2)(x_1x_3 - x_2^2) \\ (x_2x_4 - x_3^2)(x_1x_4 - x_2x_3) \\ (x_2x_4 - x_3^2)^2 \end{bmatrix}. \tag{2.9}$$

Indeed, the ideal of the congruence C (that is the secant congruence of the twisted curve) is generated by the Plücker relation together with five quadrics

$$p_{24}^2 - p_{14}p_{34} - p_{23}p_{34},\;\; p_{23}p_{24} - p_{13}p_{34},\;\; p_{23}^2 - p_{12}p_{34},\;\; p_{13}p_{23} - p_{12}p_{24},\;\; p_{13}^2 - p_{12}p_{14} - p_{12}p_{23}. \tag{2.10}$$

If we augment this ideal by the four entries of Px, where $x = [x_1, x_2, x_3, x_4]$ is an unknown world point in $\mathbb{P}^3$, then the radical of the resulting ideal is generated by the quadrics in (2.10) together with six bilinear equations that can be written in matrix-vector form as follows:

$$\begin{bmatrix} 0 & 0 & 0 & x_4 & -x_3 & x_2 \\ 0 & 0 & 0 & x_3 & -x_2 & x_1 \\ 0 & x_4 & -x_3 & 0 & 0 & x_1 \\ 0 & x_3 & -x_2 & -x_2 & x_1 & 0 \\ x_4 & 0 & -x_2 & 0 & x_1 & 0 \\ x_3 & -x_2 & 0 & x_1 & 0 & 0 \end{bmatrix} \begin{bmatrix} p_{12} \\ p_{13} \\ p_{14} \\ p_{23} \\ p_{24} \\ p_{34} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{2.11}$$

This 6×6 matrix has rank 5. The solution space of (2.11) is spanned by the vector in (2.9).

The *twisted cubic camera* (2.9) has a nice interpretation in terms of tensor decompositions. For this, we identify $\mathbb{P}^3$ with the space of symmetric $2 \times 2 \times 2$-tensors. We seek to decompose an arbitrary tensor as the sum of two rank 1 tensors. Equivalently, we seek to write a binary cubic $x_1u^3 + 3x_2u^2v + 3x_3uv^2 + x_4v^3$ as the sum of two cubes of linear

forms in u and v. Rank 1 tensors correspond to points on the twisted cubic curve. The desired representation is unique, and is given by the intersection points of the twisted cubic with the secant line $C(x)$.

Example 2.3.4. The binary cubic $3uv^2 + 3u^2v$ corresponds to the point $[0,1,1,0]$ in $\mathbb{P}^3$. Its image under (2.9) is the line $[1,1,0,1,1,1]$. This line intersects the twisted cubic at $[-2,-1+\sqrt{-3},1+\sqrt{-3},2]$ and $[-2,-1-\sqrt{-3},1-\sqrt{-3},2]$ in $\mathbb{P}^3$. From this we deduce

$$3uv^2+3u^2v \;=\; \frac{\sqrt{-3}}{3}\left(u+\left(1+\frac{\sqrt{-3}}{2}\right)v\right)^3 \;-\; \frac{\sqrt{-3}}{3}\left(u+\left(1-\frac{\sqrt{-3}}{2}\right)v\right)^3. \tag{2.12}$$

♢

Type 3: $\mathcal{F}(C)$ is a rational curve X and a line L. After a change of coordinates we may assume that the line is $L = \{[0,0,x_3,x_4] \in \mathbb{P}^3 : [x_3,x_4] \in \mathbb{P}^1\}$. The dual line $L^\vee$ parametrizes all planes in $\mathbb{P}^3$ that contain L. A natural parametrization $\mathbb{P}^1 \to L^\vee$ is given by identifying $[x_1,x_2]$ with the plane in $\mathbb{P}^3$ with dual coordinates $[x_2,-x_1,0,0]$.

To build our rational camera, we take an arbitrary rational curve X of degree β that intersects L in $\beta-1$ points. Each such curve X is given by a parametric representation

$$\mathbb{P}^1 \to X,\ [s,t] \mapsto [sf(s,t), tf(s,t), g(s,t), h(s,t)], \tag{2.13}$$

where f, g and h are arbitrary binary forms of degree $\beta-1$, β and β respectively.

Proposition 2.3.5. *For the rational camera of Type 3, the map (2.6) is given by*

$$C(x) \;=\; \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \vee \begin{bmatrix} x_1 f(x_1,x_2) \\ x_2 f(x_1,x_2) \\ g(x_1,x_2) \\ h(x_1,x_2) \end{bmatrix}. \tag{2.14}$$

Proof. The two column vectors in (2.14) represent two points in $\mathbb{P}^3$ that lie on the plane in $L^\vee$ with coordinates $[x_1,x_2]$, according to the parametrization above. We see from (2.13) that the second point lies on the curve X. Hence (2.14) is a line the intersects both L and X. □

Remark 2.3.6. The ideal of the congruence C can be described in terms of *Chow forms*. We recall that Chow form Ch_Z of an irreducible curve Z of degree γ in $\mathbb{P}^3$ is a hypersurface of degree γ in the Grassmannian $\mathrm{Gr}(1,\mathbb{P}^3)$, containing all lines in $\mathbb{P}^3$ that intersect Z (see Section 1.5.2). The ideal of C is in fact the saturation of $\langle \mathrm{Ch}_L, \mathrm{Ch}_X \rangle$ with respect to $\cap_{i=1}^{\beta-1}\langle Pu_i\rangle$, where u_i are the intersections between L and X, and Ch_L and Ch_X are the Chow forms of L and X respectively. Hence the ideal $\langle \mathrm{Ch}_L, \mathrm{Ch}_X \rangle$ represents all lines that intersect both L and X. The saturation removes $\beta-1$ extraneous components, namely the $(1,0)$-congruences of lines passing through the points u_i. We conjecture that the resulting ideal is generated by the Plücker quadric, the linear Chow form Ch_L, and

β linearly independent forms of degree β (including Ch_X). This description was observed experimentally.

Example 2.3.7. Fix $\beta = 3$ and let X be the twisted cubic curve given as in (2.13) with $f = (s-t)(s+t)$, $g = s^3$ and $h = t^3$. The ideal of X is generated by the 2×2-minors of

$$\begin{bmatrix} x_2 + x_4 & x_3 - x_1 & x_4 \\ x_3 & x_2 + x_4 & x_3 - x_1 \end{bmatrix}. \tag{2.15}$$

The line $L = V(x_1, x_2)$ meets the curve X in the two points $[0,0,1,1]$ and $[0,0,1,-1]$. The corresponding $(1,3)$-congruence C is parametrized by (2.14). The ideal of C equals

$$\begin{aligned} \langle\, p_{12}\,,\, p_{14}p_{23} - p_{13}p_{24}\,,\, p_{13}p_{14}^2 - p_{23}^2p_{24} - p_{13}p_{14}p_{34} + p_{23}p_{24}p_{34}, \\ p_{14}^3 - p_{23}p_{24}^2 - p_{14}^2p_{34} + p_{24}^2p_{34}\,,\, p_{13}^2p_{14} - p_{23}^3 - p_{13}^2p_{34} + p_{23}^2p_{34}\rangle. \end{aligned} \tag{2.16}$$

♢

Type 4: $\mathcal{F}(C)$ is a non-reduced line L. This is the degenerate case of Type 3 congruences when the binary form f is identically zero. The degree β morphism $\phi : L^\vee \to L$ promised in Theorem 2.3.3 sends $[x_1, x_2]$ to the point $[0,0,x_3,x_4] = [0,0,g(x_1,x_2),h(x_1,x_2)]$ on the line $L \subset \mathbb{P}^3$. The corresponding rational camera is given by the formula (2.14) with $f = 0$.

Example 2.3.8. Let $\beta = 3$ as in Example 2.3.7 but now with $f = 0$, $g = s^3$ and $h = t^3$. The non-reduced structure of L is the ideal $\langle x_1^2, x_1x_2, x_2^2\rangle$, obtained from (2.15) by setting $x_4 = 0$. The ideal of the resulting $(1,3)$-congruence is

$$\langle\, p_{12}\,,\, p_{14}p_{23} - p_{13}p_{24}\,,\; p_{13}p_{14}^2 - p_{23}^2p_{24}\,,\; p_{14}^3 - p_{23}p_{24}^2\,,\; p_{13}^2p_{14} - p_{23}^3\,\rangle.$$

♢

2.4 Multi-view Geometry

In this section we study the multi-view geometry of systems of rational cameras. For this, we first investigate in Section 2.4.1 the conditions for multiple lines to be all concurrent at a single point, deriving a complete algebraic characterization of the *concurrent lines variety*. We apply this result in Section 2.4.2 to describe multi-view correspondences for arbitrary rational cameras.

2.4.1 The concurrent lines variety

The *concurrent lines variety* V_n consists of ordered n-tuples of lines in $\mathbb{P}^3$ that meet in a point x. The lines containing a fixed x form a linear space of constant dimension 2 (the α-plane for x). From this we infer that V_n is irreducible of dimension $2n + 3$,

provided $n \geq 2$. We regard V_n as a subvariety in the product of projective spaces $(\mathbb{P}^5)^n$. The following result fully characterizes the prime ideal $I_n = I(V_n)$ of the concurrent lines variety in the polynomial ring of $6n$ Plücker coordinates.

Theorem 2.4.1. *Let $P_1, P_2, \ldots, P_n$ be skew-symmetric 4×4-matrices of unknowns that represent lines in $\mathbb{P}^3$, and let $P_1^*, P_2^*, \ldots, P_n^*$ be the dual matrices. The ideal I_n is minimally generated by the $\binom{n+1}{2}$ quadrics $\mathrm{trace}(P_i P_j^*)$ and the $10\binom{n}{3}$ cubics obtained as 3×3-minors of $(P_1u, P_2u, \ldots, P_nu)$ where u runs over $\{e_1, e_2, e_3, e_4, e_1{+}e_2, e_1{+}e_3, \ldots, e_3{+}e_4\}$, and e_1, e_2, e_3, e_4 represent the fundamental points in $\mathbb{P}^3$. For the reverse lexicographic order, the reduced Gröbner basis of I_n consists of $\binom{n+1}{2}$ quadrics, $12\binom{n}{3}$ cubics and $4\binom{n+1}{4}$ quartics.*

Our proof rests on computations with the computer algebra system `Macaulay2` [70].

Proof. The case $n = 2$ is easy. We begin with $n = 3$. Let P, Q, R be skew-symmetric 4×4-matrices representing three lines. These matrices have rank 2. The Plücker quadrics are

$$\mathrm{trace}(PP^*) \;=\; \mathrm{trace}(QQ^*) \;=\; \mathrm{trace}(RR^*) \;=\; 0. \tag{2.17}$$

Furthermore, the three lines are pairwise concurrent if and only if

$$\mathrm{trace}(PQ^*) \;=\; \mathrm{trace}(PR^*) \;=\; \mathrm{trace}(QR^*) \;=\; 0. \tag{2.18}$$

Using a computation with `Macaulay2`, we find that the ideal generated by the six quadrics in (2.17) and (2.18) is radical. It is the intersection of two prime ideals, each minimally generated by ten cubics in addition to (2.17) and (2.18). The first prime represents triples of lines that are coplanar, and is an extraneous component for us. The second prime is the concurrent lines variety. The cubic generators of that second prime ideal are the 3×3 minors of the 4×3 matrix (Pu, Qu, Ru), where u is a column vector in $\mathbb{R}^4$. These span a ten-dimensional space of cubics. A basis for that space is obtained by selecting the vector u from the set

$$\{\, e_1\,,\; e_2\,,\; e_3\,,\; e_4\,,\; e_1+e_2\,,\; e_1+e_3\,,\; e_1+e_4\,,\; e_2+e_3\,,\; e_2+e_4\,,\; e_3+e_4\,\}. \tag{2.19}$$

We note that the cubics for coplanar triples of lines are the 3×3-minors of the 4×3-matrix (P^*u, Q^*u, R^*u), where $u \in \mathbb{R}^4$. A basis of 10 cubics is obtained from the same set (2.19).

Using `Macaulay2`, we now compute the reduced Gröbner basis of our prime ideal from the $6 + 10 = 16$ generators with respect to the reverse lexicographic order determined by

$$p_{12}{>}p_{13}{>}p_{14}{>}p_{23}{>}p_{24}{>}p_{34} \,>\, q_{12}{>}q_{13}{>}q_{14}{>}q_{23}{>}q_{24}{>}q_{34} \,>\, r_{12}{>}r_{13}{>}r_{14}{>}r_{23}{>}r_{24}{>}r_{34}.$$

The initial monomial ideal is generated by the leading terms in the reduced Gröbner basis:

$$\begin{aligned} M_3 \;=\; \big\langle\, & p_{14}p_{23},\, q_{14}q_{23},\, r_{14}r_{23}\,,\;\; p_{34}q_{12},\, p_{34}r_{12},\, q_{34}r_{12}, \\ & p_{23}q_{13}r_{12}\,,\;\; p_{23}q_{14}r_{12}\,,\;\; p_{23}q_{14}r_{23}\,,\;\; p_{23}q_{14}r_{13}, \\ & p_{24}q_{13}r_{12}\,,\;\; p_{24}q_{14}r_{12}\,,\;\; p_{24}q_{14}r_{23}\,,\;\; p_{24}q_{14}r_{13}\,, \\ & p_{34}q_{14}r_{13}\,,\;\; p_{34}q_{14}r_{23}\,,\;\; p_{34}q_{24}r_{13}\,,\;\; p_{34}q_{24}r_{23}\,, \\ & p_{23}q_{13}q_{24}r_{13}\,,\;\; p_{23}q_{13}q_{24}r_{23}\,,\;\; p_{24}q_{13}q_{24}r_{13}\,,\;\; p_{24}q_{13}q_{24}r_{23}\,\big\rangle. \end{aligned} \tag{2.20}$$

This shows that the reduced Gröbner basis consists of 6 quadrics, 12 cubics and 4 quartics. All 22 leading terms are squarefree. This completes the proof of Theorem 2.4.1 for $n = 3$.

We next consider the case $n = 4$. A `Macaulay2` computation verifies that Theorem 2.4.1 is true here. The ideal I_4 is minimally generated by the 10 quadrics $\mathrm{trace}(P_i P_j^*)$ together with $40 = 10\binom{4}{3}$ cubics, namely the 10 cubics from I_3 for any three of the four lines. The initial ideal $M_4 = \mathrm{in}(I_4)$ is minimally generated by 10 quadratic monomials, $48 = 12\binom{4}{3}$ cubic monomials, and $20 = 4\binom{5}{4}$ quartic monomials. The quadrics and cubics come from M_3 for any three of the four lines. Among the quartics are the $16 = 4\binom{4}{3}$ quartics from M_3 for any three of the four lines. However, the reduced Gröbner basis of I_4 now also contains four quadrilinear forms. These contribute four new generators of the monomial ideal M_4:

$$p_{23}q_{13}r_{24}s_{13}\,,\;\; p_{23}q_{13}r_{24}s_{23}\,,\;\; p_{24}q_{13}r_{24}s_{23}\,,\;\; p_{24}q_{13}r_{24}s_{13}. \tag{2.21}$$

We next assume $n \geq 5$. We write $\mathcal{G}_n$ for the union of the various reduced Gröbner bases, obtained from I_4 for any four of the n lines. The set $\mathcal{G}_n$ has $\binom{n+1}{2}$ quadrics $\mathrm{trace}(P_i P_j^*)$, and it has $12\binom{n}{3}$ cubics, namely those having the 12 leading terms in (2.20), for any three lines. Finally, there are $4\binom{n+1}{4} = 4\binom{n}{3} + 4\binom{n}{4}$ quartics in $\mathcal{G}_n$. Their leading monomials are the quartics in (2.20), for any three lines, and the quartics in (2.21), for any four of the n lines.

We claim that $\mathcal{G}_n$ is the reduced Gröbner basis for the ideal $\langle \mathcal{G}_n \rangle$ it generates. This can be verified computationally with `Macaulay2` for $n \leq 7$. For $n \geq 8$, we argue as follows. Consider any two polynomials in $\mathcal{G}_n$. We must show that their S-polynomial reduces to zero upon division with respect to $\mathcal{G}_n$. If their leading monomials are relatively prime then this is automatic, by Buchberger's First Criterion [39]. Otherwise, the leading monomials have a Plücker variable in common. This means that at most seven of the n lines are involved in the two polynomials. But then their S-polynomial reduces to zero because the Gröbner basis property is already known for $n \leq 7$. A similar argument shows that no trailing term in $\mathcal{G}_n$ is a multiple of an leading term. Hence $\mathcal{G}_n$ is the reduced Gröbner basis for its ideal.

The minimal generators of the ideal $\langle \mathcal{G}_n \rangle$ are obtained from the minimal generators of I_4, for any four of the n lines. Hence $\langle \mathcal{G}_n \rangle$ is generated by the $\binom{n+1}{2}$ quadrics and the

$10\binom{n}{3}$ cubics that are listed in the statement of Theorem 2.4.1. Its leading terms are square-free.

Finally, we must prove that the ideal $\langle \mathcal{G}_n \rangle$ equals the ideal I_n we are interested in. By construction, all generators of $\mathcal{G}_n$ vanish on the concurrent lines variety $V_n = V(I_n)$. Therefore,

$$\langle \mathcal{G}_n \rangle \ \subseteq \ I_n. \tag{2.22}$$

Moreover, the initial ideal of $\langle \mathcal{G}_n \rangle$ is radical, and hence $\langle \mathcal{G}_n \rangle$ is a radical ideal. To complete the proof, all we now need is that the set $\mathcal{G}_n$ cuts out the variety V_n set-theoretically. This is equivalent to the statement that $n \geq 4$ distinct lines in $\mathbb{P}^3$ are concurrent if and only if any three of the n lines are concurrent. This is indeed the case. □

Remark 2.4.2. Suppose all P_i satisfy the Plücker constraint $\text{trace}(P_i P_i^*) = 0$. The four 3×3-minors of $(P_i u, P_j u, P_k u)$ are scalar multiples of a single trilinear polynomial t_u that expresses the condition for the planes $u \vee p_i, u \vee p_j$ and $u \vee p_k$ to be linearly dependent, *i.e.*, for p_i, p_j, p_k to admit a *transversal line* passing through u: this is exactly the trilinear relation t_p considered in Section 1.5.1. In fact, as argued in Proposition 1.5.2, three lines are concurrent if and only if they are pairwise coplanar and they admit a transversal not contained in the planes defined by any two of them. From this we deduce that V_n is cut out *set-theoretically* by $\binom{n+1}{2}$ bilinear quadrics $\text{trace}(P_i P_j^*)$ and the $4\binom{n}{3}$ trilinear cubics T_u where u runs over only $\{e_1, e_2, e_3, e_4\}$. This is confirmed by computation with `Macaulay2`.

The *multidegree* of V_n in $(\mathbb{P}^5)^n$ is a homogeneous polynomial of degree $3n-3$ in n unknowns $t_1, t_2, \ldots, t_n$ (see Section 1.5.2). Using the built-in command `multidegree` in `Macaulay2`, we found experimentally that the multidegree of the concurrent lines variety is the polynomial

$$[V_n] \ = \ (t_1 t_2 t_3 \cdots t_n)^3 \cdot \Big(4 \sum_{(i,j)} t_i^{-2} t_j^{-1} \ + \ 8 \sum_{\{i,j,k\}} t_i^{-1} t_j^{-1} t_k^{-1} \Big). \tag{2.23}$$

The first sum is over ordered pairs (i,j) with $i \neq j$; the second sum is over unordered triples $\{i,j,k\}$. After completion of our paper [149], Laura Escobar and Allen Knutson [52] found a proof for the formula (2.23). Their derivation rests on methods from representation theory.

2.4.2 Multi-image varieties

In this section, we use the concurrent lines variety V_n to characterize multi-view correspondences for n rational cameras. We fix congruences $C_1, \ldots, C_n \subset \text{Gr}(1, \mathbb{P}^3)$, where C_i has bidegree $(1, \beta_i)$ for some $\beta_i \in \mathbb{N}$. Combining their maps as in (2.6) gives

$$\mathbb{P}^3 \dashrightarrow C_1 \times \cdots \times C_n \ , \ \ x \mapsto (C_1(x), \ldots, C_n(x)). \tag{2.24}$$

The base locus of this rational map (*i.e.*, its locus of indeterminacy) is the product of the focal loci, $\mathcal{F}(C_1) \times \cdots \times \mathcal{F}(C_n)$. We define the *multi-image variety* $M(C_1, \ldots, C_n)$ to be the closure of the image of (2.24). This is an irreducible subvariety in the product of Grassmannians $\mathrm{Gr}(1, \mathbb{P}^3)^n \subset (\mathbb{P}^5)^n$. We expect the map (2.24) to be birational in almost all cases, so $M(C_1, \ldots, C_n)$ is a threefold.

The multi-image variety is clearly contained in the concurrent lines variety. In symbols,

$$M(C_1, \ldots, C_n) \subseteq V_n \cap (C_1 \times \cdots \times C_n) \subset \mathrm{Gr}(1, \mathbb{P}^3)^n. \tag{2.25}$$

Our first result in this section shows that the left inclusion in (2.25) is usually an equality.

Theorem 2.4.3. *Suppose that the n focal loci $\mathcal{F}(C_1), \ldots, \mathcal{F}(C_n)$ are pairwise disjoint. Then*

$$M(C_1, \ldots, C_n) = V_n \cap (C_1 \times \cdots \times C_n), \tag{2.26}$$

i.e., the concurrent lines variety gives an implicit representation of the multi-image variety.

Proof. By (2.25), we only need to show one direction. For any n-tuple of lines $(L_1, \ldots, L_n)$ in $V_n \cap M(C_1, \ldots, C_n)$, there exists $x \in \mathbb{P}^3$ such that $x \in L_i$ for all i. If x does not lie in any of the n focal loci then $L_i = C_i(x)$ and we are done. Otherwise, x lies in exactly one of the focal loci, say, $x \in \mathcal{F}(C_i)$. We then consider a sequence of nearby pairs $(x_\epsilon, L_{i,\epsilon})$ that converges to (x, L_i) and satisfies $x_\epsilon \in L_{i,\epsilon} \backslash \mathcal{F}(C_i)$ and $C_i(x_\epsilon) = L_{i,\epsilon}$. For each $j \in \{1, 2, \ldots, n\} \backslash \{i\}$ the locus $\mathcal{F}(C_j)$ is closed. Since it does not contain x, we can assume that it also does not contain x_ϵ. Hence $(C_1(x_\epsilon), \ldots, C_n(x_\epsilon))$ is a well-defined sequence of points in the variety $M(C_1, \ldots, C_n)$. It converges to $(L_1, \ldots, L_n)$, which therefore also lies in $M(C_1, \ldots, C_n)$. □

We now give an example for $n = 3$ that shows the necessity of the hypothesis on the focal loci in Theorem 2.4.3.

Example 2.4.4. Let $\beta_1 = \beta_2 = \beta_3 = 1$ and fix three two-slit cameras C_1, C_2, C_3 defined by

$$J = \langle p_{12}, p_{34}, q_{13}, q_{24}, r_{14}, r_{23} \rangle \subset \mathbb{R}[p_{12}, \ldots, p_{34}, q_{12}, \ldots, q_{34}, r_{12}, \ldots, r_{34}].$$

Geometrically, we partition the set of six coordinate lines in $\mathbb{P}^3$ into three pairs of disjoint lines (see Figure 2.3). Each pair defines a $(1,1)$-congruence. Note that $\mathcal{F}(C_1)$, $\mathcal{F}(C_2)$ and $\mathcal{F}(C_3)$ are distinct, but they intersect in the four coordinate points. So, the hypothesis of Theorem 2.4.3 fails.

The ideal $J + I_3$ is radical but not prime. It is the intersection of five primes, each defining a threefold in $\mathbb{P}^5 \times \mathbb{P}^5 \times \mathbb{P}^5$. One of these is the variety $M(C_1, C_2, C_3)$, with ideal

$$\begin{aligned}\langle p_{14}p_{23} - p_{13}p_{24}, q_{14}q_{23} + q_{12}q_{34}, r_{13}r_{24} - r_{12}r_{34}, p_{14}q_{23} + p_{23}q_{14}, p_{24}r_{13} + p_{13}r_{24}, q_{34}r_{12} + q_{12}r_{34}, \\ p_{23}q_{34}r_{24} + p_{24}q_{23}r_{34},\ p_{13}q_{34}r_{24} + p_{14}q_{23}r_{34},\ p_{14}q_{23}r_{24} + p_{24}q_{12}r_{34},\ p_{13}q_{23}r_{24} + p_{23}q_{12}r_{34}, \\ p_{13}q_{14}r_{24} - p_{14}q_{12}r_{34},\ \ p_{14}q_{34}r_{13} + p_{13}q_{14}r_{34},\ \ p_{14}q_{23}r_{13} - p_{13}q_{12}r_{34}, \\ p_{14}q_{23}r_{12} - p_{13}q_{12}r_{24},\ \ p_{13}q_{23}r_{12} + p_{23}q_{12}r_{13},\ \ p_{24}q_{14}r_{12} + p_{14}q_{12}r_{24}\,\rangle \ +\ J.\end{aligned}$$

The other four associated primes define coordinate 3-planes in $\mathbb{P}^5 \times \mathbb{P}^5 \times \mathbb{P}^5$. They are

$$\begin{aligned}&\langle p_{23}, p_{24}, q_{23}, q_{34}, r_{24}, r_{34}\rangle + J, \quad \langle p_{13}, p_{14}, q_{14}, q_{34}, r_{13}, r_{34}\rangle + J, \\ &\langle p_{14}, p_{24}, q_{12}, q_{14}, r_{12}, r_{24}\rangle + J, \quad \langle p_{13}, p_{23}, q_{12}, q_{23}, r_{12}, r_{13}\rangle + J.\end{aligned}$$

To understand the geometric meaning of these extraneous components, consider the last ideal. It represents all triples (L_1, L_2, L_3) where L_1, L_2, L_3 pass through $[0, 0, 0, 1]$, and each line L_i intersects one of the opposite coordinate lines, as is required for lines in C_i. ♢

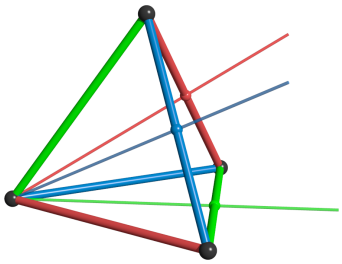


Figure 2.3: Three two-slit cameras C_1, C_2, C_3 with intersecting focal loci. The triple of concurrent lines belongs to $C_1 \times C_2 \times C_3$ but not to the multi-image variety $M(C_1, C_2, C_3)$. See Example 2.4.4 for details.

In the following we always consider congruences whose focal loci are pairwise disjoint, so the identity (2.26) holds. We present a detailed study of two special cases. First we allow arbitrary n, but assume $\beta_i \in \{0, 1\}$. We then focus on $n = 2$, but with arbitrary β_1 and β_2.

Multiple Views with Pinhole and Two-Slit Cameras. We begin with the most classical case, where $C_1, \ldots, C_n$ are pinhole cameras with distinct centers $c_1, \ldots, c_n$. Each congruence C_i is a plane in $\mathbb{P}^5$, and the map $x \mapsto C_i(x) = x \vee c_i$ is analogous to the linear projection $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ with center c_i. In the usual set-up with pinhole cameras, this map is represented by a 3×4 matrix A_i whose null-space is given by c_i, and $\mathbb{P}^2$ is identified with the image of A_i. Since A_i and $x \mapsto x \vee c_i$ have the same null-space, there exists a 6×3 matrix B_i such that $x \vee c_i = B_i A_i x$; this is of course simply the "inverse projection matrix" from Chapter 1, Section 1.2.5.

Proposition 2.4.5. *The joint image variety (see Definition 1.3.4) of the pinhole cameras $A_1, \ldots, A_n$ is isomorphic to the multi-image variety $M(C_1, \ldots, C_n)$ under the map*

$$(\mathbb{P}^2)^n \to \mathrm{Gr}(1, \mathbb{P}^3)^n, \ (u_1, \ldots, u_n) \mapsto (B_1 u_1, \ldots, B_n u_n).$$

Here, the equation (2.26) holds ideal-theoretically, i.e., *the prime ideal of the multi-view variety is the image of I_n modulo the linear equations $P_1 c_2 = \cdots = P_n c_n = 0$ that define $C_1 \times \cdots \times C_n$.*

Proof. The first statement is immediate from the discussion of the two realizations of $\mathbb{P}^2$, as the image of A_i or as the plane C_i in $\mathbb{P}^5$. The second statement about ideals is more subtle. It can be derived using the functorial set-up developed by Li [115]. □

The ideal I_n of the concurrent lines variety V_n is minimally generated by $\binom{n+1}{2}$ quadrics and $10\binom{n}{3}$ cubics in the $6n$ Plücker coordinates. We add to this the $3n$ linear equations that define $C_1 \times \cdots \times C_n$. This reduces the minimal generators to $\binom{n}{2}$ quadrics and $\binom{n}{3}$ cubics. These are the bilinearities and trilinearities from Theorem 1.3.10.

We next generalize Proposition 2.4.5 to arrangements of n_1 pinhole cameras $C_1, \ldots, C_{n_1}$ and n_2 two-slit cameras $C'_1, \ldots, C'_{n_2}$. These $n = n_1 + n_2$ cameras are assumed to satisfy the hypothesis of Theorem 2.4.3. Thus, the pinholes are distinct, the slits are pairwise disjoint, and no pinhole is allowed to lie on a slit.

Theorem 2.4.6. *The ideal of the multi-image variety $M(C_1, \ldots, C_{n_1}, C'_1, \ldots, C'_{n_2})$ is minimally generated by $3n_1 + 2n_2$ linear forms, $\binom{n_1+n_2}{2} + n_2$ quadrics, and $\binom{n_1}{3} + 3\binom{n_1}{2}n_2 + 6n_1\binom{n_2}{2} + 10\binom{n_2}{3}$ cubics in the $6n_1 + 6n_2$ Plücker coordinates on the ambient space $(\mathbb{P}^5)^{n_1+n_2}$.*

Note that for $n_2 = 0$ we recover the known ideal generators of the joint image variety (Theorem 1.3.10).

Proof. The desired ideal is obtained from I_n by adding 3 linear forms for every pinhole camera C_i and 2 linear forms for every two-slit camera C'_i. We need to examine the extent to which the generators of I_n become linearly dependent modulo these $3n_1 + 2n_2$ linear forms. For $n \leq 3$ cameras this examination amounts to computations with `Macaulay2`, one for each ordered partition (n_1, n_2) of n. For $n \geq 4$ cameras we group the minimal generators of I_n according to their degree in the $\mathbb{Z}^n$-grading. Each graded component specifies a subset of cameras of size at most three. Hence all the linear relations arise from those for $n = 3$. □

Epipolar Geometry for Rational Cameras. We now take a closer look at the case of two rational cameras C_1 and C_2. We assume that C_i is a congruence of bidegree $(1, \beta_i)$ for $i = 1, 2$ and that $\mathcal{F}(C_1) \cap \mathcal{F}(C_2) = \emptyset$. The associated multi-image variety $M(C_1, C_2)$ in $\mathbb{P}^5 \times \mathbb{P}^5$ is defined by the ideal

$$I_P(C_1) + I_Q(C_2) + \langle \mathrm{trace}(PQ^*) \rangle \quad \subset \quad \mathbb{C}[p_{12}, \ldots, p_{34}, q_{12}, \ldots, q_{34}], \tag{2.27}$$

where $I_P(C_1)$ and $I_Q(C_2)$ are respectively the ideals for C_1 and C_2 in the two sets of variables.

This set-up generalizes familiar objects from two-view geometry. For example, if p is a line in the congruence C_1, then the *epipolar curve* $\mathrm{Epi}(p)$ in C_2 consists of all lines q such that (p, q) belongs to $M(C_1, C_2)$. The ideal of $\mathrm{Epi}(p)$ in the $\mathbb{P}^5$ with coordinates $q_{12}, \ldots, q_{34}$ is given by $I_Q(C_2) + \langle \mathrm{trace}(PQ^*) \rangle$. The curve $\mathrm{Epi}(p)$ has degree $1 + \beta_2$ in Plücker coordinates (see Proposition 2.4.7 below). In particular, for pinhole cameras C_1, C_2, we recover the classical epipolar lines in two-view geometry [78]. However there is an important (and well known [173]) distinction compared to the classical case: if either C_1 or C_2 is not a pinhole cameras, then the families of curves $\mathrm{Epi}_{12} = \{\mathrm{Epi}(p) : p \in C_1\}$ and $\mathrm{Epi}_{21} = \{\mathrm{Epi}(q) : q \in C_2\}$ are *not* in general related by a one-to-one correspondence. More concretely, if q and q' both belong to $\mathrm{Epi}(p)$, then we cannot conclude that $\mathrm{Epi}(q) = \mathrm{Epi}(q')$. This follows from the fact that the ideal from Theorem 2.4.6 is not multilinear, and contrasts with the case of pinholes, where there exists a homography relating the $\mathbb{P}^1$ of epipolar lines in each image.

In traditional two-view geometry, the two camera centers in $\mathbb{P}^3$ span the "baseline", which projects onto the two epipoles. This generalizes as follows to our setting. A line L in $\mathbb{P}^3$ is a *baseline* for the two cameras C_1 and C_2 if it lies in the intersection $C_1 \cap C_2$ in $\mathrm{Gr}(1, \mathbb{P}^3)$. The baselines are precisely the loci that are contracted by the map (2.24), since for every point x in such a line L we have $(C_1(x), C_2(x)) = (L, L)$. We expect $C_1 \cap C_2$ to consist of finitely many points. Some of these points are defined over $\mathbb{C}$, and are included in the following count.

Proposition 2.4.7. *Let C_1 and C_2 be general congruences of bidegree $(1, \beta_1)$ and $(1, \beta_2)$. The epipolar curves $\mathrm{Epi}(p)$ and $\mathrm{Epi}(q)$ in $\mathrm{Gr}(1, \mathbb{P}^3)$ have degrees $1 + \beta_2$ and $1 + \beta_1$ respectively. The number of baselines in $\mathbb{P}^3$ for the camera pair (C_1, C_2) equals $1 + \beta_1\beta_2$.*

Proof. The intersection theory in the Grassmannian $\mathrm{Gr}(1, \mathbb{P}^3) \subset \mathbb{P}^5$ works as follows. A hypersurface of degree d intersects an (α, β)-congruence in a curve of degree $d\alpha + d\beta$. Two congruences of bidegrees (α_2, β_1) and (α_3, β_2) intersect in $\alpha_2\alpha_3 + \beta_1\beta_2$ points. A classical reference is Jessop's book [92]. A modern one is any introduction to *Schubert calculus*. □

We now illustrate these concepts with an example.

Example 2.4.8. Let $\beta_1 = \beta_2 = 2$ and consider the type 3 congruences C_1 and C_2 of common transversals to L_1, X_1 and L_2, X_2, where

$$L_1 = V(x_2, x_3 - x_4), \quad X_1 = V(x_1, x_2^2 + x_3^2 - x_4^2),$$

$$L_2 = V(x_1 - x_2, x_3), \quad X_2 = V(x_1^2 - x_2^2 + x_3^2, x_4).$$

Note that that the intersection of $\mathcal{F}(C_1) = L_1 \cup X_1$ and $\mathcal{F}(C_2) = L_2 \cup X_2$ in $\mathbb{P}^3$ is empty. The intersection points on the two focal loci are $L_1 \cap X_1 = \{[0, 0, 1, 1]\}$ and

$L_2 \cap X_2 = \{[1,1,0,0]\}$. The prime ideals of the two congruences in the coordinate ring of $\mathrm{Gr}(1,\mathbb{P}^3)$ are given by

$$I_P(C_1) \;=\; \langle p_{23}-p_{24},\; p_{12}^2+p_{13}^2-p_{14}^2,\; p_{12}p_{24}+p_{13}p_{34}+p_{14}p_{34},\; p_{12}p_{34}-p_{13}p_{24}+p_{14}p_{23}\rangle,$$

$$I_P(C_2) \;=\; \langle p_{13}-p_{23},\; p_{14}^2-p_{24}^2+p_{34}^2,\; p_{12}p_{14}+p_{12}p_{24}-p_{23}p_{34},\; p_{12}p_{34}-p_{13}p_{24}+p_{14}p_{23}\rangle.$$

In both expressions, the first two polynomials are the Chow forms of L_i and X_i respectively. The ideal of the two-image variety $M(C_1,C_2)$ is given by (2.27).

If we fix a point p in C_1 then its corresponding cubic curve $\mathrm{Epi}(p)$ lives in C_2, and vice versa. For example, the ideal $I_P(C_2)+\langle 3p_{12}+4p_{13}+5p_{14}-3p_{23}-3p_{24}+p_{34}\rangle$ defines the epipolar curve in $\mathrm{Gr}(1,\mathbb{P}^3)$ associated with $p=[3,4,5,-3,-3,1]$ in C_1.

The ideal $I_P(C_1)+I_P(C_2)$ defines five points in $\mathrm{Gr}(1,\mathbb{P}^3)$, These represent the five baselines. One point is $[0,1,1,1,1,0]$. It represents the line through $[0,0,1,1]$ and $[1,1,0,0]$. The other four baselines have the Plücker vectors

$$\left[\frac{5}{2}a^2-\frac{1}{2},\, a,\, -\frac{5}{2}a^3-\frac{1}{2}a,\, a,\, a,\, 1\right] \qquad \text{where} \quad 5a^4-2a^2+1=0.$$

We see that three of the five baselines are real, while the other two are defined over $\mathbb{C}$. ♢

2.5 Higher-Order Cameras

In Sections 2.3.2 and 2.4.2 we considered congruences whose point-to-line maps $x \mapsto C(x)$ are rational. However, there are important examples of real-world cameras that are associated with congruences of higher order (see *e.g.*, [173]). For example, a $(2,\beta)$-congruence associates a given point x with a pair of lines, but a physical camera might record only one line for x, because of orientation constraints. In this section, we study the geometry in $\mathrm{Gr}(1,\mathbb{P}^3)$ for two types of devices that exist in practice, namely "non-central panoramic" and "catadioptric cameras". We consider here idealized and generalized versions of these models. For example, we describe catadioptric systems that use an arbitrary smooth algebraic surface as a mirror, and a general geometric camera to record lines.

2.5.1 Panoramic Cameras

A panoramic camera enables photographs with a 360° field of view. One such panoramic device consists of a 1D-sensor measuring 2D-projections onto a fixed center, that is rotated about a vertical axis L not containing the center. The 1D-sensor travels on a circle X around the line L. The associated congruence C consists of all lines that intersect both L and X, and has bidegree $(2,2)$. Physical realizations come in two versions, depending on the orientation of the sensor (which causes precisely one of the

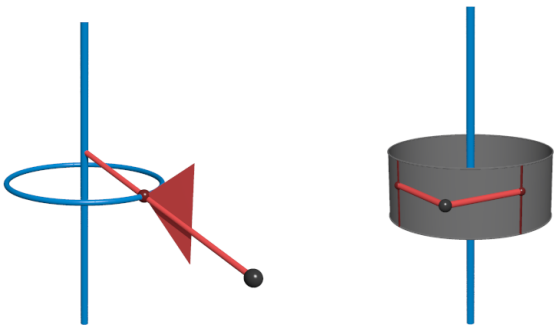


Figure 2.4: Panoramic cameras: non-central (*left*) and stereo (*right*)

two lines of C through a point x to be recorded): if the sensor points outwards then we get a *non-central panoramic camera* (see Figure 2.4, *left*); if the sensor points inwards then the camera is a *cyclograph*, a device that records a 360° representation of a single object placed in the middle.

Another system is the *stereo panoramic camera*, on the right in Figure 2.4. It is obtained by rotating a 1D-sensor about an axis parallel to the sensor. In each position the sensor records parallel lines tangent to the rotation. This is a variation of the "omnivergent" camera proposed in [165]. It produces stereo (binocular) panoramic images, since every 3D-point is observed from two sensor locations. Both of the cameras shown in Figure 2.4 are discussed in [173].

We first discuss the non-central panoramic camera. The corresponding $(2,2)$-congruence C is determined by a line L and a non-degenerate conic X, both in $\mathbb{P}^3$, such that $L \cap X = \emptyset$. Then, as above, C consists of all lines in $\mathbb{P}^3$ that intersect both L and X.

Lemma 2.5.1. *Any two such congruences are equivalent up to projective transformations of* $\mathbb{P}^3$.

Proof. Given any two pairs of disjoint conics and lines (X, L) and (X', L') in $\mathbb{P}^3$, we may always apply a homography over $\mathbb{C}$ so that $X = X' = V(x_1, x_2^2 + x_3^2 + x_4^2)$. Transformations that fix X are projectivizations of affine maps $\mathbb{C}^3 \to \mathbb{C}^3$, $\tilde{x} \mapsto A\tilde{x} + b$ such that $AA^T = \lambda \mathrm{Id}$, where $\tilde{x} = (x_2/x_1, x_3/x_1, x_4/x_1)$ are coordinates on the affine chart $U_1 = \{x_1 \neq 0\}$. These maps act transitively on points of U_1 and on points of $V(x_1) \backslash X$, so we conclude that L and L' are equivalent. If we restrict to $\mathbb{R}$, and both conics have real points, then we use $X = X' = V(x_1, x_2^2 + x_3^2 - x_4^2)$, and a similar result holds. □

Thanks to Lemma 2.5.1, we may choose $I_L = \langle x_2, x_3 \rangle$ and $I_X = \langle x_4, x_2^2 + x_3^2 - x_1^2 \rangle$ as the ideals that represent the line L and the conic X. The ideal of the congruence C in $\mathbb{R}[p_{12}, \ldots, p_{34}]$ is generated by the Chow forms of L and X together with the Plücker quadric:

$$I_C = \langle p_{23}\,,\ p_{14}^2 - p_{24}^2 - p_{34}^2\,,\ p_{12}p_{34} - p_{13}p_{24} + p_{14}p_{23} \rangle. \tag{2.28}$$

We can see that C has bidegree $(2,2)$ by counting the intersections with generic α-planes and β-planes. Since C has order higher than one, its focal locus is two-dimensional.

Proposition 2.5.2. *The focal locus $\mathcal{F}(C)$ of the non-central panoramic camera consists of the plane spanned by the conic X, taken with multiplicity 2, and a conjugate pair of complex planes that intersect in the line L. Algebraically, it is defined by the non-reduced quartic $(x_2^2 + x_3^2)x_4^2$.*

Proof. We compute the focal locus as described at the end of Section 2.2. For the congruence C given in (2.28), the quartic generator is found to be $(x_2^2 + x_3^2)x_4^2$. □

Remark 2.5.3. It is known that the focal locus of a *general smooth* $(2,2)$-congruence (defined by a general linear form and a general quadratic form in Plücker coordinates) is a *Kummer surface*: this is a classical algebraic surface of degree four with 16 nodes. However, the congruence C in (2.28) is singular: its singular locus, $V(p_{14}, p_{23}, p_{24}, p_{34})$, consists in of all lines in $\mathbb{P}^3$ that meet L and lie in the plane spanned by X. The Kummer surface degenerates to the arrangement of four planes $\mathcal{F}(C)$ in Proposition 2.5.2.

We now discuss the stereo panoramic camera. Its congruence consists of the lines that are tangent to a singular quadratic surface Q and pass through a fixed line L. For the camera on the right in Figure 2.4, the quadric Q is a cylinder around the axis and L is a line at infinity. Interestingly, we note that the stereo panoramic camera is *dual*, in the sense of projective geometry, to the non-central panoramic camera. Specifically, its congruence C^* is obtained by dualizing C in (2.28). The result is

$$I_{C^*} = \langle\, p_{14}\,,\ p_{23}^2 - p_{13}^2 - p_{12}^2\,,\ p_{12}p_{34} - p_{13}p_{24} + p_{14}p_{23}\rangle. \tag{2.29}$$

Here $I_L = \langle x_1, x_4\rangle$ and $I_Q = \langle x_1^2 - x_2^2 - x_3^2\rangle$. Clearly, C^* has bidegree $(2,2)$. As in Remark 2.5.3, C^* is singular along a line. Singular points are lines that meet L and the cone point of Q.

Proposition 2.5.4. *The focal locus $\mathcal{F}(C^*)$ of the stereo panoramic camera consists of the singular quadric Q and the plane at infinity (spanned by L and the cone point of Q), which is taken with multiplicity 2. Algebraically, it is defined by the non-reduced quartic $x_1^2(x_1^2 - x_2^2 - x_3^2)$.*

Proof. This is verified by a computation, like Proposition 2.5.2. □

2.5.2 Catadioptric Cameras

A *catadioptric camera* is an optical device that makes use of reflective surfaces (catoptrics) and lenses (dioptrics). For many applications it is desirable to have a single "effective viewpoint" [9]. This can be achieved by using a mirror that is a paraboloid or hyperboloid of revolution, placing a pinhole camera at one of the foci. In our setting, it is

natural to consider a more general catadioptric system that uses an arbitrary smooth algebraic surface S as a mirror, and a geometric camera C to record lines. We will describe the resulting line congruence in $\mathrm{Gr}(1,\mathbb{P}^3)$.

We measure angles in $\mathbb{P}^3$ according to the usual scalar product in $U_1 = \{x_1 \neq 0\}$, so that

$$\cos \angle(x,y) \;=\; \frac{x_2y_2 + x_3y_3 + x_4y_4}{\sqrt{(x_2^2+x_3^2+x_4^2)(y_2^2+y_3^2+y_4^2)}}. \tag{2.30}$$

The points $x = [0, x_2, x_3, x_4]$ and $y = [0, y_2, y_3, y_4]$ lie on the plane at infinity, $\mathbb{P}^3 \setminus U_1$, and they represent directions in U_1. Let H be the plane in $\mathbb{P}^3$ defined by $a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 = 0$. Here $a_i \in \mathbb{C}$ is allowed, but we assume that H is *non-isotropic*, meaning that $a_2^2 + a_3^2 + a_4^2 \neq 0$. With the convention above, the *reflection of a point* $y = [y_1, y_2, y_3, y_4]$ with respect to H is

$$\rho_H(y) = \begin{bmatrix} \left(\sum_{i=1}^3 a_i^2\right) y_1 \\ \left(\sum_{i=1}^3 a_i^2\right) y_2 - 2\left(\sum_{i=0}^3 a_i y_i\right) a_2 \\ \left(\sum_{i=1}^3 a_i^2\right) y_3 - 2\left(\sum_{i=0}^3 a_i y_i\right) a_3 \\ \left(\sum_{i=1}^3 a_i^2\right) y_4 - 2\left(\sum_{i=0}^3 a_i y_i\right) a_4 \end{bmatrix}. \tag{2.31}$$

This map is a linear involution of $\mathbb{P}^3$ that fixes H, and whose restriction to the real affine 3-space U_1 is the usual Euclidean reflection with respect to $H \cap U_1$. The reflection can also be applied to lines, by defining the image of $p = x \vee y$ as $\rho_H(x) \vee \rho_H(y)$. This is a linear involution of the Grassmannian $\mathrm{Gr}(1,\mathbb{P}^3)$, defined by the 6×6 compound matrix $C_2(\rho_H)$ of ρ_H.

We now consider a smooth algebraic surface S in $\mathbb{P}^3$, defined by a polynomial f of degree d. We say that two lines L and L' in $\mathbb{P}^3$ are *specular* for S if there exists a point $x \in S$ such that the tangent plane $T_x(S)$ is not isotropic, L and L' meet in x, and they are reflections of each other respect to $T_x(S)$. We define the *mirror variety* M_S to be the closure of the set of all pairs $(L, L') \in \mathrm{Gr}(1,\mathbb{P}^3)^2$ that are specular for S. For a general line L there are d lines L' such that $(L, L') \in M_S$, one for each point x in $S \cap L$. Hence the mirror variety M_S of a surface S is 4-dimensional.

To compute the defining equations of the mirror variety M_S, we first construct the ideal

$$J \;= \langle\, f,\, Px, \mathrm{trace}(PP^*), \mathrm{trace}(QQ^*) \rangle \;+\; \langle\, \wedge_2\, (q \,|\, \rho_{T_xS}(p)) \rangle.$$

This lives in $\mathbb{R}[x_1, x_2, x_3, x_3, p_{12}, \ldots, p_{34}, q_{12}, \ldots, q_{34}]$. The last summand is the ideal of 2×2-minors of a 6×2-matrix with columns q and $\rho_{T_xS}(p)$, where $\rho_{T_xS}(p)$ denotes the reflection of the line p with respect to the tangent plane $T_x(S)$. These constraints express the requirement that q is equal to that reflection. We then saturate J with respect to the isotropic ideal $I_{\mathrm{Iso}} = \langle\, (\nabla_x f)_1^2 + (\nabla_x f)_2^2 + (\nabla_x f)_3^2 \,\rangle$ and with respect to the irrelevant ideal $\langle x_1, x_2, x_3, x_4 \rangle$, before eliminating the variables x_1, x_2, x_3, x_4.

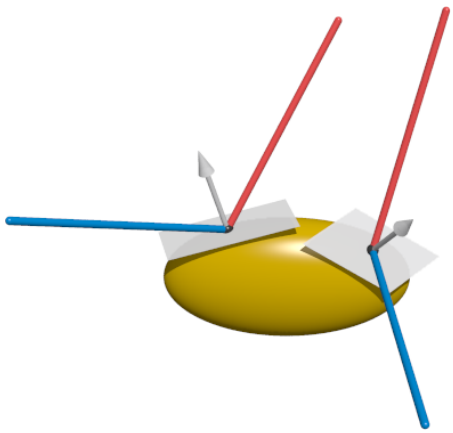


Figure 2.5: A general algebraic surface S maps a congruence C (red lines) to a "specular congruence" C_S (blue lines).

Example 2.5.5. Let S be the ellipsoid given by $f = \frac{1}{16}(x_2^2 + x_3^2) + \frac{1}{25}x_4^2 - x_1^2$. The mirror variety M_S has codimension 6 in $\mathbb{P}^5 \times \mathbb{P}^5$. Using the approach outlined above, we recover the ideal associated with M_S. We observe that its *bidegree* (see Appendix B) is $4t_0^5t_1 + 12t_0^4t_1^2 + 18t_0^3t_1^3 + 12t_0^2t_1^4 + 4t_0t_1^5$. ♢

Remark 2.5.6. The intersection of the mirror variety M_S with the diagonal Δ in $\mathbb{P}^5 \times \mathbb{P}^5$ is the *normal congruence*. These are the lines that intersect S orthogonally (we assume that we have removed components associated with isotropic tangent planes). The focal locus of the normal congruence is the *caustic surface* [32]. In the language of differential geometry, this is the union of the centers of principal curvature for S. Moreover, the order α of the normal congruence coincides with the *Euclidean distance degree* (ED degree) of the surface S. That is the number of critical points on S of the squared distance function to a generic point [48].

Let C be any congruence, representing a geometric camera. The *specular congruence* of C with respect to S is another surface C_S in the Grassmannian $\mathrm{Gr}(1, \mathbb{P}^3)$. We define C_S as the closure of the set of all lines L' for which there exist $L \in C$ and $x \in L \cap S$ such that T_xS is not isotropic and $L' = \rho_{T_xS}L$. Concretely, C_S contains the lines of C after these are reflected by S (see Figure 2.5.) Thus, C_S is the congruence associated with the catadioptric camera determined by S and C. Note that if L' is in C then there exists L such that $(L, L') \in M_S$. This implies

$$C_S \subseteq C'_S := \pi_2(M_S \cap (C \times \mathrm{Gr}(1, \mathbb{P}^3))), \tag{2.32}$$

where π_2 is the projection onto the second factor. For a general C, an appropriate application of Bertini's Theorem [73] ensures that the right hand side C'_S is irreducible, and the containment (2.32) is an equality (set-theoretically). In this case, we can compute equations for C_S by adding the equations defining the given congruence C (in the variables $p_{12}, \ldots, p_{34}$) to the ideal of the mirror variety M_S, then saturating by the irrelevant

ideal $\langle p_{12}, \ldots, p_{34} \rangle$, and finally eliminating the variables $p_{12}, \ldots, p_{34}$. We experimented with this in `Macaulay2`.

The next example shows that C'_S can have spurious components. These are removed by saturating the ideal of C'_S with respect to the Chow form $Ch_{X_{\text{Iso}}}$ where $X_{\text{Iso}} = V(I_{\text{Iso}} + \langle f \rangle)$. We note that the order and class of the specular congruence C_S depend on the relative position of S and C (and the absolute quadric $V(x_1, x_2^2 + x_3^2 + x_4^2)$). The focal locus of C_S is a *caustic by reflection* in [95], but here we do not require for the light source to be a point.

Example 2.5.7. Let S be the ellipsoid from Example 2.5.5. We first consider a catadioptric camera with mirror S and a pinhole sensor at a point P. Let us start with $P = [1, 0, 0, 3]$. The radical ideal of C'_S (computed as described above) is the intersection of two prime ideals:

$$I_1 = \langle q_{23}, 3q_{13} - q_{34}, 3q_{12} - q_{24} \rangle$$
$$I_2 = \langle q_{14}q_{23} - q_{13}q_{24} + q_{12}q_{34}, 625q_{12}^2 + 625q_{13}^2 + 256q_{14}^2 + 150q_{12}q_{24} + 9q_{24}^2 + 150q_{13}q_{34} + 9q_{34}^2 \rangle.$$

We observe that I_2 is a component of $Ch_{X_{\text{Iso}}}$. It is extraneous for us. More precisely, X_{Iso} contains two quadric curves on S, and I_2 is generated by the Plücker quadric and the Chow form of one of these curves. On the other hand, I_1 is the ideal of C_S. This $(1, 0)$-congruence is the α-plane for $Q = [1, 0, 0, -3]$. The points P and Q are the two foci of the ellipsoid S.

If we choose P randomly, then $C_S = C'_S$. Using the computation explained above, we find that the bidegree of the specular congruence C_S is $(8, 4)$. According to Josse and Pène [95], the focal locus $\mathcal{F}(C_S)$, which is the caustic by reflection of S, is a surface of degree 18.

Finally, we consider the catadioptric camera given by S together with a general $(1, 1)$-congruence (two-slit camera). The resulting specular congruence has bidegree $(12, 6)$. ♢

In closing, we wish to reiterate that the notion of order used in this chapter is always the algebraic one. The "physical" order of a catadioptric camera may be quite a bit lower, due to orientation constraints, with some of the rays reflected inside the body of the mirror.

2.6 Conclusions

We have studied in this chapter a model of geometric cameras that represents an imaging system as a line congruence inside $\text{Gr}(1, \mathbb{P}^3)$. Much of our exposition focused on cameras that are associated with congruences of order one, and that can be used to define rational projection mappings. In particular, we have classified all possible such cameras, and studied their multi-view geometry, by characterizing algebraically n-tuples of lines that

are concurrent at a point. We have also described the geometry of some devices that are associated with congruences of higher order, namely panoramic and catadioptric cameras.

While our treatment may appear as very abstract, we believe that our approach is useful for several reasons. First, our framework can be used to compare and classify general imaging models, and to deduce geometric properties of camera systems that are independent of parameterizations and choices of image coordinates. For example, the degree of "epipolar curves", or the number of "baselines", are features that can be understood using the geometry of congruences, as explained in Section 2.4.2. More concretely, our results on multi-view correspondences can be carried over to image coordinates, and thus provide a methodology for characterizing multi-view correspondences for very general systems of cameras.

Further work could be devoted to non-rational cameras, which we did not study in full generality. For example, it could be useful to gain a better understanding of the action of reflections and refractions on congruences of lines. As discussed in Section 2.5.2, mirrors or lenses act on a congruence C by mapping it to another congruence C'. Although we described a method for computing C' given C, we were not able to establish a general relation between the bidegrees of these congruences.

Finally, we mention another direction that we think is worth pursuing, that is to relate our framework with (some version of) the *plenoptic function* [2]. This function describes the radiance along all rays in a region of three-dimensional space. Mathematically, we can model it as a mapping $\mathcal{F} : \mathrm{Gr}(1, \mathbb{P}^3) \to \mathbb{R}$ or $\mathcal{F} : \mathrm{Gr}(1, \mathbb{P}^3) \to \mathbb{P}^1$.[1] In these terms, taking a picture with any type of camera means restricting $\mathcal{F}$ to the lines of the associated congruence: the output is a two-dimensional photograph of the scene represented by $\mathcal{F}$. This viewpoint also fits nicely with the topics that will be studied in Chapters 4 and 5. In particular, $\mathcal{F}$ must satisfy certain properties in order to be a "valid" plenoptic function representing a scene. For example, if we assume that $\mathcal{F}$ is a discrete function $\mathcal{F} : \mathrm{Gr}(1, \mathbb{P}^3) \to \{0, 1\}$ that produces binary images (silhouettes), then it must describe sets of lines which intersect an object in $\mathbb{P}^3$ (*e.g.*, a surface over the reals). This relates to our study of visual hulls in Chapter 4. Alternatively, we can assume that pictures of surfaces are "outlines", rather than silhouettes. In this case, $\mathcal{F}$ must describe sets of tangents to smooth surfaces. This idea will be important in Chapter 5.

[1] This is actually closer to the "4D light field" [114], since the plenoptic function has a five-dimensional parameter space, that includes the point where a ray is being sampled.

Chapter 3

Photographic Cameras

The geometric cameras studied in Chapter 2 are maps from $\mathbb{P}^3$ to $\mathrm{Gr}(1, \mathbb{P}^3)$, and are independent of image measurements. Physical "photographic" cameras, on the other hand, will always represent image points using coordinates. In this chapter, we model this kind of camera as a rational map $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ or $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$. Since any photographic camera is obtained from a parameterization of a congruence of order one, we can build on the framework of rational geometric cameras to study arbitrary photographic imaging systems and their multi-view geometry.

The material in this chapter is based on the following publications:

– Jean Ponce, Bernd Sturmfels, and Matthew Trager. "Congruences and concurrent lines in multi-view geometry". In: *Advances in Applied Mathematics* 88 (2017), pp. 62–91.

– Matthew Trager, Bernd Sturmfels, John Canny, Martial Hebert, and Jean Ponce. "General models for rational cameras and the case of two-slit projections". In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2017.

3.1 Introduction

Representing cameras as congruences of viewing rays is convenient for studying geometric properties of imaging processes, however, in practice, the use of image coordinates is very important, and is essential for the design of SfM algorithms. For this reason, we extend the geometric camera model considered in the previous chapter to include image measurements and coordinates. More precisely, we define a "photographic" camera to be a rational map $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ or $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ with the property that the pre-image of a (generic) point is a line. Any such camera is obtained by composing a rational camera $\mathbb{P}^3 \dashrightarrow C$, where C is an order-one congruence in $\mathrm{Gr}(1, \mathbb{P}^3)$, with a (birational)

"parameterization" $C \dashrightarrow \mathbb{P}^2$ or $C \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$. Thus, our photographic camera model is an analytical version of the geometric rational camera model from Chapter 2.

A natural strategy for obtaining a parameterization $C \dashrightarrow \mathbb{P}^2$ of a congruence C is to use a retinal plane H, equipped with a system of homogeneous coordinates. This allows us to map a generic line in C to the coordinates of its point of intersection with H. This strategy models the geometry of a physical device (which could record an image using a plane of photosensitive sensors), and generalizes the coordinate mapping used in pinhole cameras. However, it is also convenient to consider more general parameterizations, and particularly mappings $C \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$. Indeed, many $(1, \beta)$-congruences are naturally parameterized by $\mathbb{P}^1 \times \mathbb{P}^1$, since their focal locus consists of a line and a rational curve. For these congruences, a photographic camera $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ leads to simpler expressions. Moreover, many non-central cameras used in practice produce photographs using a 1D scanner moving along a curve, or by sampling columns or rows from sequences of 2D pictures [173]. In these cases, the projectivized image is best identified with $\mathbb{P}^1 \times \mathbb{P}^1$ (one component parameterizes "time", and the other component represents a 1D measurement).

The fact that we only consider rational (polynomial) projections allows us to use Theorem 2.3.3 to classify all photographic cameras based on the "type" of the associated congruence of order one. Moreover, we can apply the formulae for geometric rational cameras in Plücker coordinates from previous chapter to construct explicit examples of photographic cameras. The inverse of any parameterization $C \dashrightarrow \mathbb{P}^2$ or $C \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ also immediately provides the "inverse line projection", associating image points with viewing lines. This can be used together with our description of the concurrent lines variety to recover multi-view constraints for correspondences in image coordinates.

After some general considerations on photographic cameras, we present a detailed analysis of *two-slit cameras*. It has previously been noted that a two-slit projection onto a retinal plane leads to a map $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ whose coordinates are quadratic expressions [55, 201]. On the other hand, the same mapping can be described using a pair of linear projections $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1$, that is, using a bilinear map $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$. This simple bilinear form is very convenient, and allows us for example to analyze orbits of two-slit cameras under affine, similarity, and euclidean transformations, and to introduce intrinsic parameters and calibration matrices that closely resemble the traditional ones defined for pinhole cameras. Moreover, point correspondences among two views are characterized by a $2 \times 2 \times 2 \times 2$ *epipolar tensor*, and we can completely describe the relationship between such tensors and camera configurations (each tensor encodes *two* configurations of two-slit cameras). We propose "linear" and "minimal" reconstruction algorithms, and discuss strategies for self-calibration for two-slit cameras.

The study of two-slit cameras is relatively simple because of the bilinear representation, however one could develop a more general theory of photographic cameras, that could be seen as an algebraic, non-linear version of classical multi-view geometry. We do not

pursue this direction in detail here, but we describe some simple methods for manipulating photographic cameras, explaining for example how to compute forward and inverse projections of lines and points. Furthermore, we apply "multi-graded Chow forms" [131] to show the existence of *trifocal* and *quadrifocal forms*, that characterize matching points and lines in different images, and directly generalize traditional multi-focal tensors for arbitrary photographic cameras.

3.1.1 Previous work

The popularity of the pinhole projection model is due to its simplicity, and to the fact that it naturally mimics the geometry of the (human) eye. However, in recent years there has been growing interest in alternative imaging systems, that often have practical advantages. For example, compared to a pinhole camera, a non-central imaging device can collect light in a more "distributed" way, so that a single photograph can provide more information about a scene [152, 165]. A broad overview of non-standard imaging technologies, with many pointers to the literature, can be found in the monograph by Sturm, Ramalingam, Tardif, Gasparini and Barreto [173]. In that work, the authors make the distinction between "global", "local", and "discrete" models, in which each of the camera's parameters affects, respectively, the whole projection function, a local "patch" of the field of view, or a single ray recorded by the camera. From this viewpoint, we only consider global camera models in this chapter.

Among the most common techniques for generating non-perspective images is to use a moving "slit" that captures 1D images. This includes pushbroom cameras [72], but also more general "non-central panoramic" systems, which sample columns of pixels from 2D photographs recorded with a moving (traditional) camera [200, 89]. In this setting, we mention that by sampling *two* columns of pixels it is also possible to capture "stereo panoramas" from a single rotating camera [137, 165]. In our terminology, the viewing rays of this type of device form a line congruence of order two (see Section 2.5.1). It has also been observed that sampling different columns of pixels varying together with the camera location is equivalent to a *two-slit projection* model [201]. As noted by Pajdla [135], two-slit projection includes the pushbroom camera model, which corresponds to one of the two slits lying at infinity. In general, the analytical form of a two-slit projection is a quadratic map $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ [201]. Feldman, Pajdla and Weinshall used this representation to study epipolar geometry of two-slit cameras [55]. Their description is based on a 6×6 "fundamental matrix", that corresponds to a bilinear constraint on the quadratic Veronese embedding $\mathbb{P}^2 \hookrightarrow \mathbb{P}^5$ of image coordinates.

A unified representation for several types of non-central cameras is the "General Linear Camera" (GLC) model by Yu and McMillan [197]. This model is based on a two-plane parameterization of lines, and can be used to represent pinhole, two-slit, and linear oblique [136] cameras. As argued by Ponce [145], all these classes of cameras can also be understood using *linear congruences* in $\mathrm{Gr}(1, \mathbb{P}^3)$. In subsequent work [11], Batog,

Goaoc, and Ponce investigated in more detail these linear imaging systems (which are actually represented by quadratic expressions in 3D coordinates: here "linear" refers to the geometry of the line congruence). They explained, for example, the role of the retinal plane in the imaging process, and argued that the choice of this plane is not completely irrelevant, as is the case for pinhole cameras. They also described epipolar geometry for these cameras using a 6×6 fundamental matrix, or using a 4×4 fundamental matrix assuming appropriate "normalized" image coordinates. These special image coordinates will also play a role in our study of two-slit cameras. We also mention that our $2\times2\times2\times2$ epipolar tensor can be seen as an instance of "Grassmann tensors", which were studied by Hartley and Schaffalitzky [77], and is also formally identical the radial quadrifocal tensor, proposed by Thirtala and Pollefeys for cameras with radial distortion [181]. Here we reintroduce this type of tensors in the context of two-slit cameras.

Finally, we mention that Hartley and Saxena introduced in [76] a "cubic rational polynomial" camera model that is similar to the rational photographic cameras that are studied in this chapter (their model is essentially a bicubic map $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$). The motivation for this type of camera was to provide an accurate approximation for complex imaging systems, particularly those used in satellite imagery, whose real projection mappings are often non-rational. Hartley and Saxena presented an algorithm for estimating parameters of their model given world-to-image correspondences, and also discussed several strategies for dealing with noise and over-fitting issues. Interestingly, this model has since become a standard in photogrammetry, and is commonly used in commercial satellite sensors (see for example [88, 178, 199]). A different application of a similar model was proposed by Claus and Fitzgibbon [35] to approximate distortion effects in wide-angle lenses. The main difference between these rational mappings and our photographic cameras is that we only consider projections that associate image points with *lines*. Rational maps with arbitrary coefficients, such as those used in [76] and [35], will not have this property, since the pre-image of a point will typically be an algebraic curve in space: from our strict perspective, these are not models of "cameras".

3.1.2 Main contributions

The main contributions of this chapter can be summarized as follows.

- We study *photographic cameras*, that are rational mappings $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ or $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ with the property that the pre-image of generic points are lines in space (Section 3.2). By parameterizing line congruences of order one, we derive explicit representations for all such cameras.
- We present a study of *two-slit cameras* [11, 135, 201], that we model as bilinear maps $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$. We explain the relationship between this model and the traditional description of these cameras as quadratic maps $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ (Section 3.3.1). We then use the bilinear representation to introduce a pair of upper-triangular *cal-*

ibration matrices that can be identified with the orbits of two-slit cameras under similarity transforms (Section 3.3.2). We also introduce a $2 \times 2 \times 2 \times 2$ *epipolar tensor*, that encodes the projective geometry of a pair of two-slit cameras, and generalizes the traditional fundamental matrix. We use these results to propose algorithms for structure from motion and self-calibration (Section 3.3.3).

- We illustrate some properties of general photographic cameras: for example, we characterize multi-view point correspondences using a generalized version of the joint image variety (Section 3.4.1), and explain how to compute forward and inverse projections of lines and points (Section 3.4.2). We also show the existence of *multi-focal forms*, that characterize corresponding points and lines in three and four views, and directly generalize classical trifocal and quadrifocal tensors in the case of photographic cameras (Section 3.4.3).

Conventions. We return to the notation used in Chapter 1: we write $\mathbb{P}^n = \mathbb{P}(\mathbb{R}^{n+1})$ for the n-dimensional projective space over the *real* numbers, and use bold and normal font to distinguish between vectors and projective objects (*e.g.*, we write $x = [\mathbf{x}]$ for a point in $\mathbb{P}^n$). If $\mathbf{p} = (p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34})$ is a vector representing Plücker coordinates of a line in $\mathbb{P}^3$, then we write $\mathbf{p}^* = (p_{34}, -p_{24}, p_{23}, p_{14}, -p_{13}, p_{12})$ for the vector that represents the same line in dual coordinates (or, equivalently, that represents the dual of that line in primal coordinates).

3.2 Congruences and Coordinates

This section introduces our model for *photographic cameras*. After giving some preliminary definitions (Section 3.2.1), we apply our study of congruences of order one from the previous chapter to derive explicit algebraic representations for photographic cameras (Sections 3.2.2 and 3.2.3).

3.2.1 Basic definitions

We are interested in photographic devices that can be described using rational maps from $\mathbb{P}^3$ into $\mathbb{P}^2$ or $\mathbb{P}^1 \times \mathbb{P}^1$. The following is a precise definition for our model. We recall that a rational map between projective spaces is a map whose coordinates are homogeneous polynomial functions. Such a map is called "dominant" if the closure of its image is the whole codomain (*i.e.*, the map is surjective up to closure).

Definition 3.2.1. A *photographic rational camera* is a dominant rational map $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ or $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ with the property that the fiber (pre-image) of a generic point is a line in $\mathbb{P}^3$.

This clearly extends the traditional pinhole camera model studied in Chapter 1. As explained in the Introduction, we allow for the image to be the projective plane $\mathbb{P}^2$ as well as a product of projective lines $\mathbb{P}^1 \times \mathbb{P}^1$, since the latter is more natural when the imaging system consists of a moving 1D sensor, or when images are formed from 1D samples in sequences of images, as in [152, 201], for example. Moreover, we will see shortly that rational cameras associated with congruences of higher order are more easily viewed as maps into $\mathbb{P}^1 \times \mathbb{P}^1$.

Concretely, we can describe a photographic camera using a triplet $[f_1, f_2, f_3]$ of homogeneous polynomials in $\mathbb{R}[x_1, x_2, x_3, x_4]$ of the same degree, or by two such pairs $([g_1, g_2], [h_1, h_2])$. The projection maps are given by

$$\begin{aligned} &Q_1 : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2, \ [\mathbf{x}] \mapsto [f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x})], \quad \text{or} \\ &Q_2 : \mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1, \ [\mathbf{x}] \mapsto ([g_1(\mathbf{x}), g_2(\mathbf{x})], [h_1(\mathbf{x}), h_2(\mathbf{x})]) \, . \end{aligned} \tag{3.1}$$

We will indicate these mappings as $Q_1 = [f_1, f_2, f_3]$ and $Q_2 = ([g_1, g_2], [h_1, h_2])$. The *base locus* of a rational camera is the set points in $\mathbb{P}^3$ where the projection mapping is not defined. This means that the base locus of Q_1 and Q_2 as above is given by $V(f_1, f_2, f_3)$ and $V(g_1, g_2) \cup V(h_1, h_2)$ respectively. For a pinhole camera, the base locus is simply the pinhole. The *degree* of a rational camera is the degree of the f_i, or the pair of degrees of the g_i and h_i (which in general are different).

It is worth noting that the polynomials in (3.1) cannot be arbitrary, because of the assumption that the fibers must lines. More concretely, to describe the pre-image a general image point $u = [u_1, u_2, u_3]$ in $\mathbb{P}^2$ for a camera $Q_1 = [f_1, f_2, f_3]$, we consider the locus of $x = [\mathbf{x}]$ in $\mathbb{P}^3$ such that

$$\operatorname{rank} \begin{bmatrix} f_1(\mathbf{x}) & u_1 \\ f_2(\mathbf{x}) & u_2 \\ f_3(\mathbf{x}) & u_3 \end{bmatrix} = 1. \tag{3.2}$$

For generic polynomials f_1, f_2, f_3, this describes a curve in $\mathbb{P}^3$ that contains the base locus $V(f_1, f_2, f_3)$. In order for Q_1 to be a camera, we require for this curve be the union of $V(f_1, f_2, f_3)$ and a line. Analogous considerations hold for a camera $Q_2 = ([g_1, g_2], [h_1, h_2])$, for which we take the locus of $x = [\mathbf{x}]$ in $\mathbb{P}^3$ such that

$$\operatorname{rank} \begin{bmatrix} g_1(\mathbf{x}) & u_1 \\ g_2(\mathbf{x}) & u_2 \end{bmatrix} = 1 \text{ and } \operatorname{rank} \begin{bmatrix} h_1(\mathbf{x}) & v_1 \\ h_2(\mathbf{x}) & v_2 \end{bmatrix} = 1.$$

Example 3.2.2. If three quadratic forms f_1, f_2, f_3 are such that $V(f_1, f_2, f_3)$ is a twisted cubic, then $Q = [f_1, f_2, f_3]$ is a camera. Indeed (3.2) is a curve of degree four, so removing $V(f_1, f_2, f_3)$ leaves a line. For three quadratic forms to have this property they must

be expressible as the 2×2-minors of a 2×3-matrix of linear forms in x_1, x_2, x_3, x_4. For example, the map

$$x \mapsto [x_1x_3 - x_2^2, x_1x_4 - x_2x_3, x_2x_4 - x_3^2] \tag{3.3}$$

is a photographic camera. ♢

In general, characterizing which triples f_1, f_2, f_3 or quadruples g_1, g_2, h_1, h_2 of polynomials correspond rational photographic cameras is a challenging problem. We address this issue by using of our study of line congruences from Chapter 2.

3.2.2 From lines to measurements

To construct examples of rational cameras, we will use our classification of congruences from the previous chapter. A photographic rational camera $Q_1 : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ or $Q_2 : \mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ is in fact uniquely associated with a congruence of order one, that is the closure of the image of the map

$$L_{Q_1} : \mathbb{P}^2 \dashrightarrow \mathrm{Gr}(1, \mathbb{P}^3), \text{ or } L_{Q_2} : \mathbb{P}^1 \times \mathbb{P}^1 \dashrightarrow \mathrm{Gr}(1, \mathbb{P}^3),$$

associating image points with their fiber. The image of L_{Q_1} or L_{Q_2} is a congruence of order one since a general point x in $\mathbb{P}^3$ will be contained in exactly one image line (that is the set of points in $\mathbb{P}^3$ whose image under Q_1 or Q_2 is the same as for x). Conversely, given any $(1, \beta)$-congruence C, we may obtain a rational camera by composing the geometric camera $C : \mathbb{P}^3 \dashrightarrow \mathrm{Gr}(1, \mathbb{P}^3)$ with a birational "parameterization" $C \dashrightarrow \mathbb{P}^2$ or $C \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$. Here "birational" means that these maps are rational and (generically) invertible, and that their inverse is also rational.

In summary, any rational photographic camera $Q : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ or $Q : \mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ can be seen as an analytical representation of a rational geometric camera $C : \mathbb{P}^3 \dashrightarrow \mathrm{Gr}(1, \mathbb{P}^3)$, which is in turn defined by line congruence of order one. This is analogous to relationship between a pinhole projection $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ and the corresponding bundle of lines. In particular, using the terminology from Chapter 2, the *base locus* of a photographic camera Q, where the projection mapping is not defined, will always contain the *focal locus* of the associated congruence C (possibly strictly, as we will see shortly). Both the base locus and the focal locus can be seen as generalizations of the pinhole in traditional cameras. We will also say that photographic camera has *class* β if the associated congruence has bidegree $(1, \beta)$.

In order to construct parameterizations $C \dashrightarrow \mathbb{P}^2$ or $C \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ we will usually follow one of two strategies:

1. A natural approach for defining a map $C \dashrightarrow \mathbb{P}^2$ is to fix a "retinal plane" H in $\mathbb{P}^3$, equipped with a system of homogeneous coordinates. This allows us to associate with a general line ℓ in C the coordinates of the point of intersection $\ell \wedge H$ on H

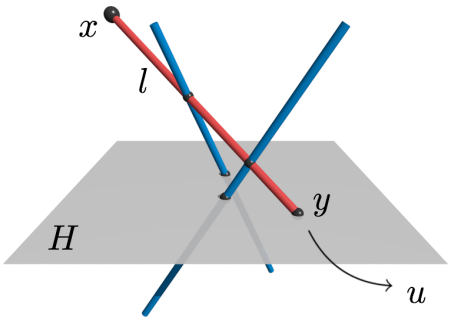


Figure 3.1: A photographic camera (in the figure, a two-slit camera) can be obtained by associating a scene point x with a visual ray l, then mapping the ray l to its intersection y with some retinal plane H, and finally using a projective coordinate system on H to express y as a point u in $\mathbb{P}^2$.

(see Figure 3.1). This coincides of course with the usual coordinate mapping used for pinhole cameras.

2. The second strategy for constructing a birational map $C \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ applies when the lines in C are naturally parametrized by pairs of points in $\mathbb{P}^1$. This happens for example when lines in C are characterized as the common transversals of two space curves (or lines).

We note that the first strategy can be made more explicit as follows. Let H is a plane in $\mathbb{P}^3$ defined parametrically by

$$\mathbb{P}^2 \to H, \quad [u_1, u_2, u_3] \mapsto [u_1\mathbf{y}_1 + u_2\mathbf{y}_2 + u_3\mathbf{y}_3] \tag{3.4}$$

where $y_1 = [\mathbf{y}_1]$, $y_2 = [\mathbf{y}_2]$, $y_3 = [\mathbf{y}_3]$ are points on H. In other words, we view the columns of the 4×3 matrix $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3)$ as vectors forming a normalized basis inducing a system of homogeneous coordinates on H. We can then define a rational map $\mathrm{Gr}(1, \mathbb{P}^3) \dashrightarrow \mathbb{P}^2$ associating a generic line ℓ with the coordinates of the point $\ell \wedge H$ in H. This map is *linear* in Plücker coordinates, and is described by the 3×6 matrix

$$\mathbf{N}_\mathbf{Y} = \begin{bmatrix} (\mathbf{y}_2 \vee \mathbf{y}_3)^* \\ (\mathbf{y}_3 \vee \mathbf{y}_1)^* \\ (\mathbf{y}_1 \vee \mathbf{y}_2)^* \end{bmatrix}. \tag{3.5}$$

In other words, the matrix $\mathbf{N}_\mathbf{Y}$ induces a linear projection $\mathbb{P}^5 \dashrightarrow \mathbb{P}^2$, with the property that a generic line ℓ in $\mathbb{P}^3$ with Plücker coordinates $[\mathbf{p}]$ intersects the plane H at the point with coordinates $u = [\mathbf{u}] = [\mathbf{N}_\mathbf{Y}\mathbf{p}]$ in $\mathbb{P}^2$. As for pinhole cameras, the parameterized retinal plane H can be used to model a grid of photosensitive sensors that record photographs.

Note that by composing a geometric camera C with a linear projection $\mathbb{P}^5 \dashrightarrow \mathbb{P}^2$ given by (3.5), we obtain a rational map Q that is not defined on the lines in C that lie on H. There are in general β such lines (the class of C), that will belong to the base locus of the photographic camera Q, but not to the focal locus of the congruence C.

Example 3.2.3. Consider the plane $H = \{x_3 - x_4 = 0\}$ in $\mathbb{P}^3$, equipped with the reference frame induced by the normalized basis $\mathbf{y}_1 = (1,0,0,0)^T$, $\mathbf{y}_2 = (0,1,0,0)^T$, $\mathbf{y}_3 = (0,0,1,1)^T$. The matrix $\mathbf{N_Y}$ takes in this case the form

$$\mathbf{N_Y} = \begin{bmatrix} 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{3.6}$$

The null-space of the corresponding linear projection is $V(p_{13}+p_{14}, p_{42}+p_{23}, p_{34})$, which characterizes lines contained in the plane H. ♢

Remark 3.2.4. The construction described above can in principle be generalized to curved "retinal surfaces". More precisely, given a rational surface S in $\mathbb{P}^3$ together with an explicit parameterization $\varphi : \mathbb{P}^2 \dashrightarrow S$, we can define, at least locally, an algebraic map from $\mathrm{Gr}(1,\mathbb{P}^3)$ to $\mathbb{P}^2$ that associates a line with the coordinates of one of its intersections with S. By composing this map with geometric cameras from Chapter 2, we can construct new analytical camera mappings. Note however that these are not photographic cameras in the sense of Definition 3.2.1, because the resulting expressions will in general be algebraic and not rational (*i.e.*, they will involve square roots). Nevertheless, this could be a good mathematical model for many optical devices (*e.g.*, compound eyes of insects).

3.2.3 Examples of photographic cameras

We next use our study of congruences from Chapter 2 together with the parameterizations discussed above to construct examples of photographic rational cameras. In particular, we consider the four types of congruences described by Theorem 2.3.3. Computations in this section were carried out in Sage [158].

Type 1: A congruence C of type 1 is an α-plane, *i.e.*, a bundle of lines that meet at a point c. The classical analytical model for a pinhole camera as a projection matrix can be derived by composing a geometric camera of this type with the map $\mathrm{Gr}(1,\mathbb{P}^3) \dashrightarrow \mathbb{P}^2$ defined by the 3×6 matrix $\mathbf{N_Y}$ from (3.5), for an arbitrary plane H and normalized basis $\mathbf{Y}$ on H. Indeed, it is easy to see that

$$\mathbf{x} \mapsto \mathbf{N_Y}(\mathbf{c} \vee \mathbf{x}) = \begin{bmatrix} (\mathbf{y}_2 \vee \mathbf{y}_3 \vee \mathbf{c})^* \\ (\mathbf{y}_3 \vee \mathbf{y}_1 \vee \mathbf{c})^* \\ (\mathbf{y}_1 \vee \mathbf{y}_2 \vee \mathbf{c})^* \end{bmatrix} \mathbf{x},$$

which represents the usual projection matrix. On the other hand, by considering different parameterizations of C, it is also possible to construct "alternative" pinhole cameras, which differ only analytically from the traditional linear projection model. For example, a bilinear map $[\mathbf{x}] \mapsto (\mathbf{A}_1\mathbf{x}, \mathbf{A}_2\mathbf{x})$ where $\mathbf{A}_1, \mathbf{A}_2$ are 2×4-matrices whose null-spaces meet is, geometrically speaking, a pinhole camera. Any pinhole model can be obtained from a traditional one by applying a "non-linear" change of image coordinates $\mathbb{P}^2 \dashrightarrow \mathbb{P}^2$ or $\mathbb{P}^2 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$.

Example 3.2.5. If $[\mathbf{x}] = [x_1, x_2, x_3, x_4]$, the following three maps

$$\begin{array}{lll} \mathbb{P}^3 \dashrightarrow \mathbb{P}^2, & [\mathbf{x}] \mapsto [x_1, x_2, x_3], \\ \mathbb{P}^3 \dashrightarrow \mathbb{P}^2, & [\mathbf{x}] \mapsto [x_2x_3,\ x_1x_3,\ x_1x_2], \\ \mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1, & [\mathbf{x}] \mapsto ([x_1, x_2], [x_1 + x_2, x_3])), \end{array} \tag{3.7}$$

represent the same geometric pinhole camera with center $c = [0, 0, 0, 1]$. ♢

Type 2: A geometric camera of type 2 records lines which are secants of a twisted cubic in $\mathbb{P}^3$. An example of a photographic camera of this type can be obtained by composing the geometric camera from (2.9) with the map $\mathrm{Gr}(1, \mathbb{P}^3) \dashrightarrow \mathbb{P}^2$ defined by the 3×6 matrix $\mathbf{N_Y}$ from (3.5). Note that this results in a photographic rational map where the coordinate polynomials have degree 4.

A different approach for constructing this type photographic camera is illustrated in Example 3.2.2: if three quadratic forms f_1, f_2, f_3 are expressible as the 2×2-minors of a 2×3-matrix of linear forms in x_1, x_2, x_3, x_4, then $V(f_1, f_2, f_3)$ is a twisted cubic, and $Q = [f_1, f_2, f_3]$ is a photographic camera of type 2. Note that in this case the coordinate polynomials have degree 2.

Example 3.2.6. Applying the 3×6 projection matrix $\mathbf{N_Y}$ from (3.6) to the geometric camera (2.9) we obtain the map

$$Q(x) = \begin{bmatrix} -x_2^3x_3 + x_1x_2x_3^2 + x_1x_3^3 + x_1x_2^2x_4 + x_2^3x_4 - x_1^2x_3x_4 - 3x_1x_2x_3x_4 + x_1^2x_4^2 \\ -x_2^2x_3^2 + x_1x_3^3 + x_2x_3^3 + x_2^3x_4 - x_1x_2x_3x_4 - x_2^2x_3x_4 - x_1x_3^2x_4 + x_1x_2x_4^2 \\ x_3^4 - 2x_2x_3^2x_4 + x_2^2x_4^2 \end{bmatrix}.$$

This describes a photographic camera whose focal locus $\mathcal{F}$ is the standard twisted cubic (the image of $[s, t] \mapsto [s^3, s^2t, st^2, t^3]$), and where the retinal plane $H = \{x_3 - x_4 = 0\}$ has a coordinate frame defined by the basis $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 1)$. The base locus of $Q(x)$ (where the projection is not defined) consists of X together with the two lines $L_1 = V(x_3, x_2)$ and $L_2 = V(x_2 - x_3, x_1 - x_3)$: these are in fact the lines of the congruence that lie on H. We point out that are only two lines rather than the expected number $\beta = 3$ because H is tangent to $\mathcal{F}$ at the point $[1, 0, 0, 0]$.

Another photographic camera associated with the same congruence is given more simply by

$$Q'(x) = \begin{bmatrix} x_3^2 - x_2x_4 \\ x_2x_3 - x_1x_4 \\ x_2^2 - x_1x_3 \end{bmatrix}. \tag{3.8}$$

The base locus for $Q(x)'$ contains only the twisted cubic $\mathcal{F}$: this coincides with the focal locus of the congruence. ♢

Type 3: A congruence C of type 3 contains a family of lines that intersect both a fixed line L and a rational curve X of degree β meets intersects L in $\beta - 1$ points (counted with multiplicity). As argued in Section 2.3.2, if L is the line $V(x_1, x_2)$, the curve X can be represented parametrically as

$$\mathbb{P}^1 \to X, \qquad [s,t] \mapsto [sf(s,t), tf(s,t), g(s,t), h(s,t)], \tag{3.9}$$

where f, g and h are arbitrary binary forms of degree $\beta - 1$, β and β respectively. The associated geometric camera $\mathbb{P}^3 \to \mathrm{Gr}(1, \mathbb{P}^3)$ is described by (Proposition 2.3.5)

$$C(x) \;=\; \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \vee \begin{bmatrix} x_1 f(x_1,x_2) \\ x_2 f(x_1,x_2) \\ g(x_1,x_2) \\ h(x_1,x_2) \end{bmatrix}. \tag{3.10}$$

Using this expression, we can construct an associated photographic camera by composing with the map $\mathrm{Gr}(1, \mathbb{P}^3) \dashrightarrow \mathbb{P}^2$ defined by $\mathbf{N_Y}$ as above. However, another (arguably simpler) strategy is instead to construct a map $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ that associates a point x with the coordinates of the points y_1 and y_2 where $C(x)$ meets X and L respectively. Here we parameterize X as in (3.9) and L by $[s,t] \mapsto [0,0,s,t]$. A simple calculation using (3.10) shows that

$$\begin{aligned} y_1 &= [x_1 f(x_1,x_2), x_1 f(x_1,x_2), g(x_1,x_2), h(x_1,x_2)], \\ y_2 &= [0, 0, g(x_1,x_2) - f(x_1,x_2)x_3, h(x_1,x_2) - f(x_1,x_2)x_4], \end{aligned}$$

so the resulting photographic camera $Q : \mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ is given by

$$Q(x) = \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} g(x_1,x_2) - f(x_1,x_2)x_3 \\ h(x_1,x_2) - f(x_1,x_2)x_4 \end{bmatrix} \right). \tag{3.11}$$

General photographic cameras $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ can be obtained by composing this expression with a birational map $\mathbb{P}^1 \times \mathbb{P}^1 \dashrightarrow \mathbb{P}^2$.

Example 3.2.7. The map

$$Q(x) = \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} x_1^2 + x_2^2 - x_1 x_3 \\ x_1 x_2 - x_1 x_4 \end{bmatrix} \right)$$

is a photographic camera $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ of type 3 with $\beta = 2$. It corresponds to the congruence of lines intersecting $L = V(x_1, x_2)$ and the conic $X(s,t) = [s^2, st, s^2 + t^2, st]$. These expressions can be obtained from (3.9) and (3.11) using $f(s,t) = s$, $g(s,t) = s^2 + t^2$ and $h(s,t) = st$.

A photographic camera $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ associated with the same congruence is given by the composition of the geometric camera (3.10) and the map $\mathrm{Gr}(1, \mathbb{P}^3) \dashrightarrow \mathbb{P}^2$ defined by the matrix $\mathbf{N_Y}$ from (3.6). The resulting expression is

$$Q'(x) = \begin{bmatrix} -x_1^3 + x_1^2 x_2 - x_1 x_2^2 + x_1 x_2 x_3 - x_1 x_2 x_4 \\ -x_1^2 x_2 + x_1 x_2^2 - x_2^3 + x_2^2 x_3 - x_2^2 x_4 \\ x_1 x_2 x_3 - x_1^2 x_4 - x_2^2 x_4 \end{bmatrix}.$$

♢

Type 4: A congruence C of type 4 can be viewed as a degeneration of a congruence of type 3, and the associated geometric map can be represented as (3.9) assuming the binary form f is identically zero. In this case, because the curve X and the line L coincide, we cannot map a point x in $\mathbb{P}^3$ to the coordinates of the two intersection points y_1, y_2 (in fact, the image (3.11) becomes one-dimensional in this case). However, we can construct photographic maps by composing (3.9) with the usual map $\mathrm{Gr}(1, \mathbb{P}^3) \dashrightarrow \mathbb{P}^2$ defined by the matrix $\mathbf{N_Y}$.

Example 3.2.8. An example of photographic camera $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ of type 4 is

$$Q(x) = \begin{bmatrix} -x_1^3 + x_1^2 x_2 - x_1 x_2^2 \\ -x_1^2 x_2 + x_1 x_2^2 - x_2^3 \\ x_1 x_2 x_3 - x_1^2 x_4 - x_2^2 x_4 \end{bmatrix}.$$

This expression is obtained from the geometric camera (3.9) with $g(s,t) = s^2 + t^2$ and $h(s,t) = st$, composed with the map $\mathrm{Gr}(1, \mathbb{P}^3) \dashrightarrow \mathbb{P}^2$ defined by $\mathbf{N_Y}$ from (3.6). ♢

3.3 Two-Slit Cameras

In this section, we focus on a particular class of photographic rational cameras, namely *two-slit cameras* [55, 72, 135, 201]. This type of camera is associated with congruences of class $\beta = 1$, and it can be viewed as the "simplest" rational camera model after pinhole

projections. Photographic two-slit cameras can in fact be conveniently represented using a bilinear model

$$\mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1, \qquad [\mathbf{x}] \mapsto ([\mathbf{A}_1\mathbf{x}], [\mathbf{A}_2\mathbf{x}]),$$

where $\mathbf{A}_1, \mathbf{A}_2$ are 2×4-matrices of full rank with disjoint null spaces. These null spaces identify the two slits of the cameras. We view each $\mathbb{P}^1$ in the image as parameterizing the planes that contain one of two slits.

We begin our discussion by explaining in Section 3.3.1 the relationship between the bilinear representation and the quadratic mapping $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ that has been previously used to model two-slit cameras [55, 201]. We then classify two-slit cameras based on their orbits under affine, similarity and euclidean transformations (Section 3.3.2). Our analysis leads to the definition of upper triangular *calibration matrices*, whose entries are invariant to similarity orbits. In Section 3.3.3, we describe two-view geometry for two-slit cameras, pointing out the existence of a $2 \times 2 \times 2 \times 2$ "epipolar tensor" that shares many similarities with the traditional fundamental matrix for pinhole cameras (although a tensor represents in general *two* configurations of cameras). We also present some algorithms for reconstruction and self-calibration based on these ideas.

3.3.1 Retinal planes and bilinear maps

Two-slit cameras record viewing rays that are the transversals to two fixed skew lines (the slits). The bidegree of the associated congruence is $(1, 1)$, since a generic plane will contain exactly one transversal to the slits, namely the line joining the points where the slits intersect the plane. As explained in Chapter 2, this type of congruence is the intersection of $\mathrm{Gr}(1, \mathbb{P}^3)$ with a 3-dimensional linear space in $\mathbb{P}^5$. The geometric camera $C : \mathbb{P}^3 \dashrightarrow \mathrm{Gr}(1, \mathbb{P}^3)$ associated with slits l_1, l_2 was described in (2.8) as a Plücker matrix

$$C(x) \;=\; (x \vee l_1) \wedge (x \vee l_2) \;=\; [\mathbf{P}_1\mathbf{x}\mathbf{x}^T\mathbf{P}_2 - \mathbf{P}_2\mathbf{x}\mathbf{x}^T\mathbf{P}_1], \tag{3.12}$$

where $\mathbf{P}_1, \mathbf{P}_2$ are the Plücker matrices associated with the slits l_1, l_2.

We now assume that we have fix a retinal plane H with the coordinate system defined by $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3)$. A rational camera $Q : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ is obtained by composing (3.12) with the 3×6 matrix $\mathbf{N}_\mathbf{Y}$ as in (3.5). A simple calculation shows that the resulting map can be written compactly as

$$Q(x) = \begin{bmatrix} \mathbf{x}^T\mathbf{P}_1\mathbf{S}_1^*\mathbf{P}_2\mathbf{x} \\ \mathbf{x}^T\mathbf{P}_1\mathbf{S}_2^*\mathbf{P}_2\mathbf{x} \\ \mathbf{x}^T\mathbf{P}_1\mathbf{S}_3^*\mathbf{P}_2\mathbf{x} \end{bmatrix}, \tag{3.13}$$

where $\mathbf{S}_1^*, \mathbf{S}_2^*, \mathbf{S}_3^*$ are dual Plücker matrices for $[\mathbf{y}_2 \vee \mathbf{y}_3]$, $[\mathbf{y}_3 \vee \mathbf{y}_1]$, $[\mathbf{y}_1 \vee \mathbf{y}_2]$ respectively [55, 201].

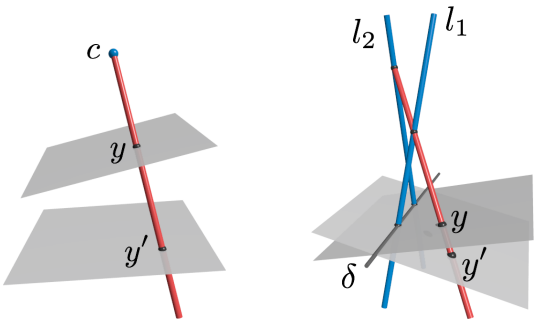


Figure 3.2: *Left:* A pinhole c induces a homography between any two retinal planes not containing c. *Right:* Two skew lines l_1, l_2 induce a homography between planes intersecting at a transversal δ to l_1, l_2.

Example 3.3.1. Let us fix the slits to be the lines $l_1 = V(x_1, x_3), l_2 = V(x_2, x_3 + x_4)$. The corresponding geometric map $\mathbb{P}^3 \dashrightarrow \mathrm{Gr}(1, \mathbb{P}^3)$ is given by

$$C(x) = \begin{bmatrix} x_1x_2 \\ 0 \\ x_1x_3 + x_1x_4 \\ -x_2x_3 \\ x_2x_3 \\ x_3^2 + x_3x_4 \end{bmatrix}. \tag{3.14}$$

Composing with the projection $\mathbf{N_Y}$ in Example 3.2.3, we obtain the formula for a rational camera with slits $\mathbf{l}_1, \mathbf{l}_2$:

$$Q(x) = \begin{bmatrix} x_1x_3 + x_1x_4 \\ 2x_2x_3 \\ x_3^2 + x_3x_4 \end{bmatrix}. \tag{3.15}$$

The same expression can be deduced using (3.13). ♢

It was noted in [55, 201] that using a different retinal plane H in (3.13) corresponds in general to composing $Q : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ with a quadratic change of image coordinates $\mathbb{P}^2 \dashrightarrow \mathbb{P}^2$. However, this transformation is in fact linear when H and H' intersect along a transversal to the slits. This follows from the following property.

Lemma 3.3.2. *Let l_1, l_2 be two skew lines in $\mathbb{P}^3$. For any point x not on the these lines, we indicate with $C(x)$ the unique transversal to l_1, l_2 passing through x. If H and H' are two planes meeting at a line δ that intersects l_1 and l_2, then the map $f : H \dashrightarrow H'$ defined, for points y not on δ, as*

$$f(y) = C(y) \wedge H', \tag{3.16}$$

can be extended to a homography between H and H'.

Proof. The result can be shown by fixing (convenient) coordinate frames on H and H', and using them to express f analytically: the result is a linear map.

We also sketch a more "geometric" argument. We need to show that a generic line m on H is mapped by (3.16) to a line on H'. If m does not intersect l_1 or l_2, the union of the common transversals to l_1, l_2, m is a quadric in $\mathbb{P}^3$. The intersection of this quadric with H' will be a plane conic. However, this intersection contains the transversal line δ, and hence it is a degenerate conic, *i.e.*, a union of two lines. Since δ does not belong to the image of (3.16), we deduce that (the closure of) image of m is a line in H'. □

This lemma also implies that two retinal planes that intersect at a transversal to the slits can define the same rational camera (using appropriate coordinate systems). Note the similarity with the traditional pinhole case, where the choice of the retinal plane is completely irrelevant since the pinhole c induces a homography between any planes H, H' not containing c. See Figure 3.2.

Contrary to the case of pinhole cameras, two-slit cameras of the form (3.13) are *not* all projectively equivalent. This can be argued geometrically by noting that the coordinates u_1, u_2 in $\mathbb{P}^2$ of the points $y_1 = l_1 \wedge H$, $y_2 = l_2 \wedge H$ (the intersections of the slits with the retinal plane) are always preserved by projective transformations of $\mathbb{P}^3$. For Batog *et al.* [10, 11], the coordinates u_1, u_2 are "intrinsic parameters" of the camera. More precisely, they are *projective intrinsics*, since they are invariant under projective transformations. Batog *et al.* also argue that choosing the points y_1, y_2 as points in the projective basis on H leads to simplified analytic expressions for the projection. Here, we develop this idea further, and observe that a two-slit camera with this kind of coordinate system can be equivalently represented using the bilinear description $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ mentioned at the beginning of the section.

For any retinal plane H, we fix a coordinate system defined by a basis $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3)$ where $[\mathbf{y}_1] = l_2 \wedge H$, $[\mathbf{y}_2] = l_1 \wedge H$ and $\mathbf{y}_3$ is arbitrary: in this case, a straightforward computation shows that (3.13) can be written in the form

$$Q(x) = \begin{bmatrix} (\mathbf{p}_1^T\mathbf{x})\ (\mathbf{q}_2^T\mathbf{x}) \\ (\mathbf{p}_2^T\mathbf{x})\ (\mathbf{q}_1^T\mathbf{x}) \\ (\mathbf{p}_2^T\mathbf{x})\ (\mathbf{q}_2^T\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \mathbf{p}_1^T\mathbf{x}/\mathbf{p}_2^T\mathbf{x} \\ \mathbf{q}_1^T\mathbf{x}/\mathbf{q}_2^T\mathbf{x} \\ 1 \end{bmatrix}, \tag{3.17}$$

where $\mathbf{p}_1 = (\mathbf{l}_1 \vee \mathbf{y}_3)$, $\mathbf{p}_2 = -(\mathbf{l}_1 \vee \mathbf{y}_1)$, $\mathbf{q}_1 = (\mathbf{l}_2 \vee \mathbf{y}_3)$, $\mathbf{q}_2 = -(\mathbf{l}_2 \vee \mathbf{y}_2)$ are vectors representing planes in $\mathbb{P}^3$. This quadratic map can be described using *two linear maps* $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1$, namely

$$Q_1(x) = \begin{bmatrix} \mathbf{p}_1^T\mathbf{x} \\ \mathbf{p}_2^T\mathbf{x} \end{bmatrix} = [\mathbf{A}_1\mathbf{x}],\ Q_2(x) = \begin{bmatrix} \mathbf{q}_1^T\mathbf{x} \\ \mathbf{q}_2^T\mathbf{x} \end{bmatrix} = [\mathbf{A}_2\mathbf{x}]. \tag{3.18}$$

In other words, (3.17) determines the 2×4 matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ up to two scale factors. The two 2×4 matrices for a two-slit camera are analogues of the 3×4 matrix representing

a pinhole camera: for example, the slits are associated with the null-spaces of these two matrices. The linear maps $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1$ correspond in fact to the "line-centered" projections for the two slits.

Conversely, given a pair of 2×4 projection matrices $\mathbf{A}_1, \mathbf{A}_2$, we can use (3.17) to obtain a (quadratic) mapping $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$. This represents a photographic two-slit camera with a retinal plane. The retinal plane is however not completely fixed: it may be any plane containing the line $\{\mathbf{p}_2^T\mathbf{x} = \mathbf{q}_2^T\mathbf{x} = 0\}$. This line passes is spanned by y_1 and y_2, and is the locus of points where (3.17) is not defined.

In summary, representing a two-slit camera as a bilinear map $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ is essentially equivalent to using a retinal plane with a reference frame in which the first two basis points are $l_1 \wedge H$, $l_2 \wedge H$. Clearly, we can assume this whenever the coordinates of $l_1 \wedge H$, $l_2 \wedge H$ (the projective intrinsic parameters in [11]) are known. In the remainder of the section, we will always represent a two-slit camera with a pair of 2×4 projection matrices, writing $Q = ([\mathbf{A}_1], [\mathbf{A}_2])$ for the camera represented by $\mathbf{A}_1, \mathbf{A}_2$. We will however sometimes also consider the associated "physical realization" with a retinal plane, using the projection $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ given by (3.17).

Remark 3.3.3. It is clear from the representation in terms of projection matrices that our two-slit model has $7 + 7 = 14$ degrees of freedom. In terms of retinal planes: there are 8 degrees of freedom corresponding to the choice of the slits, 2 for the intersection points of the retinal plane with the slits, and 4 for the choice of coordinates on the plane (since two basis points are constrained).

Example 3.3.4. The two-slit projection from Example 3.3.1 is of the form (3.17) with

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \mathbf{A}_2 = \begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}. \tag{3.19}$$

The retinal plane belongs to the pencil of planes containing $\{x_3 = x_3 + x_4 = 0\}$, *i.e.*, it is a plane of the form $x_3 - dx_4 = 0$. The choice $d = 1$ is natural since points of the form $[x_1, x_2, 1, 1]$ are mapped to $[x_1, x_2, 1]$. ♢

3.3.2 Orbits and calibration matrices

The orbit of a photographic camera under the action of projective, affine, similarity, or euclidean transformations is a family of cameras that are geometrically and analytically equivalent for that "geometry". In the following, we will refer to an orbit as a *primitive camera model*, where the word "primitive" refers to the fact that the model represents a single device up to changes of world coordinates. We next describe some primitive models for two-slit cameras. This is analogous to the discussion on pinhole cameras from Section 1.2.4.

Our analysis is based on the simple observation that applying a projective transformation $[\mathbf{T}]$ to a two-slit camera $Q = ([\mathbf{A}_1], [\mathbf{A}_2])$ yields $Q' = ([\mathbf{A}_1\mathbf{T}], [\mathbf{A}_2\mathbf{T}])$.[1]. This immediately implies that all photographic two-slit cameras represented as pairs 2×4-matrices are projectively equivalent (as noted above, this is in contrast with the quadratic model (3.13)). Indeed, given any pair of two-slit cameras $Q = ([\mathbf{A}_1], [\mathbf{A}_2])$ and $Q' = ([\mathbf{B}_1], [\mathbf{B}_2])$, a projective transformation relating them is $[\mathbf{A}^{-1}\mathbf{B}]$, where $\mathbf{A}$ and $\mathbf{B}$ are obtained by stacking $\mathbf{A}_1, \mathbf{A}_2$ and $\mathbf{B}_1, \mathbf{B}_2$ respectively (note that $\mathbf{A}$ and $\mathbf{B}$ will always be invertible because slits are disjoint). In particular, we see that our bilinear model $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ is a projective primitive camera model, with 14 degrees of freedom.

We next consider affine transformations, and look at the orbit of the two-slit camera from Example 3.3.4. It is easy to see that all cameras $Q = ([\mathbf{A}_1], [\mathbf{A}_2])$ that are equivalent to (3.19) have the form

$$\mathbf{A}_1 = \begin{bmatrix} \mathbf{m}_1^T & t_1 \\ \mathbf{m}_3^T & t_3 \end{bmatrix}, \mathbf{A}_2 = \begin{bmatrix} \mathbf{m}_2^T & t_2 \\ \mathbf{m}_3^T & t_4 \end{bmatrix}, \tag{3.20}$$

where $\mathbf{m}_i$ are arbitrary 3-vectors. Viewed as maps $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$, as in (3.17), these are two-slit cameras where the retinal plane is parallel to the slits: indeed, while this plane is not completely determined, it must contain the line $\{[\mathbf{m}_3^T, t_3]\mathbf{x} = [\mathbf{m}_3^T, t_4]\mathbf{x} = 0\}$, that is spanned by intersections of the slits with the plane at infinity. We will refer to (3.20) as a *parallel* two-slit camera. This describes a primitive affine model with 12 degrees of freedom.

We now consider the family of parallel cameras of the form

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \ \mathbf{A}_2 = \begin{bmatrix} 2\cos\theta & 2\sin\theta & 0 & 0 \\ 0 & 0 & 1 & d \end{bmatrix}. \tag{3.21}$$

for $d \neq 0$ and $0 < \theta < 2\pi$ (and $\theta \neq \pi$). From the viewpoint of euclidean geometry, the slits for this camera are at an angle of θ and distance d. Note that (3.15) is of this form, with $\theta = \pi/2$ and $d = 1$. Using (3.21) as a family of canonical euclidean devices, we can introduce expressions for the "intrinsic parameters" of two-slit cameras.

Proposition 3.3.5. *If $Q = ([\mathbf{A}_1], [\mathbf{A}_2])$ is a parallel two-slit camera (3.20), then we can uniquely write*

$$\mathbf{A}_1 = \mathbf{K}_1 \begin{bmatrix} \mathbf{r}_1^T & t_1 \\ \mathbf{r}_3^T & t_3 \end{bmatrix}, \ \mathbf{A}_2 = \mathbf{K}_2 \begin{bmatrix} \mathbf{r}_2^T & t_2 \\ \mathbf{r}_3^T & t_4 \end{bmatrix}, \tag{3.22}$$

[1] Here by "applying" we actually mean that $Q'(x) = Q(T(x))$, or $Q'(T^{-1}(x)) = Q(x)$. Exchanging the role of T and T^{-1} has no effect on our discussion.

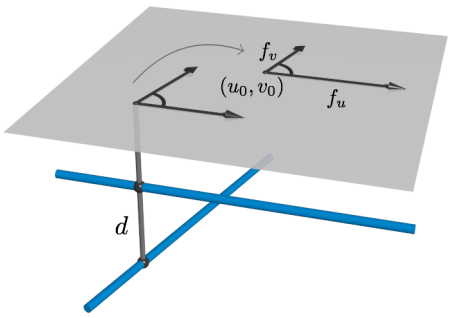


Figure 3.3: Physical interpretation of the entries the calibration matrices for parallel two-slit cameras: the parameters f_u, f_v, u_0, v_0 describe the change of retinal plane coordinates, with respect to some camera in the euclidean orbit of (3.21).

where $\mathbf{K}_1$ *and* $\mathbf{K}_2$ *are upper-triangular* 2×2 *matrices defined up to scale with positive elements along the diagonal, and* $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ *are unit vectors, with* $\mathbf{r}_3$ *orthogonal to both* $\mathbf{r}_1, \mathbf{r}_2$*. Here,* $\theta = \arccos(\mathbf{r}_1 \cdot \mathbf{r}_2)$ *is the* angle *between the slits, and* $|t_4 - t_3|$ *is the* distance *between the slits.*

Moreover, if the matrices $\mathbf{K}_1$ *and* $\mathbf{K}_2$ *are written as*

$$\mathbf{K}_1 = \begin{bmatrix} f_u & u_0 \\ 0 & 1 \end{bmatrix}, \mathbf{K}_2 = \begin{bmatrix} 2f_v & v_0 \\ 0 & 1 \end{bmatrix}, \tag{3.23}$$

then f_u, f_v *can be interpreted as "magnifications" in the* u *and* v *directions, and* (u_0, v_0) *as the position of the "principal point". See Figure 3.3.*

Proof. The decomposition exists and is unique because of RQ-decomposition of matrices [66, Theorem 5.2.3]. More precisely, if we write $\mathbf{A}_1 = [\mathbf{M}_1 \,|\, \mathbf{t}_1]$, $\mathbf{A}_2 = [\mathbf{M}_2 \,|\, \mathbf{t}_2]$, where $\mathbf{M}_1, \mathbf{M}_2$ are 2×3, then $\mathbf{K}_1$, $\mathbf{K}_2$ are the (normalized) upper triangular matrices in the RQ decomposition for $\mathbf{M}_1$, $\mathbf{M}_2$ respectively.

We next observe that for a pair canonical matrices

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \ \mathbf{A}_2 = \begin{bmatrix} 2\cos\theta & 2\sin\theta & 0 & 0 \\ 0 & 0 & 1 & d \end{bmatrix}, \tag{3.24}$$

the corresponding euclidean orbit is of the form

$$\begin{bmatrix} \mathbf{r}_1^T & t_1 \\ \mathbf{r}_3^T & t_3 \end{bmatrix}, \begin{bmatrix} 2\mathbf{r}_2^T & 2t_2 \\ \mathbf{r}_3^T & t_3 + d \end{bmatrix}, \tag{3.25}$$

where $\theta = \arccos(\mathbf{r}_1 \cdot \mathbf{r}_2)$. This follows by applying a 4×4 euclidean transformation matrix to (3.24). These cameras decompose with $\mathbf{K}_1$ being the identity and $\mathbf{K}_2 =$

diag$(2,1)$. Finally, if we indicate with $\mathbf{p}_1, \mathbf{p}_2$ and $2\mathbf{q}_1, \mathbf{q}_2$ the rows of (3.25), so that the corresponding camera can be written as $\mathbf{x} \mapsto \mathbf{u} = (\mathbf{p}_1^T\mathbf{x}/\mathbf{p}_2^T\mathbf{x}, 2\mathbf{q}_1^T\mathbf{x}/\mathbf{q}_2^T\mathbf{x}, 1)$, then the composition of $\begin{bmatrix}\mathbf{p}_1^T \\ \mathbf{p}_2^T\end{bmatrix}$, $\begin{bmatrix}\mathbf{q}_1^T \\ \mathbf{q}_2^T\end{bmatrix}$ with $\mathbf{K}_1, \mathbf{K}_2$ as in (3.23) yields the camera

$$[\mathbf{x}] \mapsto \left[f_u \frac{\mathbf{p}_1^T\mathbf{x}}{\mathbf{p}_2^T\mathbf{x}} + u_0, f_v \frac{2\mathbf{q}_1^T\mathbf{x}}{\mathbf{q}_2^T\mathbf{x}} + v_0, 1 \right]. \tag{3.26}$$

From this we easily deduce the physical interpretations of the entries of $\mathbf{K}_1$ and $\mathbf{K}_2$. □

The parameters θ and d, and the matrices $\mathbf{K}_1$ and $\mathbf{K}_2$ are clearly invariant under euclidean transformations. Moreover, within the parallel model (3.20), two cameras belong to the same euclidean orbit if and only if all of their parameters are the same. In fact, the 12 degrees of freedom of a parallel camera are split into 6 corresponding to the "intrinsics" θ, d, $\mathbf{K}_1$ and $\mathbf{K}_2$, and 6 for the "extrinsic" action of euclidean motion.[2] Compared to the traditional intrinsic parameters for pinhole cameras, we note the absence of a term corresponding to the "skewness" of the image reference frame. Indeed, the angle between the two axes must be the same as the angle between the slits, as a consequence of the "intrinsic coordinate system" (the principal directions correspond in fact to the fixed basis points y_1, y_2 on the retinal plane). On the other hand, θ and d do not have an analogue for pinhole cameras. We can think of d and θ as the "3D" intrinsic parameters, since they are distinct from the "analytic" intrinsic parameters, that are entries of the calibration matrices $\mathbf{K}_1, \mathbf{K}_2$, and differentiate (euclidean orbits of) cameras only based on analytic part of their mapping. See also Figure 3.4, that shows images taken with virtual two-slit cameras with slits forming different angles. Finally, we point out that for two-slit cameras in (3.22), the euclidean orbit and the similarity orbit do *not* coincide: this implies that when the intrinsic parameters are known, some information on the scale of a scene can be inferred from a photograph [173].

Pushbroom cameras. Pushbroom cameras are a degenerate class of projective two-slit cameras, in which one of the slits lies on the plane at infinity [55]. This is quite similar to the class affine cameras for perspective projections. Pushbrooms are handled by our projective model (3.17), but not by our affine one (3.20), where both slits are necessarily finite. We thus introduce another affine model, namely

$$\mathbf{A}_1 = \begin{bmatrix}\mathbf{m}_1^T & t_1 \\ \mathbf{0}^T & 1\end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix}\mathbf{m}_2^T & t_2 \\ \mathbf{m}_3^T & t_3\end{bmatrix}, \tag{3.27}$$

[2] Intrinsic parameters describing euclidean orbits among more general (non-parallel) two-slit cameras can also be defined, but two more parameters are required. We chose to consider only two-slits with retinal plane parallel to the slits, since this is a natural assumption, and because the parameters have a simpler interpretation in this case.

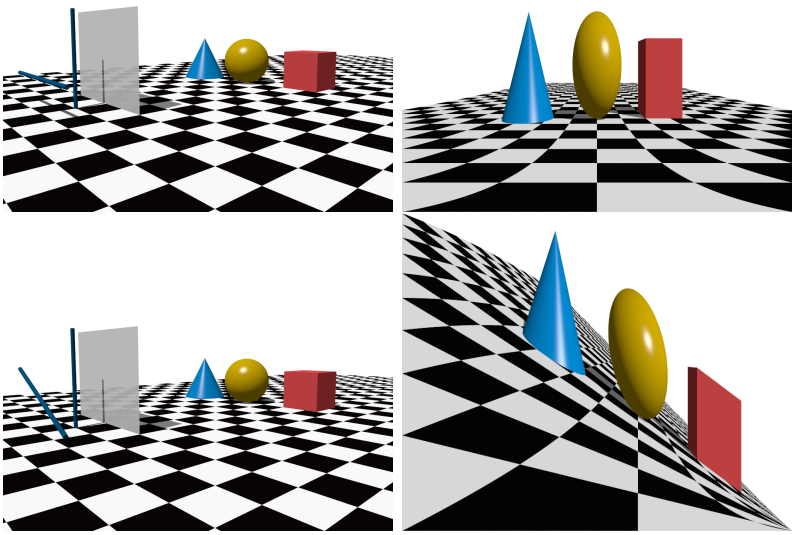


Figure 3.4: Simulated two-slit images of an artificial scene. The figures on the left show the imaging device, consisting of a retinal plane and two slits forming different angles (*top:* $\theta = \pi/2$; *bottom:* $\theta = 5/8\pi$). The figures on the right show the resulting images.

where $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$ are arbitrary 3-vectors. All such cameras are equivalent up to affine transformations, so this describes an affine model with 11 degrees of freedom. The corresponding rational cameras $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ can be written as

$$[\mathbf{x}] \mapsto \left[[\mathbf{m}_1^T, t_1]\mathbf{x}, \frac{[\mathbf{m}_2^T, t_2]\mathbf{x}}{[\mathbf{m}_3^T, t_3]\mathbf{x}}, 1 \right]. \tag{3.28}$$

This coincides with the linear pushbroom model proposed by Hartley and Gupta [72], who identify (3.28) with the 3×4 matrix with rows $[\mathbf{m}_1^T, t_1], [\mathbf{m}_2^T, t_2], [\mathbf{m}_3^T, t_3]$.

Let us now consider a family of affine pushbroom cameras of the form

$$\begin{bmatrix} \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \tag{3.29}$$

for $0 < \theta < 2\pi$ (and $\theta \neq \pi$). This represents a pushbroom camera where the direction of motion of the sensor is at an angle θ with respect to the parallel scanning planes. We use this family of canonical devices to define calibration matrices and intrinsic parameters for pushbroom cameras.

Proposition 3.3.6. *Let $Q = ([\mathbf{A}_1], [\mathbf{A}_2])$ be a pushbroom camera as in (3.27), and let us also assume that $\mathbf{m}_1$ and $\mathbf{m}_3$ are orthogonal.*[3] *We can uniquely write*

$$\mathbf{A}_1 = \mathbf{K}_1 \begin{bmatrix} \mathbf{r}_1^T & t_1 \\ \mathbf{0}^T & 1 \end{bmatrix}, \mathbf{A}_2 = \mathbf{K}_2 \begin{bmatrix} \mathbf{r}_2^T & t_2 \\ \mathbf{r}_3^T & t_3 \end{bmatrix}. \tag{3.30}$$

where $\mathbf{K}_1 = \operatorname{diag}(1/v, 1)$, $\mathbf{K}_2 = \begin{bmatrix} f & u \\ 0 & 1 \end{bmatrix}$ (with positive v and f) and $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ are unit vectors, with $\mathbf{r}_3$ orthogonal to both $\mathbf{r}_1, \mathbf{r}_2$. Here, $\theta = \arccos(\mathbf{r}_1 \cdot \mathbf{r}_2)$ is the angle *between the two slits (or between the direction of motion of the sensor and the parallel scanning planes). Moreover, v can be interpreted as the speed of the sensor, and f and u as the magnification and the principal point of the 1D projection.*

Proof. The proof is similar to that of Proposition 3.3.5. The decomposition is unique because of QR-factorization of matrices. The euclidean orbits of "canonical" pushbroom cameras have the form

$$\begin{bmatrix} \mathbf{r}_1^T & t_1 \\ \mathbf{0}^T & 1 \end{bmatrix}, \begin{bmatrix} \mathbf{r}_2^T & t_2 \\ \mathbf{r}_3^T & t_3 \end{bmatrix}.$$

All these cameras decompose with $\mathbf{K}_1, \mathbf{K}_2$ being the identity. Finally, the physical interpretation of the parameters follows by noting that composing a pushbroom camera (with rows $\mathbf{p}_1, (0,0,0,1)^T$ and $\mathbf{q}_1, \mathbf{q}_2$) with calibration matrices $\mathbf{K}_1, \mathbf{K}_2$ yields

$$[\mathbf{x}] \mapsto \left[\frac{1}{v} \mathbf{p}_1^T \mathbf{x}, f \frac{\mathbf{q}_1^T \mathbf{x}}{\mathbf{q}_2^T \mathbf{x}} + u, 1 \right].$$

From this we deduce the physical interpretations of the entries of $\mathbf{K}_1$ and $\mathbf{K}_2$. □

Similarly to parallel two-slit cameras, the entries of $\mathbf{K}_1, \mathbf{K}_2$ can be seen as the "analytic" intrinsic parameters of the pushbroom camera, while θ is a "3D" intrinsic parameter.

3.3.3 Epipolar geometry and algorithms

We next describe the epipolar geometry for two-slit cameras using a $2 \times 2 \times 2 \times 2$ *epipolar tensor*. It is worth noting that image correspondences between two-slit cameras have previously been characterized using a 6×6 [55], or a 4×4 fundamental matrix [11]. The latter approach, due to Batog *et al.* [11], is somewhat similar to ours, since it is based on the "intrinsic" image reference frame that is equivalent to our bilinear model.

[3] The result can be extended to the case of arbitrary pushbroom cameras, by allowing for the calibration matrix $\mathbf{K}_2$ to be a general 2×2 matrix with positive entries along the diagonal. In that case, however, the geometric interpretation for the parameters is less clear.

However, the tensor representation has the advantage of being easily described in terms of the elements of the four 2×4 projection matrices, in a form that closely resembles the corresponding expression for the traditional fundamental matrix.

We recall here that a pair of two-slit cameras $Q_1 = ([\mathbf{A}_1], [\mathbf{A}_2])$, $Q_2 = ([\mathbf{B}_1], [\mathbf{B}_2])$ is projectively equivalent to another pair $Q_1' = ([\mathbf{A}_1'], [\mathbf{A}_2'])$, $Q_2' = ([\mathbf{B}_1'], [\mathbf{B}_2'])$ if the two pairs of cameras are related by a projective transformation, *i.e.*, if there exists a projective transformation $T = [\mathbf{T}]$ of $\mathbb{P}^3$ such that $\mathbf{A}_i' = \mathbf{A}_i\mathbf{T}$ and $\mathbf{B}_i' = \mathbf{B}_i\mathbf{T}$ for $i = 1, 2$. An equivalence class of projectively equivalent pairs of cameras defines a "projective configuration".

Theorem 3.3.7. *Let $Q_1 = ([\mathbf{A}_1], [\mathbf{A}_2])$, $Q_2 = ([\mathbf{B}_1], [\mathbf{B}_2])$ be two general projective two-slit cameras. The set of corresponding image points $((u, v), (u', v'))$ in $\mathbb{P}^1 \times \mathbb{P}^1$ is characterized by the relation*

$$\sum_{i,j,k,l=1}^{2} f_{ijkl}\, u_i v_j u_k' v_l' = 0, \tag{3.31}$$

where $\mathbf{F} = (f_{ijkl})$ is a $2 \times 2 \times 2 \times 2$ "epipolar tensor". Its entries are

$$f_{ijkl} \;=\; (-1)^{i+j+k+l} \cdot \det\left[(\mathbf{A}_1)_{3-i}^T \;(\mathbf{A}_2)_{3-j}^T \;(\mathbf{B}_1)_{3-k}^T \;(\mathbf{B}_2)_{3-l}^T\right], \tag{3.32}$$

where $(\mathbf{A}_i)_j$ denotes the j-th row of $\mathbf{A}_i$, and similarly for $\mathbf{B}_i$.

All pairs of two-slit cameras in the same projective configuration determine the same $2 \times 2 \times 2 \times 2$ epipolar tensor $\mathbf{F}$. Conversely, in general, a valid epipolar tensor $\mathbf{F}$ determines two *distinct projective configurations of pairs of two-slit cameras.*

We note that this result is consistent with the theory presented by Hartley and Schaffalitzky in [77], where it is shown that "Grassmann tensors" (generalizations of multi-view tensors for projections in arbitrary dimension), lead to a two-fold ambiguity in the reconstruction when the image is a $\mathbb{P}^1$. A more mathematical proof of the same result is given in [90]. Our epipolar tensor is also formally identical to the "radial quadrifocal tensor", defined by Thirthala and Pollefeys in [181] for quadruples of cameras with axial symmetry (indeed, each such camera defines a linear map $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1$).

Proof. The inverse line projection $\mathbb{P}^1 \times \mathbb{P}^1 \to \mathrm{Gr}(1, \mathbb{P}^3)$ of a two-slit camera is a bilinear map given by

$$\xi(u, v) = \left[\sum_{ij} (-1)^{i+j} u_i v_j \left((\mathbf{A}_1)_{3-i} \wedge (\mathbf{A}_2)_{3-j}\right)\right] \in \mathrm{Gr}(1, \mathbb{P}^3). \tag{3.33}$$

The definition of $\mathbf{F}$ now follows immediately by imposing the condition that $\xi(u, v)$ and $\xi(u', v')$ are concurrent.

The definition of $\mathbf{F}$ is clearly invariant under projective transformations of $\mathbb{P}^3$. Hence, assuming that the vectors $(\mathbf{A}_1)_1, (\mathbf{A}_2)_1, (\mathbf{B}_1)_1, (\mathbf{B}_2)_1$ are independent (which is true generically) we can apply a change of reference frame in $\mathbb{P}^3$ so that the projection matrices have the form

$$\begin{aligned} \mathbf{A}_1 &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ c_{11} & c_{12} & c_{13} & c_{14} \end{bmatrix}, \ \mathbf{A}_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ c_{21} & c_{22} & c_{23} & c_{24} \end{bmatrix}, \\ \mathbf{B}_1 &= \begin{bmatrix} 0 & 0 & 1 & 0 \\ c_{31} & c_{32} & c_{33} & c_{34} \end{bmatrix}. \ \mathbf{B}_2 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ c_{41} & c_{42} & c_{43} & c_{44} \end{bmatrix}. \end{aligned} \tag{3.34}$$

The 16 entries of $\mathbf{F}$ are now, up to sign, the *principal minors* of the 4×4-matrix $\mathbf{C} = (c_{ij})$. More precisely, $f_{ijkl} = (-1)^{i+j+k+l} \det \mathbf{C}_{[i-1,j-1,k-1,l-1]}$ where $\mathbf{C}_{[i-1,j-1,k-1,l-1]}$ is the submatrix of $\mathbf{C}$ where the selected rows and columns correspond to the binary vector $[i-1, j-1, k-1, l-1]$ (*e.g.*, $\mathbf{C}_{[1,0,0,0]} = (c_{11})$). Thus, determining the projection matrices $(\mathbf{A}_1, \mathbf{A}_2)$, $(\mathbf{B}_1, \mathbf{B}_2)$ corresponding to the tensor $\mathbf{F}$, is equivalent to finding the entries of a 4×4-matrix given its principal minors. This problem is studied in [117]. The set of all matrices with the same principal minors as $\mathbf{C}$ have the form $\mathbf{D}^{-1}\mathbf{C}\mathbf{D}$ or $\mathbf{D}^{-1}\mathbf{C}^T\mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix. These two families of matrices, viewed as elements of (3.34), correspond to two distinct projective configurations of cameras. □

The set of all epipolar tensors forms a 13-dimensional variety in $\mathbb{P}^{15}$: this agrees with $14 + 14 - 15 = 13$, where 14 represents the degrees of freedom of two-slit cameras, and 15 is to account for projective ambiguity. Two equations are sufficient to characterize an epipolar tensor *locally*, however a result in [117] implies that a complete algebraic characterization actually requires 718 polynomials of degree 12.

Remark 3.3.8. Our study of canonical forms and calibration matrices in Section 3.3.2 leads to a natural definition of *essential tensors*: for example, an essential tensor could be defined by (3.32) where $Q_1 = ([\mathbf{A}_1], [\mathbf{A}_2]), Q_2 = ([\mathbf{B}_1], [\mathbf{B}_2])$ are all of the form (3.22) with $\mathbf{K}_1$, $\mathbf{K}_2$ being the identity. Proposition 3.3.5 then guarantees that for any pair of "parallel" two-slit cameras as in (3.20), we can uniquely write the epipolar tensor as

$$\mathbf{F}_{ijkl} = \mathbf{E}_{ijkl}(\mathbf{K}_{1\mathbf{A}})_i(\mathbf{K}_{2\mathbf{A}})_j(\mathbf{K}_{1\mathbf{B}})_k(\mathbf{K}_{2\mathbf{B}})_l, \tag{3.35}$$

where $\mathbf{E}_{ijkl}$ is an essential tensor. This closely resembles the analogous decomposition of fundamental matrices. Recovering an algebraic characterization of essential tensors, similar to the classical Demazure equations [47], could be an interesting topic for future work (the problem can also be generalized to linear projections in arbitrary dimensions).

Linear SFM. Using Theorem 3.3.7, we can design a *linear algorithm* for SfM, that proceeds as follows:

1. Using at least 15 image point correspondences, estimate $\mathbf{F}$ linearly using (3.31).
2. Recover two projective camera configurations that are compatible with $\mathbf{F}$.

Clearly, for noisy image correspondences, the linear estimate from step 1 will not be a valid epipolar tensor, since tensors of the form (3.32) are not generic (as noted above, they form a 13-dimensional variety in $\mathbb{P}^{15}$). On the other hand, it is possible to recover projection matrices from only 13 of the entries of $\mathbf{F}$, and this avoids the problem of using a valid tensor. A simple scheme for this is as follows:

1. We set out to recover the entries of a 4×4-matrix $\mathbf{C}$ given its principal minors. Since we can always replace $\mathbf{C}$ with $\mathbf{D}^{-1}\mathbf{C}\mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix, we can assume that $c_{12} = c_{13} = c_{13} = c_{14} = 1$ (at least generically). Other similar assignments are possible.
2. Elements on the diagonal and on the first column of $\mathbf{C}$ are easily computed using seven of the entries of $\mathbf{F}$:
$$c_{11} = -f_{1222};\ c_{22} = -f_{2122};\ c_{33} = -f_{2212};\ c_{44} = -f_{2221};$$
$$c_{21} = (c_{11}c_{22} - f_{1122})/c_{12};\ c_{31} = (c_{11}c_{33} - f_{1212})/c_{13};\ c_{41} = (c_{11}c_{44} - f_{1221})/c_{14}.$$
Here the elements to the right of the equal signs have already been assigned. Hence, we recover 10 entries of $\mathbf{C}$ from linear equalities.
3. The remaining six entries of $\mathbf{C}$ are pairwise constrained by six elements of $\mathbf{F}$. For example, using the minors f_{2112}, f_{1112} (corresponding to rows/columns $2, 3$ and $1, 2, 3$ of $\mathbf{C}$) we deduce that c_{32} must satisfy $ac_{32}^2 + bc_{32} + c = 0$ where
$$\begin{aligned} a &= c_{13}c_{21}, \\ b &= f_{1112} + c_{11}f_{2112} - c_{13}c_{31}c_{22} - c_{12}c_{21}c_{33}, \\ c &= c_{12}c_{31}c_{22}c_{33} - c_{12}c_{31}f_{2112}, \end{aligned}$$
and that $c_{23} = (c_{22}c_{33} - f_{2112})/c_{32}$. Similar relations hold for the pairs c_{24}, c_{42} and c_{34}, c_{43}. This leads to 8 possible matrices $\mathbf{C}$, *i.e.*, a finite number of camera configurations. Note however that the entries f_{1111} and f_{2111} of $\mathbf{F}$ were never used (which is why we can assume the tensor to be generic): in an ideal setting with no noise, exactly two of the 8 solutions will be consistent with the remaining constraints.

This approach for recovering two-slit projections from the corresponding epipolar tensor relies on some genericity assumptions (*e.g.*, we have often divided by element without verifying that it is not zero), and developing an optimal strategy for this task is outside the scope of our work. Nevertheless, we include as a proof of concept an example that applies this method on synthetic data.

Example 3.3.9. We first illustrate concretely some basic properties of the fundamental tensor. We consider the following pairs of projection matrices:

$$\begin{aligned} \mathbf{A}_1 &= \begin{bmatrix} -1 & 7 & 4 & 0 \\ 8 & -1 & 13 & 4 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 11 & 6 & -2 & 4 \\ 8 & -1 & 13 & -5 \end{bmatrix}, \\ \mathbf{B}_1 &= \begin{bmatrix} 14 & 9 & -3 & 8 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{B}_2 = \begin{bmatrix} -3 & 8 & 10 & 3 \\ 6 & 13 & 5 & 13 \end{bmatrix}. \end{aligned} \tag{3.36}$$

The pair $\mathbf{A}_1, \mathbf{A}_2$ represents a parallel finite two-slit camera, while $\mathbf{B}_1, \mathbf{B}_2$ is a pushbroom camera. The associated epipolar tensor (3.32) is

$$\begin{aligned} \mathbf{F} = &\begin{bmatrix} 0 & 0 \\ 21816 & -25650 \end{bmatrix} \begin{bmatrix} 1906 & -2090 \\ -3642 & 5510 \end{bmatrix} \\ &\begin{bmatrix} 880 & 475 \\ 18600 & -11875 \end{bmatrix} \begin{bmatrix} 97 & -380 \\ -1259 & 1425 \end{bmatrix}, \end{aligned} \tag{3.37}$$

where each 4×4 matrix represents a block $(f_{ijkl})_{kl}$ for fixed i, j. Note that f_{1111} and f_{1112} are zero, since the second rows of $\mathbf{A}_1, \mathbf{A}_2, \mathbf{B}_1$ are linearly dependent. Using the approach outlined above, we can use this tensor to recover two matrices $\mathbf{C}_1, \mathbf{C}_2$ whose principal minors are the entries of $\mathbf{F}$ (we must normalize $\mathbf{F}$ so that $f_{2222} = 1$). We use these matrices to construct two pairs of two-silt cameras, namely[4]

$$\begin{aligned} \mathbf{A}_1^1 &= \begin{bmatrix} 1. & 0. & 0. & 0. \\ -3.87 & 1. & 1. & 1. \end{bmatrix} \\ \mathbf{A}_2^1 &= \begin{bmatrix} 0. & 1. & 0. & 0. \\ -14.22 & 8.33 & -6.67 & -22.17 \end{bmatrix}, \\ \mathbf{B}_1^1 &= \begin{bmatrix} 0. & 0. & 1. & 0. \\ 0.44 & -0.28 & 0.27 & 1.14 \end{bmatrix} \\ \mathbf{B}_2^1 &= \begin{bmatrix} 0. & 0. & 0. & 1. \\ -0.86 & 0.26 & 0.15 & 0.88 \end{bmatrix}, \end{aligned} \tag{3.38}$$

[4] This is a floating point approximation of the matrices, because recovering $\mathbf{C}_1, \mathbf{C}_2$ requires solving quadratic equations. Since for the moment we do not assume any noise, it would also possible to use exact symbolic representations.

and

$$\begin{aligned}
\mathbf{A}_1^2 &= \begin{bmatrix} 1. & 0 & 0 & 0 \\ -3.87 & 1. & 1. & 1. \end{bmatrix} \\
\mathbf{A}_2^2 &= \begin{bmatrix} 0. & 1. & 0. & 0. \\ -14.22 & 8.33 & 9.25 & 4.24 \end{bmatrix}, \\
\mathbf{B}_1^2 &= \begin{bmatrix} 0. & 0. & 1. & 0. \\ 0.44 & 0.20 & 0.27 & -0.07 \end{bmatrix} \\
\mathbf{B}_2^2 &= \begin{bmatrix} 0. & 0. & 0. & 1. \\ -0.86 & -1.34 & -2.26 & 0.88 \end{bmatrix}.
\end{aligned} \tag{3.39}$$

Computing the epipolar tensor (3.31) for both of these pairs yields $\mathbf{F}$ as in (3.37). On the other hand, the two camera configurations are *not* projectively equivalent: indeed, if a projective transformation between the two existed, it would need to be the identity, because five of the eight rows coincide. It is straightforward to verify that it is in fact the second pair that corresponds to the configuration of the original cameras (3.36).

We next try to estimate the cameras (3.36) from image correspondences. We consider 70 random points in space, project them onto an image plane using (3.36) together with (3.17). We then add noise to the images of the points. Although none of the eight solutions in the method outlined above will be exactly consistent with the last two entries of $\mathbf{F}$, we can consider the two solutions that minimize an "algebraic residual" for these constraints. For image sizes of about 100×100, and noise with a standard deviation of 10^{-5}, we recover the following pairs of cameras (that should be compared with (3.38)):

$$\begin{aligned}
\mathbf{A}_1^1 &= \begin{bmatrix} 1. & 0. & 0. & 0. \\ -3.97 & 1. & 1. & 1. \end{bmatrix} \\
\mathbf{A}_2^1 &= \begin{bmatrix} 0. & 1. & 0. & 0. \\ -15.26 & 8.44 & -7.60 & -23.18 \end{bmatrix}, \\
\mathbf{B}_1^1 &= \begin{bmatrix} 0. & 0. & 1. & 0. \\ 0.42 & -0.25 & 0.27 & 1.17 \end{bmatrix} \\
\mathbf{B}_2^1 &= \begin{bmatrix} 0. & 0. & 0. & 1. \\ -0.86 & 0.25 & 0.14 & 0.88 \end{bmatrix},
\end{aligned} \tag{3.40}$$

and

$$\begin{aligned}
\mathbf{A}_1^2 &= \begin{bmatrix} 1. & 0. & 0. & 0. \\ -3.97 & 1. & 1. & 1. \end{bmatrix} \\
\mathbf{A}_2^2 &= \begin{bmatrix} 0. & 1. & 0. & 0. \\ -15.26 & 8.44 & 9.36 & 4.41 \end{bmatrix}
\end{aligned} \tag{3.41}$$

$$\mathbf{B}_1^2 = \begin{bmatrix} 0. & 0. & 1. & 0. \\ 0.42 & 0.20 & 0.27 & -0.07 \end{bmatrix}$$
$$\mathbf{B}_2^2 = \begin{bmatrix} 0. & 0. & 0. & 1. \\ -0.86 & -1.30 & -2.42 & 0.88 \end{bmatrix}. \tag{3.42}$$

♢

Minimal SFM. It is also possible to design a *13-point algorithm* that recovers projection matrices (3.34) and the corresponding tensor $\mathbf{F}$ from a minimal amount of data, namely 13 point correspondences between images. Substituting these correspondences in (3.31), we obtain an under-determined linear system, which implies that the epipolar tensor is a linear combination $\alpha T_1 + \beta T_2 + \gamma T_3$ for some T_1, T_2, T_3 that generate the corresponding null-space. Since the variety of epipolar tensors has codimension 2 in $\mathbb{P}^{15}$, we expect to find a finite number of feasible tensors in this linear space (up to scale factors). According to [117, Remark 14], the variety of epipolar tensors (that is viewed there as the projective variety for the principal minors of 4×4 matrices) has degree 28. Hence, this minimal approach should lead to 28 complex solutions for $\mathbf{F}$, and 56 projective configurations of cameras. Using the computer algebra system `Macaulay2` [70] we have verified (over finite fields) that imposing 13 general linear conditions on the 16 principal minors of the matrix $\mathbf{C}$, and fixing $c_{12} = c_{13} = c_{14} = 1$, we obtain 56 solutions $\mathbf{C}$ in the algebraic closure of the field.

Self-calibration. Any reconstruction based on the epipolar tensor will be subject to projective ambiguity. On the other hand, using results from Section 3.3.2, it is possible to develop strategies for *self-calibration*. Let us assume that we have recovered a projective reconstruction $\mathbf{A}_1^i, \mathbf{A}_2^i$ for $i = 1, \ldots, n$ for finite two-slit cameras (that we assume were originally "parallel"). We indicate with $\mathbf{Q}$ a "euclidean upgrade", that is, a 4×4-matrix that describes the transition from a euclidean reference frame to the frame corresponding to our projective reconstruction. According to Proposition 3.3.5, we may write $\mathbf{A}_1^i\mathbf{Q} = \mathbf{K}_i[\mathbf{R}_1^i \,|\, \mathbf{t}_1^i]$, $\mathbf{A}_2^i\mathbf{Q} = \mathbf{K}_2^i[\mathbf{R}_2^i \,|\, \mathbf{t}_2^i]$, where $\mathbf{R}_1^i, \mathbf{R}_2^i$ are 2×3 matrices with orthonormal rows (for simplicity, we remove the factor 2 from the canonical euclidean form). In particular, for all $i = 1, \ldots, k$, we have

$$\begin{aligned} \mathbf{A}_1^i\mathbf{Q}\boldsymbol{\Omega}^*\mathbf{Q}^T{\mathbf{A}_1^i}^T &= \mathbf{K}_1^i{\mathbf{K}_1^i}^T \\ \mathbf{A}_2^i\mathbf{Q}\boldsymbol{\Omega}^*\mathbf{Q}^T{\mathbf{A}_2^i}^T &= \mathbf{K}_2^i{\mathbf{K}_2^i}^T, \end{aligned} \tag{3.43}$$

where equality is up to scale and $\boldsymbol{\Omega}^* = \mathrm{diag}(1, 1, 1, 0)$. Geometrically, the matrix $\boldsymbol{\Omega}_Q^* = \mathbf{Q}\boldsymbol{\Omega}^*\mathbf{Q}^T$ represents the *dual of the absolute conic*, in the projective coordinates used in the reconstruction. The equations (3.43) identify in fact the *dual of the image of the absolute conic* in the two copies of $\mathbb{P}^1$. These are the set of planes containing each slit that are tangent to the absolute conic in $\mathbb{P}^3$.

We now assume that the principal points cameras are at the "origin", so that $\mathbf{K}_1^i, \mathbf{K}_2^i$ (and hence $\mathbf{K}_1^i\mathbf{K}_1^{iT}$ and $\mathbf{K}_2^i\mathbf{K}_2^{iT}$) are diagonal. Each row in (3.43) gives two linear equations in the elements of $\mathbf{\Omega}_Q^*$, corresponding to the zeros in the matrices on the right hand side. For example, imposing that the $(1,2)$-entry of $\mathbf{K}_1^i\mathbf{K}_1^{iT}$ is zero yields

$$\begin{aligned} &a_{11}a_{21}m_{11} + a_{11}a_{22}m_{12} + a_{11}a_{23}m_{13} + a_{11}a_{24}m_{14} \\ &+ a_{12}a_{21}m_{21} + a_{12}a_{22}m_{22} + a_{12}a_{23}m_{23} + a_{12}a_{24}m_{24} \\ &+ a_{13}a_{21}m_{31} + a_{13}a_{22}m_{32} + a_{13}a_{23}m_{33} + a_{13}a_{24}m_{34} \\ &+ a_{14}a_{21}m_{41} + a_{14}a_{22}m_{42} + a_{14}a_{23}m_{43} + a_{14}a_{24}m_{44} = 0, \end{aligned} \tag{3.44}$$

where $\mathbf{\Omega}_Q^* = (m_{ij})$, and the elements of $\mathbf{A}_1^i = (a_{ij})$ are known. A sufficient number of views allows us to estimate $\mathbf{\Omega}_Q^*$ linearly. Finally, from the singular value decomposition of $\mathbf{\Omega}_Q^*$, we can compute a matrix $\mathbf{Q}'$ such that $\mathbf{Q}'\mathbf{\Omega}^*\mathbf{Q}'^T = \mathbf{\Omega}_Q^*$. The matrix $\mathbf{Q}'$ is however not uniquely determined, and indeed we can actually only recover a *similarity* upgrade, since any similarity transformation will fix the absolute conic in $\mathbb{P}^3$.

Example 3.3.10. To apply our self-calibration scheme, we consider 10 cameras $\mathbf{A}_1^i, \mathbf{A}_2^i$, $i = 1, \ldots, 10$, of the form $\mathbf{A}_1^i = \mathbf{K}_1^i[\mathbf{R}_1^i \,|\, \mathbf{t}_1^i]\mathbf{Q}^{-1}$, $\mathbf{A}_2^i = \mathbf{K}_2^i[\mathbf{R}_2^i \,|\, \mathbf{t}_2^i]\mathbf{Q}^{-1}$, where $\mathbf{R}_1^i, \mathbf{t}_1^i$, $\mathbf{R}_2^i, \mathbf{t}_2^i$ are random parameters for euclidean primitive parallel cameras, $\mathbf{K}_1^i, \mathbf{K}_2^i$ are random *diagonal* calibration matrices, and $\mathbf{Q}$ is a random 4×4 matrix describing a projective change of coordinates. We also add small amounts of noise to the entries of $\mathbf{A}_1^i, \mathbf{A}_2^i$. The matrices $\mathbf{A}_1^i, \mathbf{A}_2^i$ represent a *projective* configuration of two-slit cameras. Using (3.43), we can recover an estimate for $\mathbf{\Omega}_Q^* = \mathbf{Q}\mathbf{\Omega}^*\mathbf{Q}^T$ by solving an over-constrained linear system (with 40 equations). From this, we compute a matrix $\mathbf{Q}'$ such that $\mathbf{Q}'\mathbf{\Omega}^*\mathbf{Q}'^T \simeq \mathbf{\Omega}_Q^*$. For our example, the original data was

$$\mathbf{Q} = \begin{bmatrix} 1.49 & 0.60 & -0.11 & -1.15 \\ -1.43 & 0.88 & -0.93 & 1.52 \\ -0.38 & -0.21 & 1.83 & -0.55 \\ 0.83 & -0.95 & -0.63 & 0.93 \end{bmatrix},$$

$$\mathbf{Q}\mathbf{\Omega}^*\mathbf{Q}^T = \begin{bmatrix} 1. & -0.58 & -0.34 & 0.28 \\ -0.58 & 1.42 & -0.52 & -0.55 \\ -0.34 & -0.52 & 1.36 & -0.49 \\ 0.28 & -0.55 & -0.49 & 0.77 \end{bmatrix},$$

while our estimates are

$$\mathbf{Q}' = \begin{bmatrix} -0.43 & 0.21 & 0.35 & 0. \\ 0.67 & 0.26 & 0.08 & 0. \\ -0.04 & -0.69 & 0.03 & 0. \\ -0.34 & 0.26 & -0.28 & 1. \end{bmatrix},$$

$$\mathbf{Q}'\mathbf{\Omega}^*\mathbf{Q}'^T = \begin{bmatrix} 1. & -0.59 & -0.34 & 0.29 \\ -0.59 & 1.44 & -0.51 & -0.56 \\ -0.34 & -0.51 & 1.35 & -0.48 \\ 0.29 & -0.56 & -0.48 & 0.75 \end{bmatrix}.$$

The matrices $\mathbf{Q}$, $\mathbf{Q}'$ are not close, however one easily verifies that $\mathbf{Q}^{-1}\mathbf{Q}'$ is (almost) a similarity transformation. In particular, the cameras $\mathbf{A}_1^i\mathbf{Q}', \mathbf{A}_2^i\mathbf{Q}'$, $i = 1, \ldots, 10$ are a "similarity upgrade" of the projective solution. For example, for the first of our 10 original cameras we had $\mathbf{K}_1^1 = \mathrm{diag}(4.04, 1)$, $\mathbf{K}_2 = \mathrm{diag}(1.37, 1)$, and indeed

$$\mathbf{A}_1^1\mathbf{Q}' = \begin{bmatrix} -2.07 & -1.29 & 3.23 & 13.25 \\ 0.39 & -0.91 & -0.12 & -0.08 \end{bmatrix},$$
$$\mathbf{A}_2^1\mathbf{Q}' = \begin{bmatrix} -0.49 & -0.36 & 1.24 & 2.81 \\ 0.38 & -0.91 & -0.12 & 0.53 \end{bmatrix},$$

describe a parallel two-slit camera, where the ratios between the norms of the rows (the "magnifications") are respectively 4.05 and 1.38. ♢

3.4 Properties of Photographic Cameras

After having studied two-slit cameras, we broaden our perspective and return to general rational photographic cameras. Much of the analysis carried out for pinhole cameras in Chapter 1, including the classification of orbits, the description multi-view constraints, and the study of camera configurations, can in principle be replicated for any camera model. We present in this section some first results and definitions that could be a point of departure for an in-depth treatment of all rational imaging models in the future.

In particular, we introduce in Section 3.4.1 a generalized version of the *joint image*, that characterizes point correspondences for systems of general photographic cameras. We also derive as a simple example the epipolar geometry between a pinhole and two-slit camera. In Section 3.4.2 we discuss some basic properties of rational photographic cameras, explaining in particular how to compute forward and inverse projections of lines and points. Finally, in Section 3.4.3 we introduce *trifocal* and *quadrifocal* forms, that characterize matching points and lines in different images, and generalize traditional multi-view tensors.

3.4.1 Multi-view geometry

In order to describe the multi-view geometry for arbitrary photographic cameras, we can use a generalized version of the *joint image variety* from Chapter 1. For any collection of photographic cameras $Q_1, \ldots, Q_{n_1}, Q'_1, \ldots, Q'_{n_2}$ where $Q_i : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ and $Q'_j : \mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$, we consider the closure of the image of

$$\mathbb{P}^3 \dashrightarrow (\mathbb{P}^2)^{n_1} \times (\mathbb{P}^1 \times \mathbb{P}^1)^{n_2}, \quad x \mapsto (Q_1(x), \ldots, Q_{n_1}(x), Q'_1(x), \ldots, Q'_{n_2}(x)). \tag{3.45}$$

This clearly extends the usual notion of joint image from Section 1.3 and [4, 186, 190].

The following immediate result relates the joint image and the "multi-image variety" of lines, introduced in Section 2.4.2. This is a generalization of Proposition 2.4.5 from pinhole cameras to arbitrary photographic cameras. We write C_{Q_i} and $C_{Q'_j}$ for the congruences associated with Q_i and Q'_j.

Proposition 3.4.1. *The joint image variety for $Q_1, \ldots, Q_{n_1}, Q'_1, \ldots, Q'_{n_2}$ is birationally equivalent to the multi-image variety $M(C_{Q_1}, \ldots, C_{Q_{n_1}}, C_{Q'_1}, \ldots, C_{Q'_{n_2}})$, defined in Section 2.4.2, under the map*

$$L_{Q_1} \times \cdots \times L_{Q_{n_1}} \times L_{Q'_1} \times \cdots \times L_{Q'_{n_2}} : (\mathbb{P}^2)^{n_1} \times (\mathbb{P}^1 \times \mathbb{P}^1)^{n_2} \dashrightarrow \mathrm{Gr}(1, \mathbb{P}^3)^n, \tag{3.46}$$

where $L_{Q_i}, L_{Q'_j}$ are "inverse projections" of $L_{Q_i} : \mathbb{P}^2 \dashrightarrow \mathrm{Gr}(1, \mathbb{P}^3)$, $L_{Q'_j} : \mathbb{P}^1 \times \mathbb{P}^1 \dashrightarrow \mathrm{Gr}(1, \mathbb{P}^3)$ for the congruences $C_{Q_i}, C_{Q'_j}$.

The practical importance of this observation is that we can use it together with the theory from Chapter 2 to recover multi-view constraints characterizing the joint image for any set of photographic cameras. More precisely, from Theorem 2.4.3 we deduce that, if the base loci of $Q_1, \ldots, Q_{n_1}, Q'_1, \ldots, Q'_{n_2}$ are pairwise disjoint, then the joint image variety is birationally equivalent to a slice of the *concurrent lines variety* $V_{n_1+n_2}$. Since the closure of the image of (3.46) is $C_{M_1} \times \cdots \times C_{Q_{n_1}} \times C_{Q'_1} \times \cdots \times C_{Q'_{n_2}}$, we can obtain multi-view constraints in image coordinates by replacing the Plücker variables with the coordinates of $L_{Q_i}(w)$ and $L_{Q'_j}((u, v))$ in the multilinear polynomials defining $V_{n_1+n_2}$.

Epipolar geometry: pinhole + two-slit. As a simple example (which doesn't require the concurrent lines variety but only the incidence condition for two lines) we characterize the joint-image for a pinhole camera $Q_1 : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ with $Q_1 = [\mathbf{A}]$ and a two-slit camera $Q_2 : \mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$ with $Q_2 = ([\mathbf{B}], [\mathbf{C}])$. The inverse projection $L_{Q_1} : \mathbb{P}^2 \to \mathrm{Gr}(1, \mathbb{P}^3)$ is given by (cf. Section 1.2.5)

$$[\mathbf{u}] \;\mapsto\; [u_1(\mathbf{A}_2 \wedge \mathbf{A}_3) - u_2(\mathbf{A}_1 \wedge \mathbf{A}_3) + u_3(\mathbf{A}_1 \wedge \mathbf{A}_2)], \tag{3.47}$$

where the $\mathbf{A}_i$ are row vectors of $\mathbf{A}$. The analogous map $L_{Q_2} : \mathbb{P}^1 \times \mathbb{P}^1 \to \mathrm{Gr}(1, \mathbb{P}^3)$ is

$$([\mathbf{v}], [\mathbf{w}]) \mapsto [v_1 w_1 (\mathbf{B}_2 \wedge \mathbf{C}_2) - v_1 w_2 (\mathbf{B}_2 \wedge \mathbf{C}_1) - v_2 w_1 (\mathbf{B}_1 \wedge \mathbf{C}_2) + v_2 w_2 (\mathbf{B}_1 \wedge \mathbf{C}_1)]. \tag{3.48}$$

Now if $(u, (v', w'))$ is an element in $\mathbb{P}^2 \times (\mathbb{P}^1 \times \mathbb{P}^1)$, we replace Plücker coordinates in the incidence constraint $\mathrm{trace}(\mathbf{P}\mathbf{Q}^*) = 0$ with the images of u and (v', w') under (3.47) and (3.48). From this we deduce that the joint image for Q_1 and Q_2 is characterized by the multi-linear constraint

$$\begin{aligned} \textstyle\sum_{i,j,k} f_{ijk}\, u_i v'_k w'_l &= 0, \\ f_{ijk} \;=\; (-1)^{i+j+k} \cdot \det &\begin{bmatrix} \mathbf{A}_l^T & \mathbf{A}_m^T & \mathbf{B}_{\hat{j}}^T & \mathbf{C}_{\hat{k}}^T \end{bmatrix}, \end{aligned} \tag{3.49}$$

where (i,l,m), $(j,\hat{j})$, $(k,\hat{k})$ are distinct sets of indices. This condition can be identified with a $3\times2\times2$ epipolar tensor $\mathbf{F}=(f_{ijk})$, which is of course very similar to the epipolar tensor for pairs of two-slit cameras studied in Section 3.3.1, and to the traditional fundamental matrix for pinhole cameras. A direct computation with `Macaulay2` also allows us to characterize these tensors algebraically.

Proposition 3.4.2. *The variety of* $3\times2\times2$ *epipolar tensors as in* (3.49) *has codimension one in* $\mathbb{P}^{11}$*. Its defining polynomial is*

$$\begin{aligned}
&f_{111}^2f_{212}f_{221}f_{322}^2-f_{111}^2f_{212}f_{222}f_{321}f_{322}-f_{111}^2f_{221}f_{222}f_{312}f_{322}+f_{111}^2f_{222}^2f_{312}f_{321}\\
&-f_{111}f_{112}f_{211}f_{221}f_{322}^2+f_{111}f_{112}f_{211}f_{222}f_{321}f_{322}-f_{111}f_{112}f_{212}f_{221}f_{321}f_{322}+f_{111}f_{112}f_{212}f_{222}f_{321}^2\\
&+f_{111}f_{112}f_{221}^2f_{312}f_{322}+f_{111}f_{112}f_{221}f_{222}f_{311}f_{322}-f_{111}f_{112}f_{221}f_{222}f_{312}f_{321}-f_{111}f_{112}f_{222}^2f_{311}f_{321}\\
&-f_{111}f_{121}f_{211}f_{212}f_{322}^2+f_{111}f_{121}f_{211}f_{222}f_{312}f_{322}+f_{111}f_{121}f_{212}^2f_{321}f_{322}-f_{111}f_{121}f_{212}f_{221}f_{312}f_{322}\\
&+f_{111}f_{121}f_{212}f_{222}f_{311}f_{322}-f_{111}f_{121}f_{212}f_{222}f_{312}f_{321}+f_{111}f_{121}f_{221}f_{222}f_{312}^2-f_{111}f_{121}f_{222}^2f_{311}f_{312}\\
&+f_{111}f_{122}f_{211}f_{212}f_{321}f_{322}+f_{111}f_{122}f_{211}f_{221}f_{312}f_{322}-2f_{111}f_{122}f_{211}f_{222}f_{312}f_{321}-f_{111}f_{122}f_{212}^2f_{321}^2\\
&-2f_{111}f_{122}f_{212}f_{221}f_{311}f_{322}+2f_{111}f_{122}f_{212}f_{221}f_{312}f_{321}+f_{111}f_{122}f_{212}f_{222}f_{311}f_{321}-f_{111}f_{122}f_{221}^2f_{312}^2\\
&+f_{111}f_{122}f_{221}f_{222}f_{311}f_{312}+f_{112}^2f_{211}f_{221}f_{321}f_{322}-f_{112}^2f_{211}f_{222}f_{321}^2-f_{112}^2f_{221}^2f_{311}f_{322}\\
&+f_{112}^2f_{221}f_{222}f_{311}f_{321}+f_{112}f_{121}f_{211}^2f_{322}^2-f_{112}f_{121}f_{211}f_{212}f_{321}f_{322}-f_{112}f_{121}f_{211}f_{221}f_{312}f_{322}\\
&+f_{112}f_{121}f_{222}^2f_{311}^2+2f_{112}f_{121}f_{211}f_{222}f_{312}f_{321}+2f_{112}f_{121}f_{212}f_{221}f_{311}f_{322}-f_{112}f_{121}f_{212}f_{222}f_{311}f_{321}\\
&-2f_{112}f_{121}f_{211}f_{222}f_{311}f_{322}-f_{112}f_{121}f_{221}f_{222}f_{311}f_{312}-f_{112}f_{122}f_{211}^2f_{321}f_{322}+f_{112}f_{122}f_{211}f_{221}f_{311}f_{322}\\
&+f_{112}f_{122}f_{211}f_{212}f_{321}^2-f_{112}f_{122}f_{211}f_{221}f_{312}f_{321}+f_{112}f_{122}f_{211}f_{222}f_{311}f_{321}-f_{112}f_{122}f_{212}f_{221}f_{311}f_{321}\\
&+f_{112}f_{122}f_{221}^2f_{311}f_{312}-f_{112}f_{122}f_{221}f_{222}f_{311}^2+f_{121}^2f_{211}f_{212}f_{312}f_{322}-f_{121}^2f_{211}f_{222}f_{312}^2-f_{121}^2f_{212}^2f_{311}f_{322}\\
&+f_{121}^2f_{212}f_{222}f_{311}f_{312}-f_{121}f_{122}f_{211}^2f_{312}f_{322}+f_{121}f_{122}f_{211}f_{212}f_{311}f_{322}-f_{121}f_{122}f_{211}f_{212}f_{312}f_{321}\\
&+f_{121}f_{122}f_{211}f_{221}f_{312}^2+f_{121}f_{122}f_{211}f_{222}f_{311}f_{312}+f_{121}f_{122}f_{212}^2f_{311}f_{321}-f_{121}f_{122}f_{212}f_{221}f_{311}f_{312}\\
&-f_{121}f_{122}f_{212}f_{222}f_{311}^2+f_{122}^2f_{211}^2f_{312}f_{321}-f_{122}^2f_{211}f_{212}f_{311}f_{321}-f_{122}^2f_{211}f_{221}f_{311}f_{312}+f_{122}^2f_{212}f_{221}f_{311}^2.
\end{aligned}$$

Remark 3.4.3. In [181], Thirtala and Pollefeys also point that a $3\times2\times2$ tensor as in (3.49) (which, from their perspective, represents a pinhole camera and two radial cameras) must satisfy a constraint. They derive this constraint geometrically, showing in fact that any such tensor must satisfy $\det(\mathbf{W})=0$ where $\mathbf{W}=(w_{ij})$ is a 3×3 matrix with entries

$$w_{ij}=t_{i22}t_{j11}+t_{j22}t_{i11}-t_{i12}t_{j21}-t_{j12}t_{i21}.$$

One can verify that this is the same constraint as the one given in Proposition 3.4.2. We were not aware of this result at the time of our publication [149].

3.4.2 Orbits and projections of subspaces

Given any photographic camera $Q:\mathbb{P}^3\dashrightarrow\mathbb{P}^2$ or $Q:\mathbb{P}^3\dashrightarrow\mathbb{P}^1\times\mathbb{P}^1$, changing coordinates in $\mathbb{P}^3$ amounts to replacing $Q([\mathbf{x}])$ with $Q([\mathbf{Tx}])$ for some $\mathbf{T}$ in $GL(4,\mathbb{R})$. As noted for pinhole cameras in Chapter 1, understanding the action of projective transformations on cameras is important, since a coordinate frame in 3-space is not physically defined, so a projective camera model is best viewed as a $GL(4,\mathbb{R})$-orbit (or a union of orbits).[5]

[5] As elsewhere in our presentation, we consider for concreteness the action of GL rather than $\mathbb{P}GL$. Clearly, matrices that are scalar factors of each other will act the same way.

For example, given a photographic camera Q, it is natural to ask what is the associated stabilizer group $\mathrm{Stab}(Q) = \mathrm{Stab}_{GL(4,\mathbb{R})}(Q)$, that is, the set of projective transformations that leave Q fixed. It is easy to realize that $\mathrm{Stab}(Q)$ only depends on the line congruence C_Q associated with Q. Indeed, a transformation fixes Q if and only if it preserves all lines in the congruence (not necessarily point-wise).

Computationally, we observe the following fact. Assuming that Q is a photographic camera associated with a $(1, \beta)$-congruence C_Q, then:

1. If $\beta = 0$, $\mathrm{Stab}_G(Q)$ is five-dimensional in G (including the scale factor).
2. If $\beta = 1$, $\mathrm{Stab}_G(Q)$ is two-dimensional in G (including the scale factor).
3. If $\beta > 1$, $\mathrm{Stab}_G(Q)$ is one-dimensional in G (including the scale factor). In other words, the identity is the only projective transformation of $\mathbb{P}^3$ that fixes Q.

Perhaps slightly unintuitively, this implies that the parameter space of photographic cameras of higher class is actually simpler than that of pinhole and two-slit cameras, since when $\beta > 1$ we can identify every camera in the orbit of Q with a projective transformation of $\mathbb{P}^3$. On the other hand, the action of projective transformations will in general not be transitive on the set of photographic cameras of a given class (that is, two general photographic cameras will not be projectively equivalent), even among those of the same degree. This was already noted for the quadratic model of two slit cameras in Section 3.3.

Another difference between pinhole cameras and general photographic cameras is related to changes of image coordinates. Let us assume that the image of a photographic camera Q is $\mathbb{P}^2$. Then, for general projective transformation S of $\mathbb{P}^2$, the camera $x \mapsto S(Q(x))$ will *not* be equivalent to Q under a projective transformation of $\mathbb{P}^3$. One way to argue this is that, contrary to pinhole cameras, general photographic cameras will present "focal points" in the image, where the inverse projection $\mathbb{P}^2 \dashrightarrow C_Q$ is not defined. It is easy to see that the coordinates of these points must be invariant under projective transformations of $\mathbb{P}^3$; these points are in fact the "projective intrinsics" for two-slit cameras from Section 3.3.1. Similarly, if the camera Q was defined using a retinal plane H equipped a coordinate frame, then changing this plane will generally not correspond to the action of a projective transformation on Q (neither in $\mathbb{P}^3$ nor in $\mathbb{P}^2$).

Projections and Inverse Projections of Subspaces. General photographic cameras are non-linear maps, so projections and inverse projections of linear spaces will not be linear spaces in general. This is of course not true for the inverse projections of generic image points which are, by construction, lines in $\mathbb{P}^3$. For an arbitrary camera $Q : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$, with $Q = [f_1, f_2, f_3]$, a formula for the inverse projection $\mathbb{P}^2 \dashrightarrow \mathrm{Gr}(1, \mathbb{P}^3)$ can be obtained computationally by considering the ideal I in $\mathbb{Q}[x_1, x_2, x_3, x_4, u_1, u_2, u_3]$

of 2×2 minors of

$$\begin{bmatrix} f_1(\mathbf{x}) & u_1 \\ f_2(\mathbf{x}) & u_2 \\ f_3(\mathbf{x}) & u_3 \end{bmatrix}. \tag{3.50}$$

Saturating I with respect to $\langle f_1, f_2, f_3 \rangle$ yields an ideal J that contains two linear forms in x_1, x_2, x_3, x_4 with coefficients in u_1, u_2, u_3: these two linear forms describe the pre-image of a generic point $u = [u_1, u_2, u_3]$. A similar general procedure gives the inverse projection of an image point if $Q : \mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$.

If $L = \{au_1 + bu_2 + cu_3 = 0\}$ is a line in $\mathbb{P}^2$, and $Q = [f_1, f_2, f_3]$ is any rational map, then the pre-image of L in $\mathbb{P}^3$ is a surface described simply $S_L = \{af_1 + bf_2 + cf_3 = 0\}$. In the case of a camera, this surface will always be ruled, and will contain the focal locus of the associated congruence C_Q. As the L varies in $(\mathbb{P}^2)^*$, S_L will vary a two dimensional family of surfaces, or a *linear system* [13]. This is basically a "dual" viewpoint of cameras.

Finally, for a camera $Q = [f_1, f_2, f_3]$, the projection of a general line M in $\mathbb{P}^3$ will be an algebraic curve in $\mathbb{P}^2$. The equation of this image of curve can described parametrically in terms of the Plücker coordinates of M. An explicit relation can be obtained computationally as follows. We assume that we know (or have previously computed) the geometric camera associated with Q, in the form of a six vector $\mathbf{q}_x$ of polynomial expressions in $x_1, \ldots, x_4$, representing the Plücker coordinates of the image line of $x = [x_1, x_2, x_3, x_4]$. We now consider the ideal I in $\mathbb{Q}[u_1, u_2, u_3, x_1, \ldots, x_4, p_{12}, \ldots, p_{34}]$ generated by 2×2 minors of (3.50), together with the coordinates of $\mathbf{Px}$, where $\mathbf{P}$ is the Plücker matrix of variables p_{ij}. We then saturate I with respect to the ideal of 2×2 minors of $[\mathbf{q}_x \,|\, \mathbf{p}]$ where $\mathbf{p} = (p_{ij})$. Eliminating from the resulting ideal the variables x_1, x_2, x_3, x_4 yields an ideal whose generators contain a single polynomial in u_1, u_2, u_3 with coefficients in p_{ij}.

Example 3.4.4. If $Q = [f_1, f_2, f_3]$ is the twisted cubic camera described in (3.8), then the projection of a line with Plücker coordinates $p = [\mathbf{p}]$ is the conic in $\mathbb{P}^2$ defined by the equation

$$u_1^2 p_{12} - u_1 u_2 p_{13} + u_1 u_3 p_{14} + u_2^2 p_{23} - u_1 u_3 p_{23} - u_2 u_3 p_{24} + u_3^2 p_{34}. \tag{3.51}$$

Note that this system of conics is linearly parameterized by the Plücker coordinates of the line that is being projected. ♢

Photographic cameras and the Veronese embedding. A useful way to view photographic cameras using the *Veronese embedding*. We recall that the Veronese embedding $\mathcal{V}_{n,d} : \mathbb{P}^n \to \mathbb{P}^{\binom{n+d}{d}-1}$ maps a point x in $\mathbb{P}^n$ to the point in $\mathbb{P}^{\binom{n+d}{d}-1}$ whose coordinates are all possible monomials of total degree d evaluated at the coordinates of x. The image of $V_{n,d}$ is the *Veronese variety*. For example, for $n = 1, d = 3$, we have $V_{3,1} : [s, t] \mapsto [s^3, s^2t, st^2, t^3]$, and the Veronese variety is the twisted cubic in $\mathbb{P}^3$.

If $Q : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ is a photographic camera of degree d, we consider the Veronese embedding $\mathcal{V}_{3,d} : \mathbb{P}^3 \to \mathbb{P}^N$ with $N = \binom{3+d}{3} - 1$. The action of Q can always be seen as the composition of $\mathcal{V}_{3,d}$ with a *linear* projection $Q_{\mathcal{V}} : \mathbb{P}^N \dashrightarrow \mathbb{P}^2$ and, in particular, the projection $Q_{\mathcal{V}}$ uniquely determines Q. In other words, we can represent the non-linear map Q on $\mathbb{P}^3$ as a linear map $Q_{\mathcal{V}}$ on $\mathbb{P}^N$. The action of Q corresponds to the action of $Q_{\mathcal{V}}$ restricted the Veronese variety in $\mathbb{P}^N$.

Representing a non-linear map as a linear map in a higher dimensional space can be useful in practice, and indeed the Veronese embedding has been used for calibration of rational cameras [35, 76]. However, it is important to note that while all linear projections $Q_{\mathcal{V}} : \mathbb{P}^N \dashrightarrow \mathbb{P}^2$ describe rational maps $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ of degree d, from our perspective these maps will *not* be cameras in general. In fact, the condition for a rational map to be a photographic camera can be stated in terms of how the inverse projections of $Q_{\mathcal{V}}$ intersect the Veronese variety in $\mathbb{P}^N$. It would be interesting to see whether this idea can be used to directly characterize rational cameras.

3.4.3 Multifocal forms

We now present a generalization of multifocal tensors for nonlinear photographic cameras. We will make use of the general theory of multi-graded Chow forms that we introduced in [131]. This theory was already applied in Section 1.5.2 to describe traditional multifocal tensors for pinhole cameras. We refer to that section for a very brief summary on Chow and multi-graded Chow forms, and to our paper [131] for more details.

We assume that $Q_1, \ldots, Q_n$ are photographic cameras $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ (in fact, most of what follows only requires that Q_i are rational maps). We recall that the joint image variety $M_n = M(Q_1, \ldots, Q_n)$ in $(\mathbb{P}^2)^n$ is defined as the closure of the image of

$$\mathbb{P}^3 \dashrightarrow (\mathbb{P}^2)^n, \quad x \mapsto (Q_1(x), \ldots, Q_n(x)).$$

Just as for pinhole cameras, the joint image is birationally equivalent to $\mathbb{P}^3$, and thus has dimension 3, whenever at least two of the cameras are geometrically distinct. As argued in Section 3.4.1, an algebraic characterization of M_n can always be derived from the concurrent lines variety defined in Chapter 2.

On the other hand, we can also use multi-graded Chow forms to describe the joint image. As explained in Section 1.5.2, this requires choosing $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)$ so that $0 \leq \beta_i \leq 2$ and $\sum_{i=1}^n \beta_i = 4$. For any such $\boldsymbol{\beta}$, we consider

$$Z_{M_n,\boldsymbol{\beta}} = \{(L_1, \ldots, L_k) \,|\, M_n \cap (L_1 \cap \ldots \cap L_k) \neq \emptyset\} \subset \mathrm{Gr}(n-\beta_1, \mathbb{P}^2) \times \ldots \times \mathrm{Gr}(n-\beta_k, \mathbb{P}^2).$$

This set contains n-tuples $(L_1, \ldots, L_n)$, where each L_i is a linear space of codimension β_i in $\mathbb{P}^2$, and for which there exists a correspondence $(u_1, \ldots, u_n)$ in M_n with u_i in L_i. According to Theorem 1.5.5, the set $Z_{M_n,\boldsymbol{\beta}}$ has "nice" properties when $\sum_{i \in I} \beta_i \leq 3$

for all $I \subsetneq \{1, \ldots, k\}$, which means $\beta_i > 0$ for all i and $n \leq 4$. In this case, the set $Z_{M_n,\boldsymbol{\beta}}$ will be described by a single equation $F_{M_n,\boldsymbol{\beta}}$ in primal and dual coordinates image coordinates.

Just as for pinhole cameras, there are only three possibilities:

- $n = 2$, $\boldsymbol{\beta} = (2, 2)$: in this case $F_{M_n,\boldsymbol{\beta}}$ is a polynomial describing corresponding points (PP) in two images.
- $n = 3$, $\boldsymbol{\beta} = (2, 1, 1)$ (and permutations): in this case $F_{M_n,\boldsymbol{\beta}}$ is a polynomial describing "point-line-line" (PLL) correspondences across three images.
- $n = 4$, $\boldsymbol{\beta} = (1, 1, 1, 1)$: in this case $F_{M_n,\boldsymbol{\beta}}$ is a polynomial describing "line-line-line-line" (LLLL) correspondences for four images.

From the general theory of multi-graded Chow forms, we also know that $F_{M_n,\boldsymbol{\beta}}$ characterizes the joint image M_n (*i.e.*, even for $n = 3$ and $n = 4$ the form encodes point correspondences). In summary, we have the following result.

Theorem 3.4.5. *For any type of photographic cameras (rational maps)* $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$*, there exist* bifocal, trifocal and quadrifocal forms*, which describe corresponding linear spaces of type* $(PP), (PLL), (LLLL)$ *across* $2, 3, 4$ *images respectively. The knowledge of these forms is equivalent to the knowledge of the joint image* M_n $(n = 2, 3, 4)$ *of the cameras.*

The forms in Theorem 3.4.5 generalize traditional multifocal tenors in multiview geometry. An obvious difference here is the fact that multi-view forms are in general not multilinear, but only multihomogeneous in each set of image variables. On the other hand, Theorem 1.5.5 describes how the multidegree of a Chow form $F_{M_n,\boldsymbol{\beta}}$ (as a multihomogeneous polynomial) can be computed from the multidegree of the M_n (as a variety in $(\mathbb{P}^2)^n$).[6] Recent results by Escobar and Knutson on the multidegree of multi-image varieties [52] might lead to a general formula for the multidegree of generalized joint images, and for the degrees of the associated multifocal forms. For the moment, the multidegree of M_n can be computed directly in simple cases.

Example 3.4.6. Consider the following three maps $Q_i : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$:

$$Q_1(x) = \begin{bmatrix} x_1 + x_4 \\ x_2 + x_4 \\ x_3 + x_4 \end{bmatrix}, \; Q_2(x) = \begin{bmatrix} x_1^2 + 2x_1x_3 + x_3^2 \\ x_1x_3 + x_1x_4 + x_3x_4 + x_4^2 \\ x_1^2 + x_1x_3 + x_1x_4 + x_3x_4 \end{bmatrix}, \; Q_3(x) = \begin{bmatrix} x_1^3 - x_1x_2x_4 \\ x_1^2x_2 + x_2^3 - x_2^2x_3 \\ x_1^2x_2 - x_2^2x_4 \end{bmatrix}.$$

These are three photographic cameras: Q_1 is a pinhole $(1, 0)$-camera, Q_2 is a two-slit $(1, 1)$-camera, and Q_3 is a $(1, 2)$-camera. These examples were chosen to have disjoint focal loci. Using the command `multidegree` in `Macaulay2`, we can verify that the multidegree of the joint image is

$$2T_1^2T_2 + 3T_1^2T_3 + T_1T_2^2 + 6T_1T_2T_3 + T_1T_3^2 + 3T_2^2T_3 + 2T_2T_3^2. \tag{3.52}$$

[6] Unfortunately the term "multidegree" has different meanings which are both necessary in this setting.

According to Theorem 1.5.5, the trifocal form for these cameras for $\boldsymbol{\beta} = (2,1,1)$ has degrees (d_1, d_2, d_3) given by the coefficients for degrees $(1,1,1), (0,2,1), (0,1,2)$ in (3.52), so $(d_1, d_2, d_3) = (6,3,2)$. Indeed, a computation shows that the trifocal form for these cameras is a polynomial with 579 terms that starts with

$$T = u_1^6 r_1^3 s_1^2 - 2u_1^5 u_3 r_1^3 s_1^2 - 8u_1^4 u_2 u_3 r_1^3 s_1^2 + \ldots + 9u_1^3 u_3^3 r_3^3 s_3^2.$$

Here the variables u_1, u_2, u_3 are projective coordinates representing points in the first image, while r_1, r_2, r_3 and s_1, s_2, s_3 are dual projective coordinates representing lines in the second and third image. ♢

Finally, classical multifocal tensors uniquely encode a configuration of pinhole cameras, and it is natural to ask whether this important property holds for non-linear multifocal forms as well. To address this question we need to introduce an appropriate "configuration space" for general rational cameras. For this we consider a set $\mathcal{Q}_1, \ldots, \mathcal{Q}_n$ of *projective orbits* of photographic cameras (*i.e.*, $\mathcal{Q}_i = \{Q_i T \,|\, T \in \mathbb{P}GL(4,\mathbb{R})\}$ for some camera Q_i), which we view as n fixed "camera models". We then define the set of *configurations* for these models as

$$\mathcal{C}(\mathcal{Q}_1, \ldots, \mathcal{Q}_n) = (\mathcal{Q}_1 \times \ldots \times \mathcal{Q}_n)/\mathbb{P}GL(4,\mathbb{R}).$$

This generalizes the space of configurations of pinhole cameras. Indeed, a configuration is always defined as the product of independent projective orbits, up to global projective equivalence.

It is easy to see that if $(Q_1, \ldots, Q_n)$ and $(Q'_1, \ldots, Q'_n)$ are projectively equivalent n-tuples, the joint images $M(Q_1, \ldots, Q_n)$ and $M(Q'_1, \ldots, Q'_n)$ are the same. This is because a joint image is independent of the choice of coordinates in $\mathbb{P}^3$ (see also Proposition 1.3.15 for the case of pinhole cameras). This implies that we can associate a configuration in $\mathcal{C}(\mathcal{Q}_1, \ldots, \mathcal{Q}_n)$ with a joint image, *i.e.*, there is a map

$$\begin{aligned} \mathcal{J}(\mathcal{Q}_1, \ldots, \mathcal{Q}_n) : \mathcal{C}(\mathcal{Q}_1, \ldots, \mathcal{Q}_n) &\to \{\text{varieties in } (\mathbb{P}^2)^n\}, \\ [Q_1, \ldots, Q_n] &\mapsto M(Q_1, \ldots, Q_n), \end{aligned} \tag{3.53}$$

where $[Q_1, \ldots, Q_n]$ denotes the configuration (equivalence class) of $(Q_1, \ldots, Q_n)$. If all $\mathcal{Q}_i$ are pinhole camera models, this association is injective (Theorem 1.3.16). In general, this may not be the case: indeed, if Q_1, Q_2 are a pair of two-slit cameras, then we have already noted that the joint image $M(Q_1, Q_2)$ corresponds in general to *two* projective configurations (Theorem 3.3.7). The following, however, is true.

Proposition 3.4.7. *Let $Q_1, \ldots, Q_k$ be photographic cameras where k is either 2, 3 or 4. We write of $\mathcal{Q}_1, \ldots, \mathcal{Q}_k$ for the corresponding "projective camera models".*

1. *Any multifocal form (as defined in Theorem 3.4.5) is invariant to changes of projective coordinates in $\mathbb{P}^3$.*

2. *If the map* (3.53) *is injective for the camera models* $\mathcal{Q}_1, \ldots, \mathcal{Q}_k$*, then the multifocal form uniquely represents* $[Q_1, \ldots, Q_k]$ *inside the configuration space* $\mathcal{C}(\mathcal{Q}_1, \ldots, \mathcal{Q}_k)$.

Proof. Both statements follow immediately from the fact that the knowledge of a multi-view form is equivalent to the knowledge of the the joint image. □

We conjecture that apart from the case where $n = 2$ and Q_1, Q_2 are two-slit cameras, the map (3.53) is usually injective.

3.5 Conclusions

We have introduced a general model of photographic cameras, that are described by rational functions, and that are associated with congruences of order one. Among the simplest examples of this model are two-slit cameras viewed as bilinear projections $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$. We have studied such cameras, analyzing the corresponding projective orbits, and describing the associated epipolar geometry. A similar methodology can in principle be applied to any photographic camera. While cameras described by rational mappings have previously been considered [35, 76], these works did not explicitly require image points to be projections of 3D lines.

Overall, the goal of this chapter was to provide a framework for studying properties of general photographic devices in a unified way. In particular, we believe that it is important to distinguish between two main parts of the imaging process: the "geometric" part, which characterizes the set of light rays that are recorded by the camera, and the "analytic" part, that describes how these rays are mapped to image coordinates. Moreover, it is useful to emphasize and study the action of (subgroups of) projective transformations on classes of cameras, particularly for the purpose of parameterizing camera configurations in multi-view reconstruction. For pinhole cameras, for example, affine changes of image coordinates have the same effect as a projective transformation of space. As argued in Section 3.4.2, this property is not true for arbitrary photographic cameras, and this makes the projective framework arguably less useful in practice: in this sense, for calibrating systems of general cameras, solving for euclidean motions using the method described by Sturm [172] is probably simpler than applying projective methods. On the other hand, using a projective framework is much more convenient for understanding the geometry of rational mappings and of the corresponding cameras.

In the future, we hope to understand the relation between configurations of photographic cameras and the corresponding joint image varieties, as discussed at the end Section 3.4.3. Although the use of trifocal and quadrifocal forms for camera calibration would probably be very difficult (given the large number of coefficients that would need to be estimated), whether or not a joint image determines a unique configuration of cameras is a central theoretical question for multi-view reconstruction. On the more

"practical" side, we would like to extend our photographic model to include *non-rational* mappings, particularly to be able to study catadioptric imaging systems. As explained in the previous chapter, this might require deriving an effective description for the action of refractions and reflections on line congruences.

Chapter 4

Consistency of Image Sets and Visual Hulls

This chapter studies the geometric relationship between 3D objects and their projections in multiple images. For most our discussion, we let an object be an arbitrary subset S of $\mathbb{P}^3$, so that its images for n different cameras are subsets $T_1, \ldots, T_n$ of $\mathbb{P}^2$. Being projections of the same object in space imposes constraints on image sets, leading to a notion of *geometric consistency* that resembles (and generalizes) the traditional "correspondence" for image points. Conversely, image sets can be "triangulated" to a 3D shape, that is known as the *(finite) visual hull*. We study the theory of consistency and projective visual hulls, pointing out connections with line geometry and duality. Some of these ideas are put into practice in a simple algorithm for computing the visual hull.

The material in this chapter is based on the following publications:

– Matthew Trager, Martial Hebert, and Jean Ponce. "Consistency of silhouettes and their duals". In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2016.

– Boris Bukh, Xavier Goaoc, Alfredo Hubard, and Matthew Trager. "Consistent sets of lines with no colorful incidence". In: *International Symposium on Computational Geometry*. 2018.

4.1 Introduction

What images can be obtained by taking pictures of the same 3D object? Conversely, what are the 3D objects that can produce a given set of images? In this chapter, we investigate this type of questions from a purely geometric standpoint, disregarding pho-

tometric information, and treating objects and their images as simple "sets" in projective spaces.

We begin our discussion by introducing a simple notion of "geometric consistency" for image sets. More precisely, for fixed cameras $P_1, \dots, P_n$, we say that sets $T_1, \dots, T_n$ in $\mathbb{P}^2$ are consistent if there exists a set S in $\mathbb{P}^3$ such that $P_i(S) = T_i$ (Figure 4.1). This setup can be seen as an extension of classical multi-view geometry, which provides conditions for *points* (and sometimes *lines*) to be consistent (*i.e.*, to correspond) in terms of given camera projections [53, 78]. This also expands an analogy initiated by the "generalized epipolar constraint" introduced in [8], that essentially characterizes consistency for two image sets. In our study, we point out several results that hold for arbitrary image sets (or sometimes convex image sets), that have identical counterparts in the theory of point correspondences.

Given a family of consistent image sets, the largest set in space that projects onto them is the associated *visual hull* [12]. In the literature, the term "visual hull" has also been used for another related but different concept: for Bottino and Laurentini [108], the visual hull of a set S in space is the largest set whose images from *all* viewpoints are the same as for S. In our presentation, we differentiate between these two notions, calling them the "finite" and the "complete" visual hull, respectively. After introducing both types of visual hulls in a very general setting, we point out a close analogy with (projective) convexity. Loosely speaking, the the visual and convex hull of a set in space can both be obtained from exactly the same construction, with the latter using "line-centered" projections $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1$ instead of pinhole projections.

Regarding the relationship between consistency and visual hulls, we note that a family of consistent image sets always provides a *complete*, albeit *implicit*, representation of a finite visual hull. This is similar to the idea of representing 3D points by their projections, or, equivalently, identifying 3D space with the joint image (see Section 1.3). Moreover, this viewpoint suggests a general strategy for investigating three-dimensional geometric features directly from image data. For example, the combinatorial components that have previously been defined for finite visual hulls ("cone strips", "intersection curves" and "triple points" [57]) admit simple formal descriptions in image space. This leads to a simple but effective strategy for computing an explicit representation of the boundary of the finite visual hull (an arbitrarily dense point cloud, that can be turned into a mesh using standard geometry processing algorithms [14]).

The definitions of visual hulls and image consistency do not require metric information, which is why we adopt a projective setting. This eliminates various degenerate situations and allows, for example, to unify the cases of orthographic and perspective projections (as in traditional multi-view geometry).[1] Another advantage of the projective language

[1] The most significant difference between the euclidean and projective frameworks in our setting is that in the latter case visual cones are *two-sided*. An alternative (but perhaps less natural) approach would have been the use of *oriented projective geometry* [170].

is that it provides homogeneous formulations for *duality* and *line geometry*. In particular, we can exploit the fact that perspective projections are related to *planar sections* in the dual space, to define a very natural "dual" notion of consistency, expressing conditions for a family of planar sets to be sections of the same object. Finally, we point out that for most of our presentation we do not use any regularity assumption on the observed sets (silhouettes) and 3D shapes: similar to "volumetric approaches" in shape-from-silhouettes [121, 34, 168], this bypasses the difficulties that arise when considering an idealized setting for smooth surfaces.

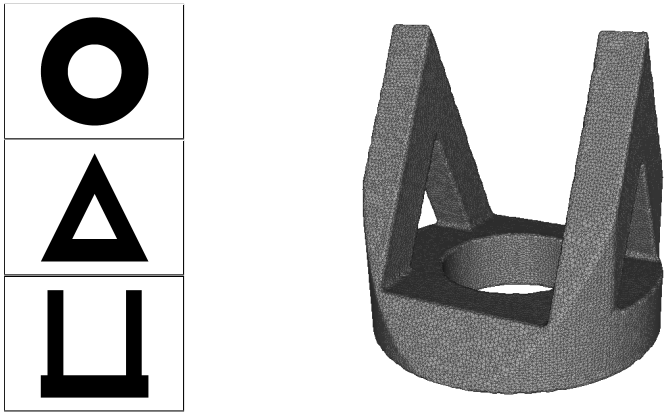


Figure 4.1: Geometrically consistent silhouettes are feasible projections of a 3D object.

4.1.1 Previous work

Consistency and duality. The most widespread application of consistency (for silhouettes) has been for designing alternatives to point-based methods for camera calibration, required for dealing with smooth and textureless surfaces. Indeed, ever since the seminal work of Rieger [156], the problem of estimating camera motion or calibration parameters using only silhouettes has received considerable attention: see, *e.g.*, [24, 62, 81, 94, 125, 167, 195]. Albeit with some variations, all these methods exploit (more or less directly) the geometric constraints provided by the *epipolar tangencies* [8]. In addition to camera calibration, silhouette consistency has been enforced explicitly for other tasks; for example multi-view segmentation [29], or 3D-reconstruction [41] and recognition [111]. Another interesting "artistic" application is discussed in [127]. There exists limited theoretical work on silhouette consistency, and it is always restricted to special situations. In particular, the problem of determining whether a family of silhouettes can correspond to a real object is considered in [19, 91, 196], but results are only given for the case of orthographic projections and somewhat restricted camera motion. Some theoretical properties, such as the fact that epipolar tangency conditions do not imply global consistency, can be found in [24, 33].

The duality between projections and planar sections (in 3-space) is well known in a variety of different settings. For orthographic projections, this principle is most naturally described for smooth surfaces (taking sections of the Gauss sphere [191]), or for convex bodies (using convex polarity [161]). For perspective projections, it is typical to consider smooth or algebraic surfaces, and define the dual surface as the union of tangent planes at smooth points inside the dual projective space $(\mathbb{P}^3)^*$. The convex projective setting discussed here is probably less common, although it is very similar to the polarity defined in terms of *oriented* projective geometry in [170]. Despite the use of different mathematical languages (and different hypotheses, particularly on the smoothness of the projected object), these approaches are all very closely related.

Visual hulls. The idea of approximating a shape by intersecting the visual cones associated with many silhouettes was, remarkably, already popular in the 1800s [171]. In the modern literature, the problem was first addressed in 1974 in Baumgaurt's PhD thesis [12]. Bottino and Laurentini studied the "complete" visual hull for polyhedral [108], smooth [21], and piecewise smooth [20] surfaces in $\mathbb{R}^3$. They showed for example that the boundary of the complete visual hull of a smooth object contains ruled surface patches, corresponding to two types of "event surfaces" ("tritangent" and "edge") [21]. This viewpoint is actually closer to the topic of the next chapter, where we will study projections of smooth (algebraic) surfaces as "contours"; here we focus on more primitive set-theoretic relations between sets of points, lines, and planes.

The concept of "image-based" visual hulls for reconstruction was introduced by Matusik *et al.* [123], who proposed an efficient method for rendering visual hulls without explicitly constructing a 3D model, exploiting instead epipolar geometry to reproject points across different images. A limitation of this method is that it is view-dependent, *i.e.*, rendering a novel view requires running the entire algorithm again. However, the basic ray casting procedure used to compute "intersection intervals" along visual rays (see Section 4.6.2), has inspired subsequent work for computing explicit models of the visual hull by similarly reducing 3D computations to 2D [60, 59, 122]. Our strategy for recovering a point cloud representation of the visual hull also exploits the image-based procedure from [123]; however, compared to [60, 59, 122], it is simpler, since it does not deal with the problem of recovering the polyhedral structure, and because it exploits image contours *as well as* silhouettes viewed as regions.

The combinatorial structure of the finite visual hull was noted by Lazebnik, Boyer, and Ponce in [109]. Moreover, similar to our projective framework, Lazebnik, Furukawa, and Ponce [110] use *oriented projective geometry* for dealing with visual hulls. Contrary to [110], however, our approach does not require a global notion of orientation, since "region correspondences" across multiple images are well defined in purely projective spaces.

4.1.2 Main contributions

The following is a brief summary of the main contributions in this chapter.

- We study a notion of *geometric consistency* for arbitrary image sets and projective cameras. We also define a more general form of k-consistency, and investigate the relationship between pairwise ($k = 2$) consistency, epipolar tangencies, and global consistency (Sections 4.2.2 and 4.2.3).
- We restate the notion of consistency in terms of duality, expressing the condition for planar sets to be sections of the same object. For convex silhouettes, we show that the dual of the visual hull coincides with the convex hull of the dual image of the silhouettes (Proposition 4.3.6).
- We describe the "complete" and the "finite" visual hulls for arbitrary sets in projective space. We point out connections with notions of convexity and duality in $\mathbb{P}^n$ and $\mathrm{Gr}(1, \mathbb{P}^3)$.
- We present an image-based description of the combinatorial features of the finite visual hull, and propose a simple but efficient method for recovering an arbitrarily dense point cloud on the boundary of the visual hull. We also introduce a "hardness" function that can be computed directly from image data and provides an estimate of the quality of the visual hull reconstruction.

Conventions. Throughout the chapter, we let $\mathbb{P}^n = \mathbb{P}(\mathbb{R}^{n+1})$. For any pinhole camera $P : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ and set of points T in $\mathbb{P}^2$, the associated *visual cone* is $P^{-1}(T)$, where P^{-1} denotes the pre-image set (this definition excludes the pinhole from the visual cone, however this will be irrelevant for our discussion).

4.2 Consistency of Image Sets

In this section, we introduce our notion of geometric consistency for arbitrary sets in different images. The definition we give is very natural, and similar concepts have previously been used to introduce "incoherence" measures for silhouettes [24, 81]. After presenting some basic definitions in Section 4.2.1, we consider in Section 4.2.2 the important case of two silhouettes, pointing out how the consistency condition is basically equivalent to the popular "epipolar tangency" constraint, but only applied to *extremal* tangents. We then investigate in more detail the difference between pairwise consistency and general consistency in Section 4.2.3. We conclude in Section 4.2.4 with a discussion on the topology of the visual hull for two views, which further clarifies the relationship between consistency and epipolar tangents.

4.2.1 Basic definitions

Let $P_1, \ldots, P_n$ be n projective cameras $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ with distinct pinholes $c_1, \ldots, c_n$, and let $T_1, \ldots, T_n$ be a family of sets in $\mathbb{P}^2$, which we view as belonging to different images. We allow these sets to be completely arbitrary: for example, each set T_i could be a finite collection of points, curves or regions in each image. For each $i = 1, \ldots, n$, we write $C_i = P_i^{-1}(T_i)$ for the visual cone associated with T_i.

Definition 4.2.1. The sets $T_1, \ldots, T_n$ are *consistent* (relative to the cameras $P_1, \ldots, P_n$) if there exists a non-empty set $R \subseteq \mathbb{P}^3 \setminus \{c_1, \ldots, c_n\}$ such that $P_i(R) = T_i$ for all $i = 1, \ldots, n$.

If each of $T_1, \ldots, T_n$ is a single point, then consistency reduces to the usual notion of point correspondence (see Definition 1.3.1). Extending this analogy, consistent image sets can be seen as n-tuples of projections of all possible objects in space. Consistency can be equivalently defined using the joint image variety $M(P_1, \ldots, P_n)$ in $(\mathbb{P}^2)^n$ for the cameras $P_1, \ldots, P_n$ (see Definition 1.3.4). More precisely, $T_1, \ldots, T_n$ are consistent if and only if the set

$$R' = M(P_1, \ldots, P_n) \cap (T_1 \times \ldots \times T_n) \subset (\mathbb{P}^2)^n$$

projects surjectively onto each T_i. Note that, contrary to the case of point correspondences, the set R in Definition 4 is in general not unique. However, we can always associate $T_1, \ldots, T_n$ with the *maximal* set that projects onto them. This set is given by $H = \bigcap_i C_i$ where each $C_i = P_i^{-1}(T_i)$ is a visual cone.

Definition 4.2.2. If $T_1, \ldots, T_n$ are consistent for the cameras $P_1, \ldots, P_n$, then $H = \bigcap_i C_i$ is the *finite visual hull* associated with $T_1, \ldots, T_n$.

It is actually customary to define the visual hull simply as the intersection of the cones $\bigcap_i C_i$ for arbitrary, not necessarily consistent sets (or "silhouettes"). We will discuss this ambiguous terminology (and explain the meaning of the word "finite") later in the chapter. Here we only note that if $\bigcap_i C_i$ is a non-empty intersection of arbitrary visual cones, then it can be viewed as the visual hull associated with the subsets $\tilde{T}_i = P_i(\bigcap_j C_j) \subseteq T_i$, which will always be consistent. In fact, consistency is clearly equivalent to the fact that $\tilde{T}_i = P_i(\bigcap_j C_j) = T_i$ for all $i = 1, \ldots, n$, or to $T_i \subseteq P_i(\bigcap_j C_j)$, since the opposite inclusion is always true. We collect a few other simple but useful properties.

Proposition 4.2.3. *Let $T_1, \ldots, T_n$ be arbitrary sets in $\mathbb{P}^2$.*

1. *$T_1, \ldots, T_n$ are consistent if and only if for each $i = 1, \ldots, n$, and for all $u_i \in T_i$, the visual ray $P^{-1}(u_i)$ intersects $\bigcap_{j \neq i} C_j$.*

2. *$T_1, \ldots, T_n$ are consistent if and only if*

$$T_i \subseteq P_i\left(\bigcap_{j\neq i} C_j\right), \quad \forall i \in \{1, \ldots, n\}. \tag{4.1}$$

   See Figure 4.2.

3. *If $T_1, \ldots, T_n$ are consistent, then any subfamily $T_{i_1}, \ldots, T_{i_s}$ is consistent (for the associated cameras $P_{i_1}, \ldots, P_{i_s}$).*

Proof. The first property follows from the fact that $T_i \subseteq P_i(\bigcap_j C_j)$ can be expressed as $P_i^{-1}(u_i) \cap \bigcap_j C_j \neq \emptyset$ for all $u_i \in T_i$, which in turn is equivalent to $P_i^{-1}(u_i) \cap \bigcap_{j\neq i} C_j \neq \emptyset$, since $P_i^{-1}(u_i) \subseteq C_i$. The second and third properties are consequences of the first one. □

The last point in this result suggests the following natural definition.

Definition 4.2.4. The sets $T_1, \ldots, T_n$ are k-consistent (for some $k \leq n$) if every subfamily of sets $T_{i_1}, \ldots, T_{i_k}$ is consistent.

Proposition 4.2.3 implies that consistency (or n-consistency) implies k-consistency for every k. Much of our discussion in the following sections will be aimed at understanding whether some version of the converse holds, *i.e.*, if k-consistency can be used to deduce consistency.

This might be a good moment to point out that the notion of geometric consistency discussed in this chapter is somewhat independent from a more intuitive (but less formal) concept of "similarity" of appearance. Indeed, consistent silhouettes may actually look completely different (as in the example in Figure 4.1), while, on the other hand, almost identical silhouettes may be geometrically inconsistent. Thus, the concept might be well suited for being used alongside more traditional feature-based methods for recognition.

Finally, for the rest of this section we will make the following assumption for all n-tuples of closed image sets $T_1, \ldots, T_n$ and cameras $P_1, \ldots, P_n$:

(A) *For each camera center c_i and visual cone C_j, with $i \neq j$, c_i does not belong to C_j.*

This condition is useful for excluding various degenerate situations.[2]

[2]This condition is not actually necessary for all of our results, and weaker assumptions may often be considered. However, for the sake of simplicity, we give a single condition that is valid throughout the section.

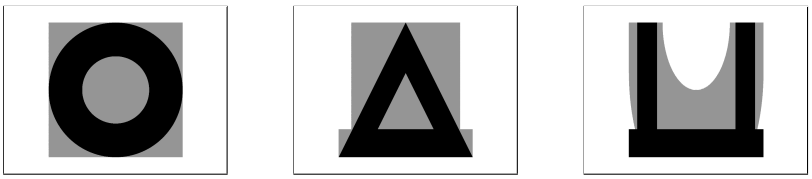


Figure 4.2: The silhouettes from Figure 4.1 are consistent (relative to three orthogonal orthographic projections), since each silhouette is contained in the reprojection (in gray) of the intersected visual cones associated with the other views. See Proposition 4.2.3.

4.2.2 Pairwise consistency

Let us assume that we are given only two image sets T_1, T_2 and two cameras P_1, P_2. According to Proposition 4.2.3, the sets T_1, T_2 are consistent if and only if

$$T_1 \subseteq P_1(C_2), \quad \text{and} \quad T_2 \subseteq P_2(C_1). \tag{4.2}$$

In other words, we require for each set to be contained in the projection of the visual cone associated with the other one. Pairwise consistency is closely related to the popular *epipolar tangency constraint* [8, 195]. Indeed, we can restate the condition (4.2) in terms of epipolar geometry as follows. We recall here that an "epipolar line" in an image is any line that contains an epipole (the the projection of the other pinhole); the set of epipolar lines in the two images are in "epipolar correspondence", *i.e.*, they are related by a fixed homography $\mathbb{P}^1 \to \mathbb{P}^1$.

Proposition 4.2.5. *Two image sets T_1, T_2 are consistent if and only if the set of epipolar lines in the first image which intersect T_1 is in epipolar correspondence with the set of epipolar lines in the second image which intersect T_2.*

Proof. The statement can be seen as a consequence of the first property in Proposition 4.2.3. In fact, the epipolar correspondence condition guarantees that for all $i = 1, 2$, and for every point $u_i \in T_i$ there exists at least one corresponding point $u_j \in T_j$ $(j \neq i)$, so that triangulating all pairs of associated points (*i.e.*, intersecting the cones C_1, C_2) we obtain a set $R \subseteq \mathbb{P}^3 \setminus \{c_1, c_2\}$ that projects exactly onto T_1 and T_2. Note that assumption A is equivalent to the fact that, in each image, the epipole lies outside of the given set. □

If we assume that T_1, T_2 are connected regions bounded by smooth curves, then pairwise consistency essentially reduces to the fact that *extremal epipolar tangents* (*i.e.*, "outermost" epipolar lines that are tangent to the contours) are in epipolar correspondence [24]. See Figure 4.3. On the other hand, pairwise consistency *does not* require non-extremal epipolar tangents to be matched. In fact, perhaps somewhat unintuitively, the visual hull generated by two silhouettes with non-extremal epipolar tangencies (matched or

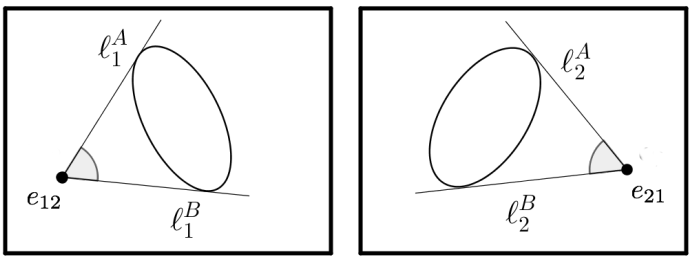


Figure 4.3: Parwise consistency: the sets of epipolar lines intersecting each silhouette must be in epipolar correspondence. In particular, the *extremal lines* lines ℓ_1^A, ℓ_2^A and ℓ_1^B, ℓ_2^B must correspond.

unmatched) will always present a complicated topology (see Figure 4.4). The difference between the two examples in the figure is that in the case of unmatched internal epipolar tangencies (*top*), *any* object which is consistent with the silhouettes must project with occlusions (because the inconsistent "branches" are necessarily projections of different parts of the observed object); on the other hand, when internal epipolar tangencies are matched (*bottom*), there exists an object (different from the visual hull) that projects onto the silhouettes without occlusions. We refer to Section 4.2.4 for a more detailed discussion.

Returning to the case of an arbitrary number of sets, we have previously observed that if $T_1, \ldots, T_n$ are consistent, then they are 2-consistent, or "pairwise consistent" (see Definition 4.2.4). The converse however is not true: this was pointed out in [24, 33], and is illustrated by our example in Figure 4.6. On the other hand, the following result shows that pairwise consistency implies an "approximate" version consistency in the case of three sets. This can be seen as a generalization of the fact that three non-coplanar visual rays that intersect pairwise are all concurrent.

Proposition 4.2.6. *Let T_1, T_2, T_3 be connected closed sets that are pairwise consistent. If the pinholes of the cameras c_1, c_2, c_3 are not collinear, and if each visual cone C_i does not intersect the plane spanned by c_1, c_2, c_3, then $\bigcap_i C_i$ is not empty.*

Proof. It is sufficient to prove that, say, $T_1 \cap P_1(C_2 \cap C_3)$ is not empty. Let $R = P_1(C_2) \cap P_1(C_3)$. The assumptions on the pinholes guarantee that R is a quadrilateral (it is the intersection of two connected projected cones; see Figure 4.5). From pairwise consistency (4.2), we know that T_1 is tightly "inscribed" in R, meaning that $T_1 \subseteq R$ and T_1 intersects of the four edges of R. The same holds for $P_1(C_2 \cap C_3)$: indeed $P_1(C_2 \cap C_3) \subseteq P_1(C_2) \cap P_1(C_3)$ holds, and $P_1(C_2 \cap C_3)$ intersects all edges of R since C_2, C_3 are themselves pairwise consistent (an extremal epipolar line in $P_1(C_2)$ is the projection of a line in C_2 which must intersect $C_2 \cap C_3$). The claim follows from continuity arguments: for example, by considering paths in T_1 and $P_1(C_2 \cap C_3)$ connecting different pairs of opposite edges. □

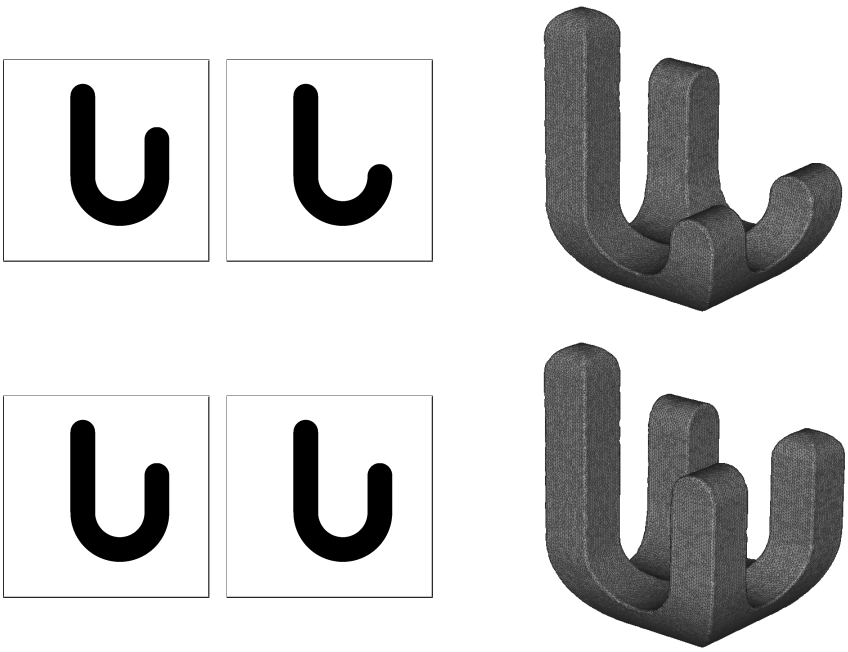


Figure 4.4: The visual hull generated by two "hook-shaped" silhouettes always presents self-occlusions, whether the internal epipolar tangents are unmatched (*top*) or matched (*bottom*). In these examples we considered orthogonal orthographic projections, but the behavior is completely general.

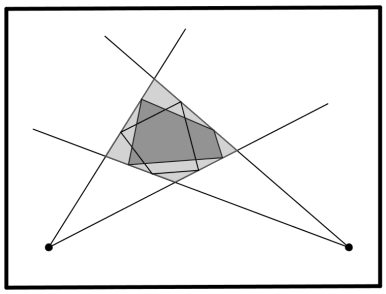


Figure 4.5: Proof of Proposition 4.2.6. The set T_1 (dark gray) and $P_1(C_2 \cap C_3)$ are both "inscribed" in the quadrilateral $P_1(C_2) \cap P_1(C_3)$ (light gray), and must thus intersect.

4.2.3 From pairwise to general consistency

We have seen that pairwise consistency does not imply general consistency [24, 33]: in fact, as shown by the example in Figure 4.6, the two notions are not equivalent even in the restricted case of convex image sets (contrary to a claim in [24]). Interestingly, however, we can show that for convex sets (*i.e.*, sets that are convex in an affine chart: see Section 4.3 for a discussion on projective convexity) general consistency is actually implied by 3-consistency. A similar fact was pointed out in [196].

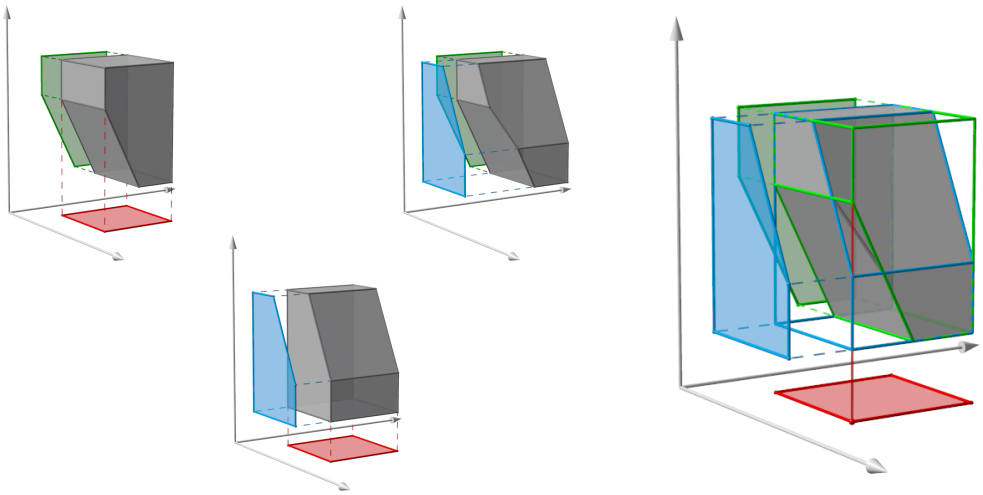


Figure 4.6: Three silhouettes that are 2-consistent but not globally consistent for three orthogonal projections. Each of the first three figures shows a three-dimensional set that projects onto two of the three silhouettes. The fourth figure illustrates that no set can project simultaneously onto all three silhouettes: the highlighted red image point cannot be lifted in 3D, since no point that projects onto it belongs to the pre-images of both the blue and green silhouettes.

Proposition 4.2.7. *If $T_1, \dots, T_n$ are convex sets that are 3-consistent (*i.e.*, T_i, T_j, T_k are consistent for every $\{i, j, k\} \subseteq \{1, \dots, n\}$) then $T_1, \dots, T_n$ are also consistent.*

Proof. From Proposition 4.2.3, it is sufficient to prove that for every $i = 1, \dots, n$, and for all $u_i \in T_i$, we have $P^{-1}(u_i) \cap \left(\bigcap_{j \neq i} C_j\right) \neq \emptyset$. Because of convexity, the sets $P_i^{-1}(u_i) \cap C_j$ are intervals; moreover, they intersect pairwise because of the assumption of 3-consistency: this implies that they all intersect.[3] See Figure 4.7. □

[3] To be precise, it should be noted that, although in projective space, none of the cones C_j for $j \neq i$ contain c_i (because of our assumption (A)). This means that we can treat the sets $P_i^{-1}(u_i) \cap C_j$ as intervals on an affine line.

This statement closely resembles "Helly-type" theorems in computational geometry [184], and is a generalization of the fact that point correspondence is always implied by triplet-wise point correspondence. An inspection of the proof also shows that it is not actually necessary for the sets T_i to be convex, but only for their intersections with epipolar lines (rather than all lines) to be intervals.

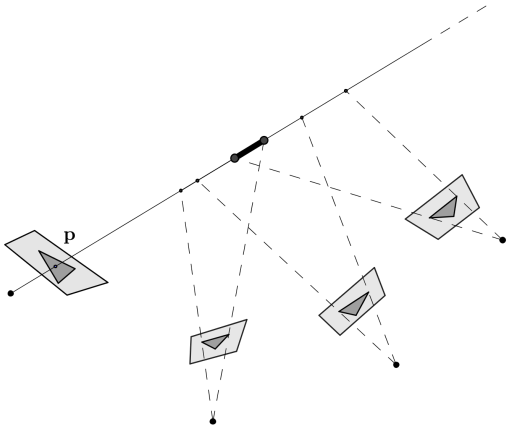


Figure 4.7: Proof of Proposition 4.2.7: the visual ray corresponding to p meets the other visual cones in intervals that intersect pairwise, and thus must all intersect.

Returning to the case of arbitrary image sets, we see from (4.2) that pairwise consistency is equivalent to

$$T_i \subseteq \bigcap_{j \neq i} P_i(C_j), \quad i = 1, \ldots, n. \tag{4.3}$$

Moreover, if the sets T_i are consistent there is a chain of inclusions $T_i \subseteq P_i\left(\bigcap_{j\neq i} C_j\right) \subseteq \bigcap_{j\neq i} P_i(C_j)$. In fact, as pointed out in [24], pairwise and general consistency can be seen as the same formal condition except for the inverted order of the projection and intersection of the visual cones.

It is also natural to ask how consistency constrains the shape of the image sets. For this, we observe that if $T_1, \ldots, T_n$ are consistent for $P_1, \ldots, P_n$, then a set, say T_1, may be replaced by *any* strictly larger one $\tilde{T}_1$ without affecting consistency, provided that $\tilde{T}_1 \subseteq P_1\left(\bigcap_{j\neq 1} C_j\right)$: indeed, this guarantees that condition (4.1) remains satisfied in each image. For example, as shown in Figure 4.8, we are allowed to modify the shape of a set T_i with a small protrusion, in the neighborhood of any point that does not belong to the boundary of $P_i\left(\bigcap_{j\neq i} C_j\right)$. On the other hand, points belonging to the intersection of T_i with the boundary of $P_i\left(\bigcap_{j\neq i} C_j\right)$ are more constrained: we will refer to these points as "tangential points", since the associated visual rays are "tangent" to the visual hull associated with the remaining silhouettes $\bigcap_{j\neq i} C_j$. A full understanding of tangential

points and their relationship with consistency requires further study. For the moment, we point out that the visual ray corresponding to a tangential point will typically graze the surface of one of the remaining visual cones, so that the ray will correspond to an epipolar tangent (not necessarily extremal) for the associated image (Figure 4.9, *left*). In general, it is also possible for the ray to intersect the set $\bigcap_{j\neq i} C_j$ only at an "intersection curve" [110] (so that it is not tangent to any cone) giving rise to a "tangential triple point" on the visual hull surface (Figure 4.9, *right*). However, it is not hard to realize that this second case will not occur for generic projections of smooth solids, because points on a smooth surface are never visible from three generic viewing directions.

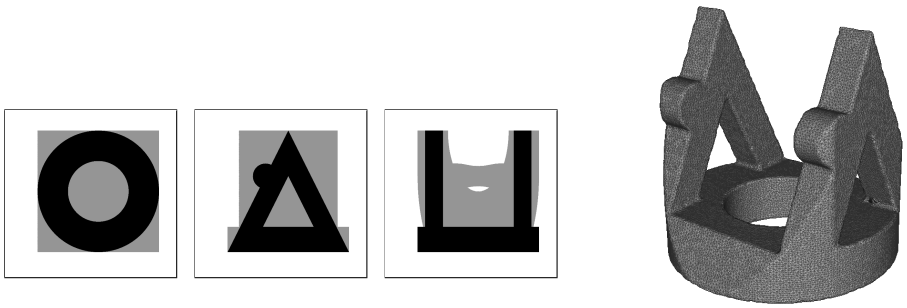


Figure 4.8: If we modify the second silhouette from Example 4.1 so that the new silhouette is still contained in the reprojection of the visual hull associated to the other views, consistency is preserved. Note that the reprojection onto the third view is not the same as in the previous case, however condition (4.1) remains satisfied since the new reprojected set is larger.

4.2.4 Topology of the visual hull from two views

In this section, we study the topology of the visual hull arising from two consistent closed sets. This clarifies our discussion in Section 4.2.2, particularly in reference to Figure 4.4. The main tool for our analysis will be the *(circular) Reeb graph* [16].

Definition 4.2.8. Let X be a topological space, and let $f : X \to S^1$ be a continuous function, where S^1 denotes unit circle. Define an equivalence relation $\sim$ on X where $p \sim q$ whenever p and q belong to the same *connected component* of a level set $f^{-1}(c)$ for some $c \in S^1$. The *circular Reeb graph* is defined as the quotient space $X / \sim$ endowed with the quotient topology.[4]

In practice, the circular Reeb graph is obtained simply by contracting every connected component of a level set to a point: see Figure 4.10. For general functions, the (circular)

[4]Traditionally, the Reeb graph is defined for real-valued functions $f : X \to \mathbb{R}$, however for our purposes the "circular" definition is more natural. Note however any function $f : X \to S^1$ that is not surjective may actually be viewed as real-valued.

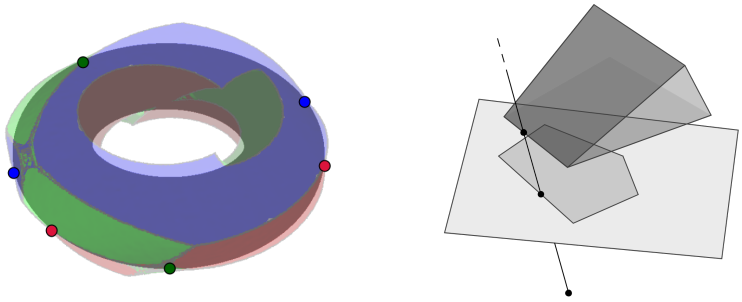


Figure 4.9: Tangential points are the intersections of a silhouette T_i with the boundary of $P_i\left(\bigcap_{j\neq i} C_j\right)$. *Left:* the silhouette of a torus and the reprojection of the visual hull generated by three different views. The six tangential points are epipolar tangencies (colors indicate projections of cone surfaces and the associated epipolar tangency points). *Right:* a tangential point may also be the projection of a "tangential triple point": in this case, the associated viewing ray is not tangent to any visual cone.

Reeb graph will be an actual graph (1-manifold with singularities).

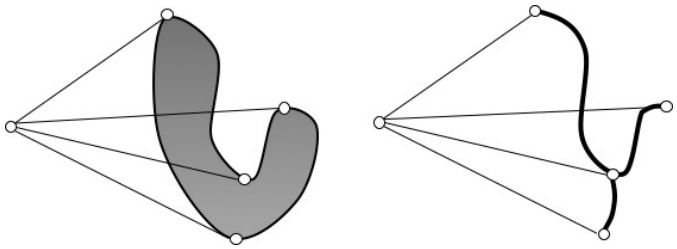
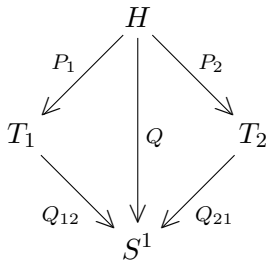


Figure 4.10: A plane region (*left*), and the associated Reeb graph (*right*), with respect to a projection (that may be viewed as a map to S^1).

Let us now consider two cameras P_1, P_2 with distinct centers c_1, c_2, two consistent closed sets T_1, T_2, and their associated visual hull H. We also indicate with $\ell = c_1 \vee c_2$ the line passing through the centers: we assume that H does not intersect ℓ (this is guaranteed by our assumption (A) from Section 4.2.1). Let us choose an arbitrary parameterization of the pencil of planes through ℓ, so that the pencil may be seen as a topological circle $\mathbb{P}^1 \cong S^1$. We thus can define a projection map $Q : \mathbb{P}^3 \setminus \ell \to \mathbb{P}^1 \cong S^1$, by associating a point p with the plane passing through p in the pencil of planes through ℓ. The parameterization of the planes through ℓ also induces parameterizations of the epipolar lines in each image: we thus can define two projection maps $Q_{ij} : \mathbb{P}^2 \setminus \{e_{ij}\} \to \mathbb{P}^1 \cong S^1$ that associate a point in image i with the corresponding epipolar line relative to the j-th epipole e_{ij}. We have the following commutative diagram:



The visual hull H is actually the *fiber product* of the maps $Q_{ij} : T_i \to S^1$ [129]: in other words, H is in bijection with

$$T_1 \times_{S^1} T_2 = \{(u, v) \in T_1 \times T_2 \,|\, Q_{12}(u) = Q_{21}(v)\} \tag{4.4}$$

(and, in fact, one can also show that H is topologically equivalent to $T_1 \times_{S^1} T_2$ with the topology induced by the product topology). As a consequence, we obtain following simple but useful result:

Proposition 4.2.9. *If G_{T_1}, G_{T_2} are the Reeb graphs of T_1, T_2 for Q_{12}, Q_{21} respectively, then the Reeb graph of H for $\mathcal{P}$ is given by the (topological) fiber product of $G_{T_1} \times_{S^1} G_{T_2}$, that is:*

$$G_H = \{(\overline{u}, \overline{v}) \in G_{T_1} \times G_{T_2} \,|\, Q_{12}(u) = Q_{21}(v)\}, \tag{4.5}$$

where $\overline{u}, \overline{v}$ denote the level sets associated with u and v respectively.

Proof. From (4.4) we see that $Q^{-1}(c) \cong Q_{12}^{-1}(c) \times Q_{21}^{-1}(c)$, which implies that the claim is true set-theoretically. To show that the result also holds topologically, we first note that the diagram illustrated above passes to the quotient on the associated Reeb graphs, so by universality of the fiber product we know that there exists a continuous map $\alpha : G_H \to G_{T_1} \times_{S_1} G_{T_2}$, given by $\alpha(\overline{p}) = (\overline{P_1(p)}, \overline{P_2(p)})$. To prove that α is an homeomorphism we only need to show that it is an open map. This is a simple verification: if $V \subseteq G_H$ is an open set, it corresponds to an open set $\hat{V} \subseteq H$, which is mapped via P_1, P_2 to two open sets $\hat{V}_1, \hat{V}_2$ in T_1, T_2 respectively (since projections are open maps), which pass to the quotient as open sets V_1, V_2 in H_{T_1}, H_{T_2}. □

We also point out that the projections induce maps on the Reeb graphs $\overline{P}_1 : G_H \to G_{T_1}$, $\overline{P}_2 : G_H \to G_{T_2}$ that are *surjective* (because of consistency). In fact, it is easy to realize that the same property must hold more generally for any set $S \subseteq \mathbb{P}^3 \setminus \{c_1, c_2\}$ such that $P_1(S) = T_1$ and $P_2(S) = T_2$:

Proposition 4.2.10. *Let $S \subseteq \mathbb{P}^3 \setminus \{c_1, c_2\}$ such that $P_1(S) = T_1$ and $P_2(S) = T_2$, and let G_S be the associated Reeb graph for the map $\mathcal{P} : S \to S^1$ (as previously defined). Then the projections P_1, P_2 induce surjective continuous maps $\overline{P}_1 : G_S \to G_{T_1}$, $\overline{P}_2 : G_S \to G_{T_2}$.*

Proof. The maps $\overline{P}_i : G_S \to G_{T_i}$, $i = 1, 2$ defined by $\overline{P}_i(\overline{p}) = \overline{P_i(p)}$ are surjective because of the assumption that $P_i(S) = T_i$. □

We can now use the Reeb graph to clarify Figure 4.4 in Section 4.2.2. For example, Figure 4.11 shows two silhouettes T_1, T_2 and their Reeb graphs G_{T_1}, G_{T_2} (we assume rectified epipolar geometry), together with the Reeb graph of the associated visual hull R_H, which is easily computed using (4.5): points in G_H represent possible pairs in G_{T_1}, G_{T_2} that are at the same "height".

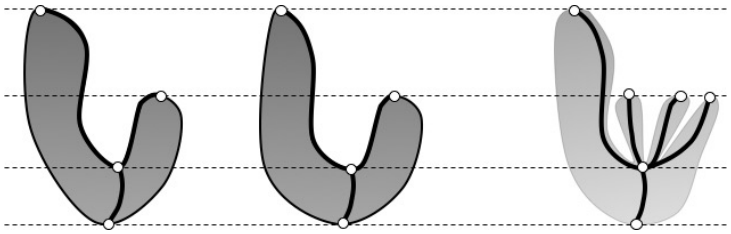


Figure 4.11: Two silhouettes, with inscribed Reeb graphs associated to the epipolar projections (*left* and *center*) and the Reeb graph of their visual hull (*right*).

Figure 4.12 shows that when the Reeb graphs G_{T_1}, G_{T_2} have non-extremal maxima of different heights (which represent unmatched internal epipolar tangencies), then H must have some level sets with at least *three* connected components (*i.e.*, the visual hull must have least three "local maxima") in order for surjective continuous projections to G_{T_1}, G_{T_2} to exist (Proposition 4.2.10). On the other hand, if the silhouettes had matching internal epipolar tangencies, two connected components would have been sufficient.

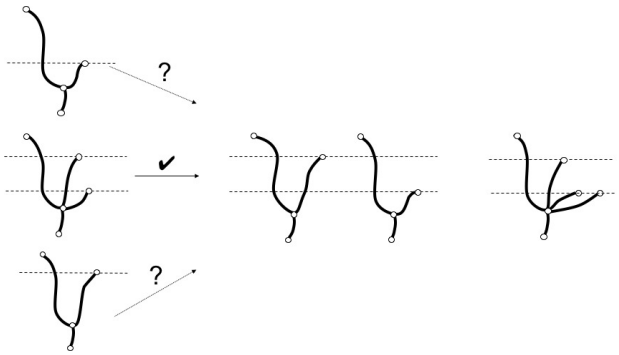


Figure 4.12: A three dimensional set S projecting onto T_1, T_2 (that have unmatched internal epipolar tangencies) must present level sets with *three* connected components (*left* and *center*). However, the visual hull of T_1, T_2 actually has level sets with *four* connected components (*right*).

Finally, Figure 4.13 shows that Reeb graph for the visual hull of two connected silhouettes may be disconnected; this implies that the visual hull itself is also disconnected. This observation is actually closely related to the so-called "mountain-climbing" problem: this problem asks for conditions on functions describing two-dimensional "mountains", so that two "climbers" in different locations can coordinate their movements to meet while always staying at the same height [194].

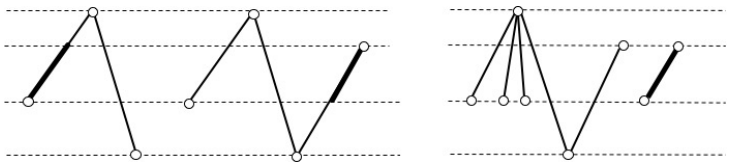


Figure 4.13: *Left*: the Reeb graphs of two connected consistent silhouettes. *Right*: the disconnected Reeb graph of their visual hull. The highlighted component of the graph associated with the visual hull corresponds to pairs of matching points on the highlighted segments on the graphs on the left.

4.3 A Dual View of Consistency

In this section we revisit the notion of consistency from the viewpoint of *duality*. In particular, we discuss the relationship between consistency for projections, as presented in the previous section, and a different notion of consistency for planar sections of solids. We focus in our presentation on the case of convex sets, since the duality is much simpler in this setting. Section 4.3.1 is an informal introduction to convex duality in the projective setting; the interested reader can find more precise definitions in Section 4.3.3; our main results are given in Section 4.3.3.

4.3.1 Duality

The basis of many similar notions of duality is the fact that points and hyperplanes in some n-dimensional space can play symmetric roles. For example, in $\mathbb{R}^n$ any hyperplane through the origin can be represented with its orthogonal vector. In projective space $\mathbb{P}^n$, all hyperplanes correspond to points of the dual projective space $(\mathbb{P}^n)^*$; in fact, any k-dimensional linear subspace H in $\mathbb{P}^n$, can be associated with an $(n-k)$-dimensional linear subspace H^* in $(\mathbb{P}^n)^*$ (see Appendix A). The identification of hyperplanes in $\mathbb{P}^n$ with points in the dual space $(\mathbb{P}^n)^*$ can be used to introduce notions of duality for more general objects in $\mathbb{P}^n$. If S is a smooth or algebraic hypersurface in $\mathbb{P}^n$, (the closure of) the set of tangent hyperplanes at points of S forms a *dual hypersurface* $S^\vee$ in $(\mathbb{P}^n)^*$ [180]. In this setting, however, the dual hypersurface will typically have self-intersections: for example, in the case of plane curves, crossings in the dual correspond to bitangents of

the original curve. For this reason, we will consider here duality for *(properly) convex sets* in projective space [65]. In the following, we will say that a set K in $\mathbb{P}^n$ is *convex* if there exists an affine chart so that K is compact and convex in the usual affine sense. This is equivalent to a more analytical definition, for which K is convex if it is the projectivization of a *cone* $\hat{K}$ in $\mathbb{R}^{n+1}$ that is convex, closed, and pointed (*i.e.*, it does not contain any line) [23]. To any convex set K in $\mathbb{P}^n$, we can now associate a *dual set* K° in $(\mathbb{P}^n)^*$, that can be characterized geometrically as the closure of the set of *all hyperplanes H such that $H \cap K$ is empty*. Alternatively, the same set K° is the projectivization of the *polar cone* $(\hat{K})^\circ$ of a convex cone $\hat{K}$ in $\mathbb{R}^{n+1}$ that corresponds to K.

For an appropriate choice of coordinates, convex duality in projective space coincides with the usual concept of polarity for affine sets [23]. However, while many properties of affine polarity only apply for convex sets containing the origin (in particular, biduality: $K^{\circ\circ} = K$), the projective framework is completely "homogeneous", and does not require similar conditions. Moreover, this setting is useful for dealing with general perspective projections. All the results in Section 4.3.3 can be understood geometrically and proven using purely synthetic arguments, by exploiting the characterization of duality as the set of "complementary" hyperplanes. More formal analytical proofs can be given using cones in $\mathbb{R}^{n+1}$.

4.3.2 Convex cones and polarity

We now clarify some technical aspects on convex duality in the projective setting. We first recall some notions on *convex cones*. More details can be found in [23, 161].

A set C in a real vector space V is a *cone* if $x \in C$ implies $\lambda x \in C$ for all $\lambda \geq 0$. A cone C is *pointed* if it contains no line, *i.e.*, if $x \in C$ and $-x \in C$ then $x = 0$. Finally, a cone C is *solid* if it has a non-empty interior.

For any cone $C \subseteq V$, we can define the *polar cone* $C^\circ \subseteq V^*$ (we denote by V^* the space of linear functionals on V):

$$C^\circ = \{\varphi \in V^* \,|\, \varphi(\mathbf{x}) \leq 0, \;\; \forall \mathbf{x} \in C\} \tag{4.6}$$

We should note that many authors only consider $V = \mathbb{R}^N$ and identify $\mathbb{R}^N$ and $(\mathbb{R}^N)^*$ by using the standard scalar product; for our purposes it is useful to keep the two spaces distinct. For any cone C, the polar cone C° is closed and convex. Moreover, the following properties hold (see [23]):

- If C is solid, then C° is pointed.
- If the closure of C is pointed, then C° is solid.
- $C^{\circ\circ}$ is the closure of the convex hull of C. In particular, if C is convex and closed, $C^{\circ\circ} = C$.

Convex sets in projective space. Let us recall here that an *affine chart* in projective space $\mathbb{P}^n$ is simply the complement of any hyperplane $A = \mathbb{P}^n \setminus H$: in practice, A may be seen as a copy of affine space in $\mathbb{P}^n$. In fact, the complement of the hyperplane $H = \{[\mathbf{x}] \,|\, \varphi(\mathbf{x}) = 0\}$ in $\mathbb{P}^n$ can be written as $A = \{[\mathbf{x}] \in \mathbb{P}^n \,|\, \varphi(\mathbf{x}) = 1\}$, which is in bijective correspondence with $\{\mathbf{x} \in \mathbb{R}^{n+1} \,|\, \varphi(\mathbf{x}) = 1\}$ in $\mathbb{R}^{n+1}$. In the following, it will be convenient to say that a set $T \subseteq \mathbb{P}^n$ is *finite* if its closure is contained in some affine chart (*i.e.*, if the closure of T does not intersect all hyperplanes in $\mathbb{P}^n$).

For any set $S \subseteq \mathbb{R}^{N+1}$, we can consider the projectivization $\mathbb{P}S = \{[\mathbf{x}] \in \mathbb{P}^n \,|\, \mathbf{x} \in S\} \subseteq \mathbb{P}^n$, consisting of all classes of vectors up to scale that contain elements of S. Conversely, if $T \subseteq \mathbb{P}^n$ is *connected* and *finite*, there are exactly two *pointed* cones C_1, C_2 such that $\mathbb{P}C_i = T$, and these are such that $C_1 = -C_2$. We will say that C_i is a cone *over* T. We now give the following definition:

Definition 4.3.1. A closed finite set $K \subseteq \mathbb{P}^n$ is *convex* if there exists a closed pointed convex cone $C \subseteq \mathbb{R}^{n+1}$ such that $K = \mathbb{P}C$. See Figure 4.14.

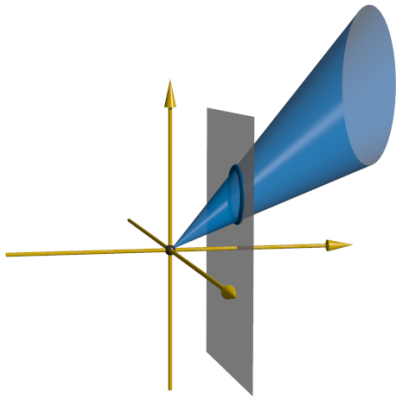


Figure 4.14: A convex set in $\mathbb{P}^n$ is the projectivization of a closed pointed convex cone in $\mathbb{R}^{n+1}$

As mentioned in Section 4.3.1, the following fact holds.

Proposition 4.3.2. *A finite set $K \subseteq \mathbb{P}^n$ is convex if and only if there exists a affine chart where K is compact and convex in the affine sense.*

Proof. If C is a pointed convex cone in $\mathbb{R}^{n+1}$, there exists $\varphi \in (\mathbb{R}^{n+1})^*$ such that $\varphi(\mathbf{x}) < 0$ for all $\mathbf{x} \in C \setminus \{0\}$. Thus, if $K = \mathbb{P}C$, K is contained in the affine chart $\{[\mathbf{x}] \in \mathbb{P}^n \,|\, \varphi(\mathbf{x}) = 1\}$, and in this chart it is compact and convex as an affine set. Conversely, if K is compact and convex in $\{[\mathbf{x}] \in \mathbb{P}^n \,|\, \varphi(\mathbf{x}) = 1\}$, then $C = \{\mathbf{x} \in \mathbb{R}^{n+1} \,|\, [\mathbf{x}] \in K, \varphi(\mathbf{x}) \geq 0\}$ is a closed pointed convex cone in $\mathbb{R}^{n+1}$ such that $\mathbb{P}C = K$. □

For any finite and connected set $T \subseteq \mathbb{P}^n$, we define the *convex hull* in $\mathbb{P}^n$ of T as $\mathbb{P}C$ where $C \subseteq \mathbb{R}^{n+1}$ is the convex hull of one of the two pointed cones over T (the projectivization

makes the choice of C irrelevant). Finally, we can define the *dual* $T^\circ \subseteq (\mathbb{P}^n)^*$ of a finite set $T \subseteq \mathbb{P}^n$ as $\mathbb{P}C^\circ$ for a pointed cone over C over S (again, the choice is irrelevant). Note that since Definition 4.3.1 requires a convex set to be the projectivization of *pointed* convex cones, it is possible that the dual of a convex set $K \subseteq \mathbb{P}^n$ may not be convex: indeed, if $C \subseteq \mathbb{R}^{n+1}$ is a convex cone with empty interior, then C° is not pointed. The next result shows that the convex dual of a set $T \subseteq \mathbb{P}^n$ coincides with the closure of the hyperplanes that do not meet T in $\mathbb{P}^n$:

Proposition 4.3.3. *Let $T \subseteq \mathbb{P}^n$ be a finite connected set and let C be a pointed cone over T. A hyperplane $[\varphi] \in (\mathbb{P}^n)^*$ belongs T° if and only if $\varphi(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in C$ or $\varphi(\mathbf{x}) \leq 0$ for all $\mathbf{x} \in C$.*

Proof. By definition $T^\circ = \mathbb{P}C^\circ$ and $C^\circ = \{\varphi \in (\mathbb{R}^n)^* \,|\, \varphi(\mathbf{x}) \leq 0, \forall \mathbf{x} \in C\}$: it is clear now that if $[\varphi] \in \mathbb{P}C^\circ$ then necessarily $\varphi \in C^\circ$ or $-\varphi \in C^\circ$. □

4.3.3 Duality and visual hulls

Given a camera $P : \mathbb{P}^3 \setminus \{c\} \to \mathbb{P}^2$ with pinhole c we write $P^\vee : (\mathbb{P}^2)^* \to (\mathbb{P}^3)^*$ for the *dual embedding* that associates lines in $\mathbb{P}^2$ with planes in $\mathbb{P}^3$ through c (see Section 1.2.5). The image of $P^\vee$ is a plane in the dual space $(\mathbb{P}^3)^*$, namely the dual $c^\vee$ of the pinhole c (the set of planes containing c). The following result describes perspective projections of convex sets from the dual point of view.

Proposition 4.3.4. *Let P be a perspective projection with center c, and let $K \subset \mathbb{P}^3 \setminus \{c\}$ and $L \subset \mathbb{P}^2$ be a convex sets. Then*

$$P(K) = L \text{ if and only if } P^\vee(L^\circ) = K^\circ \cap c^\vee. \tag{4.7}$$

Indeed, the geometric intuition for (4.7) is that a line in $\mathbb{P}^2$ does not meet L if and only if its preimage plane in $\mathbb{P}^3$ does not meet K. An analytical proof is as follows.

Proof. It is enough to show the analogous property for cones in $\mathbb{R}^4$. For any finite connected set $S \subseteq \mathbb{P}^n$, we denote by $\bar{S}$ an arbitrary pointed cone in $\mathbb{R}^{n+1}$ over S. We also indicate with $\bar{P}$ a linear projection map $\bar{P} : \mathbb{R}^4 \to \mathbb{R}^3$ associated with a perspective projection P on $\mathbb{P}^3$. We need to show that $\bar{P}(\bar{K}) = \bar{L}$ is equivalent to $\bar{P}^\vee(\bar{L}^\circ) = \bar{K}^\circ \cap \mathbf{c}^\perp$, where $\bar{P}^\vee : (\mathbb{R}^3)^* \to (\mathbb{R}^4)^*$ is the dual map associated with $\bar{P}$ (*i.e.*, the map $\varphi \mapsto \varphi \circ \bar{P}$ for $\varphi \in (\mathbb{R}^3)^*$), and $\mathbf{c}^\perp$ denotes the hyperplane in $(\mathbb{R}^4)^*$ of all linear functionals that vanish on the null-space of $\bar{P}$. Indeed, we have that:

$$\begin{aligned}
&\bar{L} = \{\bar{P}(\mathbf{x}) \,|\, \mathbf{x} \in \bar{K}\} \\
&\Leftrightarrow \bar{L}^\circ = \{\varphi \in (\mathbb{R}^3)^* \,|\, \varphi(\bar{P}(\mathbf{x})) \leq 0, \;\; \forall \mathbf{x} \in \bar{K}\} \\
&\Leftrightarrow \bar{L}^\circ = \{\psi \in (\mathbb{R}^4)^* \,|\, \psi(\mathbf{x}) \leq 0, \;\; \forall \mathbf{x} \in \bar{K}\} \cap \mathrm{Im}(\bar{P}^\vee) \\
&\Leftrightarrow \bar{L}^\circ = \bar{K}^\circ \cap \mathbf{c}^\perp,
\end{aligned}$$

where we used the fact that $\mathrm{Im}(\bar{P}^\vee) = \mathbf{c}^\perp$. □

This result shows that *projecting* a convex object in $\mathbb{P}^3$ is equivalent to *taking a planar section* of its dual. We also refer to [146], where Ponce and Hebert used a similar idea in a smooth setting to investigate the qualitative relationship between image contours and projective shapes. Motivated by Proposition 4.3.4, we introduce a notion of "sectional consistency". We assume that we are given projective maps $J_1, \ldots, J_n$, where each $J_i : \mathbb{P}^2 \to \mathbb{P}^3$ identifies $\mathbb{P}^2$ with a plane π_i in $\mathbb{P}^3$. The planes $\pi_1, \ldots, \pi_n$ are distinct.

Definition 4.3.5. A family $L_1, \ldots, L_n$ of sets in $\mathbb{P}^2$ is said to be *sectionally consistent* (relative to the embeddings $J_1, \ldots, J_n$) if there exists $K \subseteq \mathbb{P}^3$ such that $K \cap \pi_i = J_i(L_i)$ for all $i = 1, \ldots, n$.

This notion of sectional consistency is "dual" to geometric consistency from Definition 4.2.1.

Proposition 4.3.6. *A family $L_1, \ldots, L_n$ of convex sets in $\mathbb{P}^2$ is consistent for a set of projections $P_1, \ldots, P_n$ if and only if $L_1^\circ, \ldots, L_n^\circ$ are sectionally consistent for the embeddings $P_1^\vee, \ldots, P_n^\vee$. Moreover, if consistency holds, and H is the* finite visual hull *associated with $L_1, \ldots, L_n$, then $H = K^\circ$, where K is the* convex hull *of $P_1^\vee(L_1^\circ), \ldots, P_n^\vee(L_n^\circ)$.*[5]

Proof. The first claim is a direct consequence of (4.7). For the second part, we can use the fact that for arbitrary closed convex cones $K_1, \ldots, K_n$ in $\mathbb{R}^N$

$$[Conv(K_1 \cup \ldots \cup K_n)]^\circ = K_1^\circ \cap \ldots \cap K_n^\circ. \tag{4.8}$$

See for example [161]. In fact, it is sufficient to take $K_i = P_i^\vee(L_i^\circ)$, and observe that K_i° is the visual cone C_i. This follows from the following relations for cones in $\mathbb{R}^4$ (we use the same notation as in the proof of Proposition 4.3.4):

$$\begin{aligned}(\bar{P}_i^\vee(\bar{L}_i^\circ))^\circ &= \{\mathbf{x} \in \mathbb{R}^4 \,|\, \psi(\mathbf{x}) \leq 0 \ \forall \psi \in \bar{P}_i^\vee(\bar{L}_i^\circ)\} \\ &= \{\mathbf{x} \in \mathbb{R}^4 \,|\, \varphi(\bar{P}(\mathbf{x})) \leq 0 \ \forall \varphi \in \bar{L}_i^\circ\} \\ &= \{x \in \mathbb{R}^4 \,|\, \bar{P}(\mathbf{x}) \in \bar{L}_i\}.\end{aligned}$$

Note that while $\{\mathbf{x} \in \mathbb{R}^4 \,|\, \bar{P}(\mathbf{x}) \in \bar{L}_i\}$ is a convex cone in $\mathbb{R}^4$, its projectivization coincides with the two sided projective cone C_i, which is *not* convex for Definition 4.3.1 (in fact, the set $P_i^\vee(L_i^\circ)$ has empty interior). □

Consistency is arguably more intuitive in its dual formulation, since planar sets in space are conceptually easier to grasp than families of cones. For example, the pairwise consistency constraint for two convex silhouettes T_1, T_2 dualizes to the fact that $P_1^\vee(T_1^\circ)$ and $P_2^\vee(T_2^\circ)$ have the same intersection with the "dual baseline" $(c_1 \vee c_2)^\vee$: see Figure 4.15.

[5] The convex hull of a closed set S contained in an affine chart in $\mathbb{P}^n$ is well-defined only when the set S is connected: this is true for $P_1^\vee(L_1^\circ) \cup \ldots \cup P_n^\vee(L_n^\circ)$ since assumption (A) guarantees that $P_i^\vee(L_i^\circ) \cup P_j^\vee(L_j^\circ)$ is connected for all i, j.

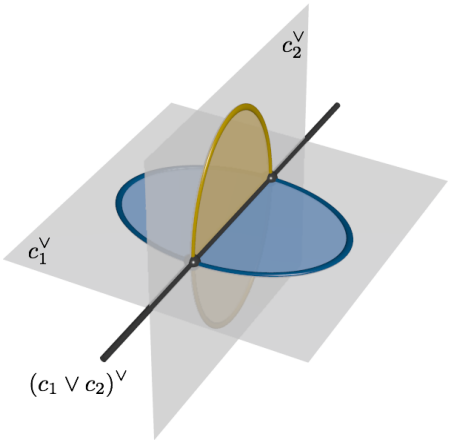


Figure 4.15: Dual pairwise-consistency. The dual images of the silhouettes $P_1^\vee(T_1^\circ)$ and $P_2^\vee(T_2^\circ)$ are sectionally consistent if and only if they have the same intersection with the "dual baseline" $(c_1 \vee c_2)^\vee$.

We also point out a duality between the "combinatorial" structure of the visual hull H and of its dual $K = H^\circ$: by interpreting the supporting planes of H as points on the boundary of K, we see that a plane π supports a cone (patch) C_i if and and only if, in the dual space, it represents a point belonging to the *planar curve* that is the boundary of $P_i^\vee(L_i^\circ)$; on the other hand, if a supporting plane π meets the visual hull at the intersection of *two* cones C_i, C_j (*i.e.*, at an *intersection curve*, see Section 4.6), then it is dual to a point on a *ruled patch*, joining the two boundaries of $P_i^\vee(L_i^\circ)$ and $P_j^\vee(L_j^\circ)$; finally, if π supports H at the intersection of three cones C_i, C_j, C_k (*i.e.*, a *triple point*), then it represents a point of the dual hull K that belongs to a *planar patch*, spanning across the three boundaries of $P_i^\vee(L_i^\circ), P_j^\vee(L_j^\circ), P_k^\vee(L_k^\circ)$ (note that ruled and planar patches are typical for convex hulls of curves in space [99]). See Figure 4.16.

The notion of sectional consistency given in Definition 4.3.5 is reminiscent of questions in geometric tomography [63], or stereology [87]. In tomography, for example, the duality between projections and sections is well studied, but typically in an affine setting that considers only orthographic projections. Nevertheless, it is quite possible that tools from these neighboring fields could provide interesting new insight for problems in computer vision.

4.4 Consistency and Line Incidences

We now re-introduce consistency as a property of families of lines in 3D. Our previous definition of consistency did not in fact depend on the representation of the image sets in terms of coordinates $\mathbb{P}^2$, but only on the incidence properties of the associated visual rays.

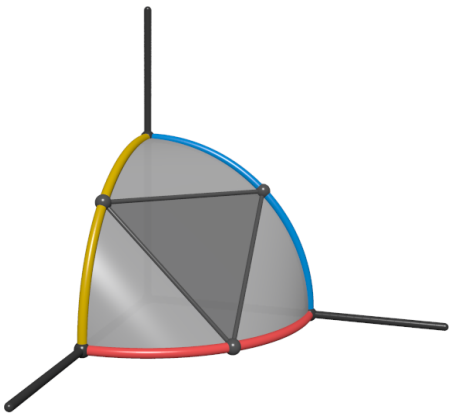


Figure 4.16: Structure of the "dual visual hull". The boundary of the dual visual hull is composed of planar curves, ruled surface patches and planar triangular patches: each of these are associated with specific components of the visual hull. See text for details.

After giving some basic definitions in Section 4.4.1, we summarize in Section 4.4.2 some results from our paper [28], where we investigated the relationship between consistency and k-consistency for *finite* families of lines.

4.4.1 Consistency for lines

If $\mathcal{L}_1, \ldots, \mathcal{L}_n$ are n sets of lines $\mathbb{P}^3$, we define an *n-incidence* as an n-tuple of lines $\ell_1, \ldots, \ell_n$ that are all concurrent and where each ℓ_i belongs to $\mathcal{L}_i$. More generally, for any subset of indices $I = \{i_1, \ldots, i_k\} \subset \{1, 2 \ldots n\}$ we say that a *k-incidence for I* is a k-tuple of lines $\ell_{i_1}, \ldots, \ell_{i_k}$ that are all concurrent and where ℓ_{i_j} belongs to $\mathcal{L}_{i_j}$.

Definition 4.4.1. The sets of lines $\mathcal{L}_1, \ldots, \mathcal{L}_n$ are *consistent* if every line in $\mathcal{L} = \bigcup_{i=1}^{n} \mathcal{L}_i$ belongs to a n-incidence.

Definition 4.4.2. The sets $\mathcal{L}_1, \ldots, \mathcal{L}_n$ are *k-consistent* if any subfamily $\mathcal{L}_{i_1}, \ldots, \mathcal{L}_{i_k}$ is consistent. In other words, $\mathcal{L}_1, \ldots, \mathcal{L}_n$ are k-consistent if for every subset $I = \{i_1, \ldots, i_k\} \subset \{1, 2 \ldots n\}$, every line in $\mathcal{L}_{i_1} \cup \ldots \cup \mathcal{L}_{i_k}$ belongs to a k-incidence for I.

It is easy to see that the image sets $T_1, \ldots, T_n$ in $\mathbb{P}^2$ are consistent (resp. k-consistent) for some cameras $P_1, \ldots, P_n$ in the sense of Definition 4.2.1 (resp. Definition 4.2.4) if and only if the families of lines $\mathcal{L}_i$ associated with the visual cones $C_i = P_1^{-1}(T_i)$ are consistent in the sense of Definition 4.4.1 (resp. Definition 4.4.2). This is completely analogous to the fact that the correspondence of points across different images can be viewed as a geometric condition on the concurrence of visual rays. Note that we did not require for each family of lines $\mathcal{L}_i$ to be concurrent at a point, even though this will always be the case for families of visual rays coming from pinhole cameras: the more

general setting can be useful for studying consistency for non-central imaging systems like the ones described in Chapters 2 and 3.

In [28] we studied the consistency for families of lines from the viewpoint of *discrete geometry*. In particular, our main result in that paper shows that it is possible to construct *finite* but *arbitrarily large* families of lines $\mathcal{L}_1, \ldots, \mathcal{L}_n$ that are k-consistent but have not even one $(k+1)$-incidence. Our discussion in [28] mainly focuses on mathematical questions, that are not always directed connected to vision. On the other hand, consistency in this finite setting is related to multi-view point correspondences, and might be useful for resolving ambiguities in SIFT matching. For this reason, we summarize without proofs some of the main results of [28] in the following subsection. The interested reader is referred to the paper for more details.

4.4.2 k-consistency with no colorful incidence

We have seen in Figure 4.6 that there exist image sets (and thus families of lines) that are pairwise consistent but not globally consistent. The following example illustrates that is also possible to construct arbitrarily large finite families of concurrent lines $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3$ that are 2-consistent but have no 3-incidences.

Example 4.4.3. Let c_0, c_1, c_2 be three non-collinear points in $\mathbb{P}^3$. For any $n \geq 1$, we construct a sequence of points $x_0, x_1 \ldots, x_{3n}$ so that x_0 is arbitrary, and x_{i+1} is a generic point on the line $\ell_{(i,i+1)} = c_{(i \bmod 3)} \vee x_i$. For each $j \in \{0, 1, 2\}$ define $\mathcal{L}_j$ to be the set of lines $\ell_{(i,i+1)}$ with $i \equiv j \pmod 3$. These families are 2-consistent but have no 3-incidences. ♢

For $k \geq 3$, showing the existence of families of lines that are k-consistent but have no $(k+1)$-incidence requires much more work. In the rest of this section, it will convenient to refer to each family $\mathcal{L}$ a *color*. An n-incidence will sometimes be referred to as a *colorful incidence*. Two proofs of the following result are given in [28].

Theorem 4.4.4. *For any $k \geq 3$, there exist arbitrarily large families $\mathcal{L}_1, \ldots, \mathcal{L}_{k+1}$ of concurrent lines in $\mathbb{P}^3$ that are k-consistent but have no colorful incidence.*

Both proofs of this result in [28] construct families of lines $\mathcal{L}'_1, \ldots, \mathcal{L}'_{k+1}$ in $\mathbb{R}^{k+1} \subset \mathbb{P}^{k+1}$ with desired properties (k-consistent with no colorful incidence) and with color classes consisting of parallel lines selected from a regular grid. A generic projection to $\mathbb{P}^3$, preserves incidences and therefore the properties of the construction. To obtain the families $\mathcal{L}'_1, \ldots, \mathcal{L}'_{k+1}$ in $\mathbb{R}^{k+1}$, we use two different strategies: the first approach is probabilistic, while the second approach uses linear algebra over finite vector spaces. In both constructions, every color class is concurrent. The probabilistic argument is asymptotic and proves the existence of configurations where every line is involved in many k-incidences for every choice of $k-1$ other colors. The algebraic construction is explicit and is minimal in the sense that removing any line breaks the k-consistency. In

the second explicit approach, each family of lines $\mathcal{L}_i$ has 2^{k^2-k-1} lines per set, which means 32 lines of each color class for $k = 3$ (see Figure 4.17).

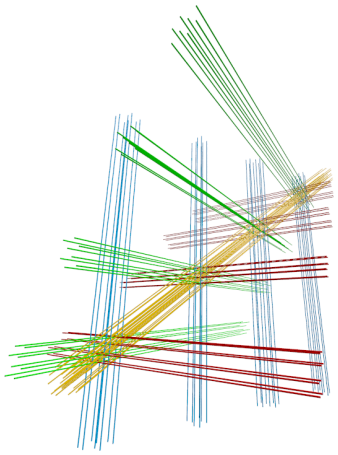


Figure 4.17: Four families of lines $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3, \mathcal{L}_4$ obtained using the explicit algebraic construction from [28] for $k = 3$ (projecting a regular 4-grid to $\mathbb{P}^3$). The four color classes are 3-consistent but have no colorful incidence. Each family $\mathcal{L}_i$ has 32 lines.

Constructions with few lines. For $k = 3$, the smallest configuration in our constructive proof of Theorem 4.4.4 has 32 lines per color. In fact, if we restrict ourselves to small configurations, k-consistency does imply the existence of some colorful incidences.

Theorem 4.4.5. *Let $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2 \cup \mathcal{L}_3 \cup \mathcal{L}_4$ be a 3-consistent colored set of lines in $\mathbb{P}^3$ with no colorful incidence and concurrent colors. If $|\mathcal{L}| < 24$, then $\mathcal{L}$ is contained in a 2-plane.*

We do not know whether the number 24 in this result is optimal. Interestingly, if we drop the assumption that all color classes are collinear then smaller examples exist: for example, we can show that there are exactly *two* configurations of 12 lines with four colors each of size 3 that are 3-consistent but have no colorful incidence.

Theorem 4.4.6. *Let $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2 \cup \mathcal{L}_3 \cup \mathcal{L}_4$ be a 3-consistent colored set of lines in $\mathbb{P}^3$ with no colorful incidence. If every color class has size 3, and $\mathcal{L}$ is not contained in a 2-plane, then it is one of the two configurations shown in Figure 4.18.*

The two configurations in Figure 4.18 are very closely related to the classical projective configurations Desargues and Reye, respectively (see [85, Chapter III]). The proof of Theorem 4.4.6 proceeds in two steps: first we show that there are only two possible combinatorial "incidence structures" so that each line meets exactly one line of the other three colors. Then we show that both these incidence structures can be realized geometrically in $\mathbb{P}^3$ (as illustrated in Figure 4.18). The second step can be addressed

algorithmically, by enforcing appropriate algebraic constraints on coordinates using a computer algebra system (see, *e.g.*, [175]).

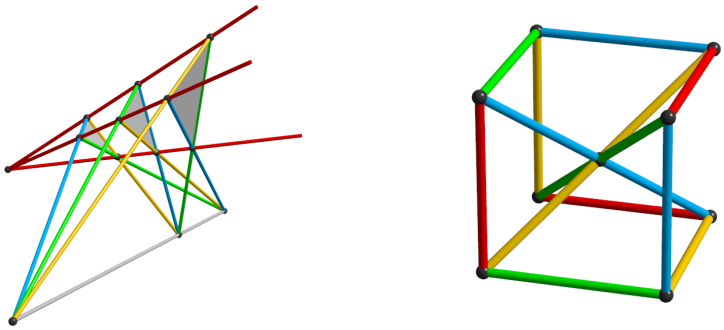


Figure 4.18: Two non-planar examples of 12 lines in 4 colors that are 3-consistent and have no 4-incidence. *Left*: A variation of Desargues' configuration. *Right*: A subset of the configuration of Reye; note that triples of parallel lines intersect at infinity.

4.5 Visual Hulls: Theory

In this section, we introduce a basic theory of visual hulls in projective space. We begin with an informal discussion, explaining the distinction between "finite" visual hulls, defined from a finite set of images, and previously considered in Sections 4.2 and 4.3 (see Definition 4.2.2), and "complete" visual hulls, associated with objects in space. Precise definitions are given in Section 4.5.2, where we also introduce a duality between sets in $\mathbb{P}^3$ and in $\mathrm{Gr}(1, \mathbb{P}^3)$. This duality is very similar to the duality between convex projective sets used in Section 4.3, and indeed visual hulls and convex hulls can be defined using almost the same construction.

4.5.1 What is a visual hull?

As mentioned in the introduction to the chapter, the term "visual hull" is often used for two different but related concepts. According to Laurentini [108], the visual hull of an object R is the largest set containing R whose opaque perspective projections are the same as those of R for *all* viewpoints.[6] More commonly, however, the term "visual hull" refers to the intersection of the visual cones obtained by observing R from a *finite* number of viewpoints. This object is the largest set whose perspective projections are

[6]Because he represents viewing rays as half-lines rather than lines, Laurentini actually defines an "internal" and an "external" (complete) visual hull, for distinguishing whether or not the viewpoint is allowed to lie inside the convex hull of the observed object. This distinction will not be relevant for us.

Figure 4.19: An object (*left*), its complete visual hull (*center*), and its finite visual hull associated with four viewpoints (*right*). Note that the finite visual hull is not topologically equivalent to the original object.

the same as those of R *only from the fixed viewpoints*. To emphasize the distinction between the two notions, we use *complete visual hull* for the former, and *finite visual hull* for the latter. The complete visual hull of an object is, in general, quite similar to the original shape, the main difference being "filled in" concavities. The finite visual hull, on the other hand, will gradually "converge" to the complete visual hull as the number of views increases, but a particular instance may be significantly different from the original object being observed (see Figure 4.19). As remarked in Section 4.2.2, the finite visual hull is also not necessarily the most "intuitive" shape associated with the given projections, especially for a small number of views.

The theory of visual hulls has been studied for polyhedral [108], smooth [21], or piecewise smooth [20] surfaces in $\mathbb{R}^3$. Here, we consider instead arbitrary *sets* in $\mathbb{P}^3$. Extending definitions from surfaces to sets does not require additional effort and, as we will argue shortly, the more general viewpoint can be useful for dealing with noisy (non-consistent) silhouettes. Regarding the projective setting, it is noted in [110] that the construction of the visual hull only relies on *contacts* between viewing rays and the observed object, and such properties can be considered independently of the euclidean or affine structure on 3D space. In [110], Lazebnik *et al.* actually make use of *oriented projective geometry* [170] in order to triangulate cone strips, since they exploit the notion of "near" and "far" intersection edges relative to the viewing direction. Oriented projective geometry is indeed necessary (and useful) to model the fact that that visible points lie "in front" of a camera. On the other hand, this setting requires some awkward constructions, since every geometric entity is doubled to take into account the possible orientations. Our approach is arguably simpler, and does not require orientation.

4.5.2 Visual and convex hulls

Let S be a set in $\mathbb{P}^3$, which for simplicity we assume to be closed, connected and with non-empty interior. For any point c in $\mathbb{P}^3$ not in S, we write $(S \,|\, c)$ for the set of lines through c that meet S:

$$(S \,|\, c) = \{\ell \in \mathrm{Gr}(1, \mathbb{P}^3) \mid c \in \ell, \ell \cap S \neq \emptyset\} \subset \mathrm{Gr}(1, \mathbb{P}^3). \tag{4.9}$$

The set $(S \,|\, c)$ can be viewed as the image of S for the *geometric camera* associated with the pinhole c (see Chapter 2). With this notation, we are interested in recovering an approximation of S from $(S \,|\, c_1), \ldots, (S \,|\, c_k)$. This is equivalent to recovering S from some perspective images in $\mathbb{P}^2$, assuming that all the camera parameters are known.

Definition 4.5.1. The *complete visual hull* associated with S is the set $VH(S)$ of points x in $\mathbb{P}^3$ such that all lines through x meet S:

$$VH(S) = \{x \in \mathbb{P}^3 \,|\, x \in \ell \in \mathrm{Gr}(1, \mathbb{P}^3) \Rightarrow \ell \cap S \neq \emptyset\} \subset \mathbb{P}^3. \tag{4.10}$$

From this definition we see that if $S' = VH(S)$ is the complete visual hull of S, then $(S' \,|\, c) = (S \,|\, c)$ for all points c in $\mathbb{P}^3$ (because the set of lines meeting S is the same as the set of lines meeting S'), and that S' is the largest set with this property. Note also that taking the visual hull can be viewed a "closure operator", *i.e.*, $VH(\,VH(S)) = VH(S)$.

Intuitively, the visual hull resembles the more conventional notion of the convex hull. We can make this idea precise by noting that the convex hull of a connected closed set S with non-empty interior in $\mathbb{P}^3$ can be defined geometrically as the set $CH(S)$ of points x in $\mathbb{P}^3$ such that all *planes* through x meet S:

$$CH(S) = \{x \in \mathbb{P}^3 \,|\, x \in H \in (\mathbb{P}^3)^* \Rightarrow H \cap S \neq \emptyset\} \subset \mathbb{P}^3. \tag{4.11}$$

When S is contained in an affine chart $\mathbb{P}^3 \setminus H$ of $\mathbb{P}^3$, this is equivalent to the convex hull in the traditional sense of convex geometry. Moreover, we can extend the notation from (4.9) so that if ℓ is a line and S is a closed connected set in $\mathbb{P}^3$, then $(S \,|\, \ell)$ is the set of planes through ℓ that meet S. The convex hull $S' = CH(S)$ of a set S now has the property that $(S' \,|\, \ell) = (S \,|\, \ell)$ for all lines ℓ in $\mathbb{P}^3$, and it is the largest set with this property.

Our definitions of visual and convex hulls suggest a more general notion of "k-convex hull" of a set S in $\mathbb{P}^n$ (with $k \leq n$), which consists of points x in $\mathbb{P}^n$ such that every $(n-k)$-dimensional linear space through x meets S. By construction, this is the largest set whose image under any linear projection $\mathbb{P}^n \dashrightarrow \mathbb{P}^k$ is the same as for S. Similar notions have indeed been introduced, although to our knowledge always in a Euclidean setting (see for example [67, 106]). In our opinion, this is less convenient for dealing geometrically with linear projections.

Returning to case of visual hulls, the problem of recovering a set S in $\mathbb{P}^3$ from a finite number of projections $(S \,|\, c_1), \ldots, (S \,|\, c_n)$ leads to the following natural definition. As before, S is a closed and connected set, and $c_1, \ldots, c_n$ are points in $\mathbb{P}^3$ not beloning to S.

Definition 4.5.2. The *finite visual hull* associated with S and $c_1, \ldots, c_n$ is the set of points x in $\mathbb{P}^3$ such that the lines $\ell_i = x \vee c_i$, for $i = 1, \ldots, n$, all meet S:

$$FVH(S \,|\, c_1, \ldots, c_n) = \{x \in \mathbb{P}^3 \,|\, (x \vee c_i) \cap S \neq \emptyset, i = 1, \ldots, n\} \subset \mathbb{P}^3. \tag{4.12}$$

By construction, the finite visual hull $FVH(S \,|\, c_1, \ldots, c_n)$ is the largest set in $\mathbb{P}^3$ such that $(S' \,|\, c_i) = (S \,|\, c_i)$ for all $i = 1, \ldots, n$ (rather than for *all* viewpoints c, as was case for the complete visual hull). The relationship with the notion of finite visual hull from Definition 4.2.2 is clear: if S is a set in $\mathbb{P}^3$ that projects onto image sets $T_1, \ldots, T_n$ in $\mathbb{P}^2$ for some cameras $P_1, \ldots, P_n$ with pinholes $c_1, \ldots, c_n$, then $T_1, \ldots, T_n$ are consistent and their finite visual hull for $P_1, \ldots, P_n$ in the sense of Definition 4.2.2 is $FVH(S \,|\, c_1, \ldots, c_n)$.

It is also interesting to note that the largest set in $\mathbb{P}^3$ such that $(S' \,|\, \ell_i) = (S \,|\, \ell_i)$ for a finite number of *lines* $i = 1, \ldots, n$ is a *polyhedron*. Hence, in a rather precise sense, a finite visual hull can be seen as a "generalized polyhedron". See also [110] and Section 4.3, where the combinatorial polyhedral structure of the visual hull is discussed.

Remark 4.5.3. If c is a point in $\mathbb{P}^3$ and ℓ is a line containing c, then $(S \,|\, c)$ determines $(S \,|\, \ell)$: indeed, a plane H belongs to $(S \,|\, \ell)$ if and only if there exists a line in $(S \,|\, c)$ contained in H. In particular, given two points c_1, c_2 the sets of lines $\mathcal{L}_1 = (S \,|\, c_1)$ and $\mathcal{L}_2 = (S \,|\, c_2)$ must induce the same set of planes $(S \,|\, (c_1 \vee c_2))$. This is essentially a geometric way of expressing the *consistency* of $\mathcal{L}_1, \mathcal{L}_2$ (see Definition 4.4.1).

Finally, visual hulls can also be understood using a form of "generalized duality". As explained in Section 4.3, we can define the *convex dual* of a finite connected set S in $\mathbb{P}^3$ as the closure of the set of all hyperplanes H such that $H \cap K$ is empty:

$$S^\circ = \mathrm{cl}\{H \in (\mathbb{P}^3)^* \,|\, \forall x \in S,\ x \notin H\} \subset (\mathbb{P}^3)^*, \tag{4.13}$$

where we write cl for the topological closure operator. Identifying $\mathbb{P}^3$ and $(\mathbb{P}^3)^*$, we can symmetrically define the dual for sets in $(\mathbb{P}^3)^*$. Now if S is closed with a non-empty interior, then $CH(S) = S^{\circ\circ}$ holds. This leads to a similar definition for the visual hull: we associate with a set S in $\mathbb{P}^3$ the (closure of the) set of lines that *do not* intersect S:

$$S^\lambda = \mathrm{cl}\{\ell \in \mathrm{Gr}(1, \mathbb{P}^3) \,|\, \forall x \in S\ x \notin \ell\} \subset \mathrm{Gr}(1, \mathbb{P}^3). \tag{4.14}$$

Conversely, we associate with a set of lines $\mathcal{L}$ in $\mathrm{Gr}(1, \mathbb{P}^3)$ the set of points that *do not* belong to any of the lines on $\mathcal{L}$:

$$\mathcal{L}^\mu = \mathrm{cl}\{x \in \mathbb{P}^3 \,|\, \forall \ell \in \mathcal{L}\ x \notin \ell\} \subset \mathbb{P}^3. \tag{4.15}$$

If S is closed with non-empty interior, then with these definitions $VH(S) = S^{\lambda\mu}$. This duality also suggests a notion of *convexity in the Grassmannian* $\mathrm{Gr}(1, \mathbb{P}^3)$, where convex sets in $\mathrm{Gr}(1, \mathbb{P}^3)$ are "double duals" $\mathcal{L}^{\mu\lambda}$ of sets of lines $\mathcal{L}$. Thus, convex sets of lines can

be viewed as counterparts inside the Grassmannian of visual hulls in space. A different notion of convexity in Grassmannian was described by Goodman and Pollack in [68, 69]: for them, however, convex sets of lines are defined as "duals" of families of convex sets in affine space. Although we will not investigate the geometric properties of these notions generalized duality and convexity, we point out that if S is a region bounded by a smooth surface Z, then the boundary of S^λ in $\mathrm{Gr}(1, \mathbb{P}^3)$ will be contained in the 3-fold of lines that are tangent to Z (the "Hurwitz 3-fold", see Chapter 5).

4.6 Visual Hulls: Practice

In this section we discuss some "practical" aspects related to consistency and visual hulls. In particular, we will show how (almost) consistent silhouettes can be used as implicit representations of a finite visual hull.

We begin by discussing an important issue that arises in practice, namely that consistency is an unstable feature and will not be exactly satisfied in the presence of noise (Section 4.6.1). We note that for dealing with this problem it is actually quite useful that we do not restrict ourselves to smooth surfaces. In Section 4.6.2, we explain how implicit image-based representations can be used to recover explicit 3D features, including a dense point cloud on the surface of the visual hull. Finally, we introduce in Section 4.6.3 a "hardness" function for points on the visual hull surface, that provides an upper bound the distance between a point and the actual (unknown) surface that is being observed. This function can be used to estimate how well the finite visual hull approximates the complete visual hull of the observed object.

Examples in this section are obtained using real data from Y.Furukawa.[7]

4.6.1 How consistent are projections of the same object in practice?

In the previous sections, we introduced finite visual hulls from two slightly different perspectives: starting from a set S in space and a finite number of viewpoints (Definition 4.5.2), and also without specifying S, simply as an intersection of visual cones (Definition 4.2.2). These definitions are essentially equivalent, since for any set of cameras $P_1, \ldots, P_n$ with centers $c_1, \ldots, c_n$, the finite visual hull $FVH(S \,|\, c_1, \ldots, c_n)$ of an "object" S coincides with the intersection of the visual cones associated with the "silhouttes" $T_i = P_i(S)$. The exact relationship between, objects, finite visual hulls, and silhouettes is illustrated in Figure 4.20. The important fact to note here is that there is a *one-to-one correspondence* between all possible finite visual hulls and all families of *consistent* silhouettes. This observation allows us to treat consistent silhouettes as representations of finite visual hulls.

[7] Available at `http://www.cse.wustl.edu/~furukawa/research/visual_hull/index.html`.

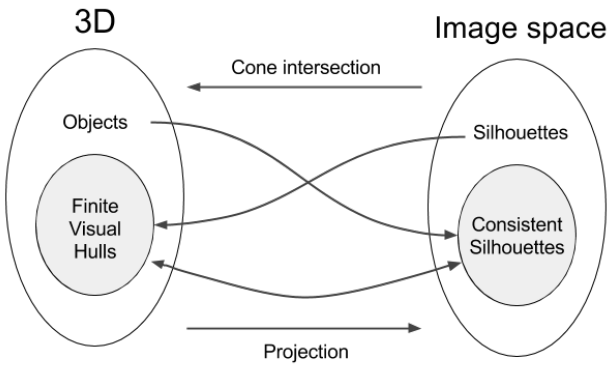


Figure 4.20: Assuming that a set of cameras has been fixed, the association between consistent silhouettes and finite visual hulls is bijective.

Silhouette consistency is however unstable for arbitrarily small measurement and calibration errors, which means that in practice silhouettes recovered separately in each image will never be exactly consistent, even if they are generated by the same object. In fact, as argued in [59, 60, 110], real contours associated with measurement errors lack the true epipolar tangencies predicted in an idealized setting. Accordingly, "epipolar tangents" to silhouettes either miss the contour or cut it transversally (Figure 4.21, *left*) and, likewise, the associated cone strips do not intersect at frontier points, but appear to be interrupted. On the other hand, as previously noted (and illustrated in Figure 4.20), naively intersecting the visual cones associated with a set of inconsistent silhouettes $T_1, \ldots, T_n$, yields in fact the finite visual hull associated with the reprojected silhouettes $\tilde{T}_i = P_i(H) \subseteq T_i$, which are, by construction, exactly consistent. These reprojected silhouettes $\tilde{T}_i$ will present sharp cuts, occurring where an epipolar tangent intersected the original silhouette (see Figure 4.21, *right*). Importantly, these cuts will not occur in idealized projections of smooth surfaces, but they are admissible in images of polyhedra (as in the "exact polyhedral visual hulls" discussed in [60]) or for more general sets considered here. In this regard, the boundaries of the reprojected consistent silhouettes $\tilde{T}_i = P_i(H) \subseteq T_i$ will present "degenerate" but *exact* epipolar tangencies, in the form of a point in one contour and a segment in the other. Such features correspond to special "frontier segments" on the visual hull surface, which are the intersections between the cone strips associated to the new silhouettes (in particular, the new cone strips are *not* disconnected). See Figure 4.21 (*right*). In conclusion, by not restricting ourselves to objects bounded by smooth surfaces, we obtain a more robust approach for dealing with finite visual hulls.[8]

[8] One could argue that the reprojection is not necessarily the "best" way to enforce consistency for noisy silhouettes. Ideally, one might try to compute a finite visual hull associated with inconsistent silhouettes $T_1, \ldots, T_n$ that is "optimal" for an appropriate measure (such as Hausorff distance between T_i and $\tilde{T}_i$). We will not deal with this issue here.

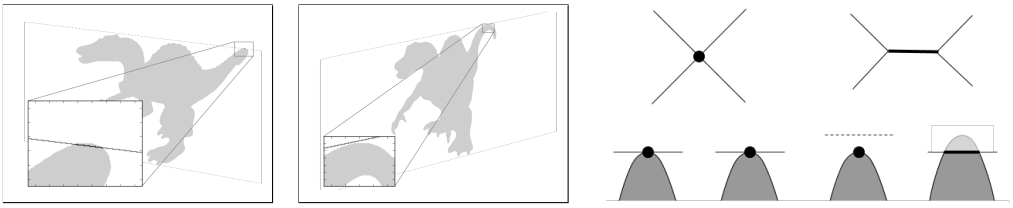


Figure 4.21: Inconsistencies in real data. *Left:* epipolar tangencies in real images do not correspond exactly. *Right:* In the presence of noise, frontier points degenerate to "frontier segments" on the visual hull.

4.6.2 From silhouettes to finite visual hulls

We now let $T_1, \ldots, T_n$ be arbitrary silhouettes in n images, and assume that n projection matrices $P_1, \ldots, P_n$ have been fixed. It will be convenient to write $H = (T_1, \ldots, T_n)$ for the finite visual hull associated with $T_1, \ldots, T_n$:

$$H = (T_1, \ldots, T_n) = \{x \,|\, P_i(x) \in T_i,\ i = 1, \ldots, n\}. \tag{4.16}$$

As noted above, this representation is also equivalent to a unique one $H = (\tilde{T}_1, \ldots, \tilde{T}_n)$, where the sets $\tilde{T}_i$ are exactly consistent (taking $\tilde{T}_i = P_i(H)$).

Equation (4.16) effectively provides a complete, albeit *implicit*, characterization of a finite visual hull and, indeed, this description is actually used by volumetric approaches [121, 34, 168] to verify whether a voxel belongs to the visual hull or not. "Surface based" methods [110, 122, 60], on the other hand, typically do not use (4.16) directly, and recover an explicit representation (usually a mesh) of the visual hull boundary by combining the geometric information provided by the *contours* in each image. However, we argue in this section that an implicit description (*i.e.*, a family of silhouettes and camera matrices) can be used to characterize directly the boundary of the visual hull, and, in fact, all of its combinatorial features. For this, it will be convenient to think of visual hulls in terms of the *joint image*, which we studied in Section 1.3. We recall here that the joint image M_n is an algebraic variety in $(\mathbb{P}^2)^n$ (or $(\mathbb{R}^2)^n$, if we consider only finite points) that contains all *point correspondences* $(u_1, \ldots, u_n)$ for the cameras $P_1, \ldots, P_n$. Here we will use the joint image mainly as a conceptual tool, treating it as an "image-based replica" of 3D space. Indeed, a point x is uniquely associated with the n-tuple of image points $(P_1(x), \ldots, P_n(x))$, and vice-versa: by definition, a correspondence $(u_1, \ldots, u_n)$ may always be triangulated to a 3D point (any two of the image points are actually sufficient for this). If the camera centers are not all collinear, this association is completely bijective.[9] In this section, we assume that the cameras $P_1, \ldots, P_n$ are fixed, and always identify a 3D point with the n-tuple of its projections,

[9] In the case of collinear centers one needs to exclude the 3D points belonging to the baseline, which are all mapped to the product of the epipoles. We will disregard this very special degenerate case.

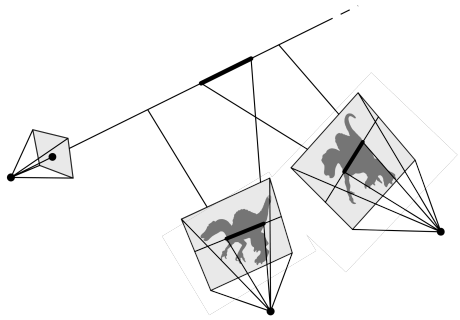


Figure 4.22: A viewing ray intersects the visual hull at "intersection intervals", which may be computed using epipolar geometry without an explicit model of the visual hull.

writing $x = (u_1, \ldots, u_n)$ where $u_i = P_i(x)$, so that $(u_1, \ldots, u_n)$ is an element of the joint image M_n. This notation is of course consistent with (4.16), since triangulating a point correspondence is effectively an intersection of visual cones.

We also point out that if we identify 3D space with the joint image, a visual hull $H = (T_1, \ldots, T_n)$ for a family of arbitrary (not necessarily consistent) image sets $T_1, \ldots, T_n$, can be characterized as

$$H = (T_1, \ldots, T_n) = M_n \cap (T_1 \times \ldots \times T_n). \tag{4.17}$$

Indeed, this relation expresses the fact that that $x = (u_1, \ldots, u_n)$ belongs to H if and only if u_i belongs to T_i for all $i = 1, \ldots, n$.

Intersection intervals using epipolar geometry. In this section, we will make use of a fundamental procedure from [123] that computes, for a given point (pixel) in a reference image, the intersections of the associated ray with the visual hull (for convenience we will refer to these sets as "intersection intervals"). The basic idea is to compute the intersection of the viewing ray and each visual cone by intersecting the epipolar line associated with the original point and the silhouette that generated the cone (thus reducing the computation from 3D to 2D). After this operation is carried out for each visual cone, the resulting intervals are lifted back to the viewing ray (or, equivalently, to an epipolar line in a reference image) and intersected. The whole procedure does not require an explicit representation of the visual hull. See Figure 4.22.

Identifying the combinatorial structure of the visual hull. The implicit description of equation (4.16) characterizes a finite visual hull as the set of points which project inside the given silhouettes. Identifying 3D space with the joint image, this is equivalent to considering all image correspondences in which each image component belongs to the

silhouette, as in equation (4.17). In particular, it is possible to use multi-view geometry to investigate the topological and combinatorial structure of the visual hull.

In the following, we write ∂S_i to indicate the boundary of S_i.

- The *cone strips* of $H = (T_1, \ldots, T_n)$ can be characterized as:

$$\begin{aligned} \partial_i H =& \{\ x \mid P_i(x) \in \partial T_i,\ P_j(x) \in T_j\ \ \forall j\} \\ =& \{\ (u_1, \ldots, u_n) \mid u_i \in \partial T_i,\ u_j \in T_j\ \ \forall j\}, \end{aligned} \tag{4.18}$$

  for $i = 1, \ldots, n$, since these are the points belonging to H that project onto (at least) one of the silhouette contours.

- The *intersection curves* are the intersections between two cone strips:

$$\begin{aligned} \partial_{ij} H =& \{\ x \mid P_i(x) \in \partial T_i,\ P_j(x) \in \partial T_j,\ P_k(x) \in T_k\ \ \forall k\} \\ =& \{\ (u_1, \ldots, u_n) \mid u_i \in \partial T_i, u_j \in \partial T_j,\ u_k \in T_k\ \ \forall k\} \end{aligned} \tag{4.19}$$

  for all $i, j = 1, \ldots, n$, and $i \neq j$ (note that such sets might be empty).

- The *triple points* are the intersections of three (or more) cone strips:

$$\begin{aligned} \partial_{ijk} H =& \{\ x \mid P_i(x) \in \partial T_i, P_j(x) \in \partial T_j, P_k(x) \in \partial T_k,\ P_l(x) \in T_l\ \forall l\} \\ =& \{\ (u_1, \ldots, u_n) \mid u_i \in \partial T_i, u_j \in \partial T_j, u_k \in \partial T_k,\ u_l \in T_l\ \forall l\} \end{aligned} \tag{4.20}$$

  for all distinct i, j, k in $\{1, \ldots, n\}$.

Frontier points can also be characterized in image space, since these are points that project onto pairs of "epipolar tangencies". However, as mentioned in Section 3.2, epipolar tangencies are unstable and will typically be degenerate in the case of real noisy data (Figure 4.21).

The image-based characterizations given above imply that the combinatorial components of the visual hull can be computed by walking along each silhouette contour and inspecting the intersection intervals (Figure 4.23, *top*). For example, points on the intersection curves are the *endpoints* of all the intersection intervals for each point on an image contour, since the extremes belong to the visual hull and project onto (at least) two of the contours. Although this strategy does not immediately provide the connectivity of the points, it is simpler (and more efficient) than the tracing procedure described in [110], where the curves generated by all pairs of cameras are first computed and then "clipped" based on whether they actually belong to the visual hull or not.[10] Figure 4.23 (*bottom*) shows the set of all intersection curves [110] computed using 24 silhouettes.

[10] Moreover, a tracing procedure of the vertices of the intersection intervals on the viewing rays can be designed, along lines similar to [110], but without the additional cost of computing the part of intersection curves lying outside the final visual hull, or a special treatment for frontier points.

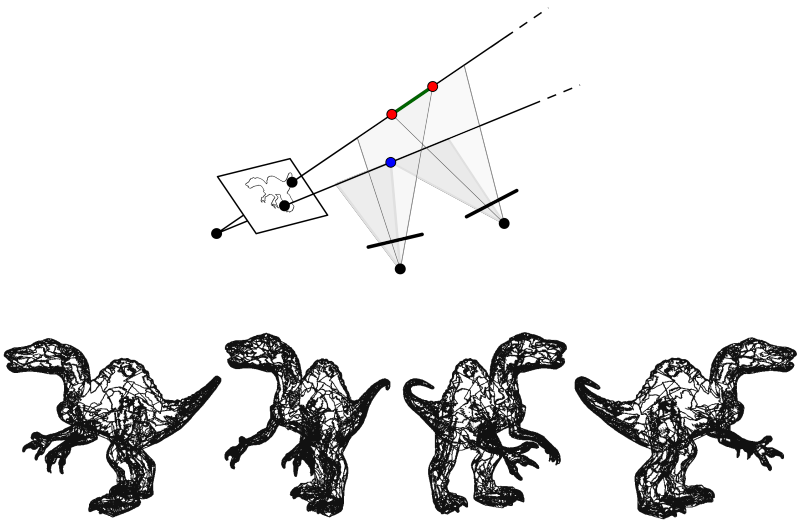


Figure 4.23: *Top:* Recovering the combinatorial components from intersection intervals for points on a contour. Using the characterizations given in the text, we see that the green segment contains general points on a cone strip (they project onto one contour), the two red points belong to an intersection curve (they project onto two contours), and the blue point is a triple point (it projects onto three contours). *Bottom:* Four views of the *intersection curves*, computed using 24 silhouettes.

From silhouettes to point clouds. We have observed that the combinatorial structure of the visual hull can be understood from intersection intervals at all points on the contours. In particular, the cone strips, which compose the visual hull's surface, are the union of *all* points on these intervals (see Figure 4.23). An explicit characterization of the surface of the visual hull can thus be obtained by *sampling* the intersection intervals for points on contours. This procedure, which requires the choice of two sampling parameters n and d, can be summarized as follows:

for all silhouettes S **do**
  Sample one out of every $n \geq 1$ points on the contour of S.
  for p in the sampled points in the contour of S **do**
    Compute the intersection intervals $I(p)$ on the ray from p (cf. Section 4.1).
    for Each interval J in $I(p)$ **do**
      Sample points at regular intervals of width d in J.
      Store the sampled points together with the extremes of J.
    end for
  end for
end for

Boyer and Franco [25] actually used a similar strategy to recover points on the surface of the visual hull, however they only considered *endpoints* of the intersection intervals, so they did not obtain an arbitrarily dense cloud (in fact, such points always project onto two contours, and thus lie on intersection curves).[11] The sampling is simple but robust, since the contours are not assumed to be smooth or polygonal, and the silhouettes used as input need not be exactly consistent. Figures 4.24–4.28, show some results. Finally, the point cloud can be sampled with any desired density, so it may be converted to a surface mesh using standard geometry processing algorithms (see Figure 4.24, *bottom* where we used a "ball-pivot" algorithm [14]).

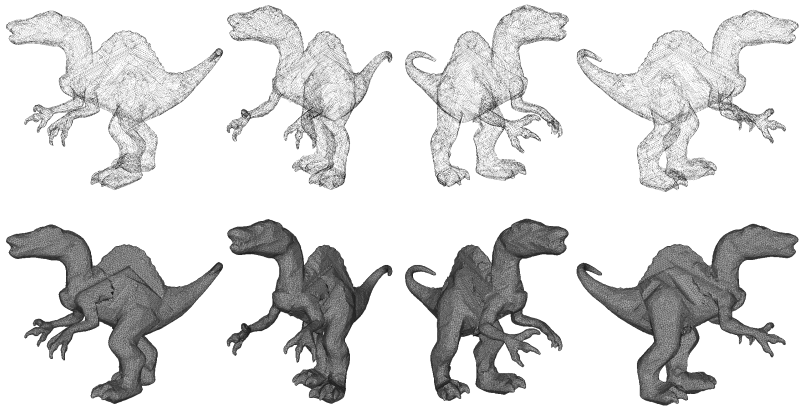


Figure 4.24: *Top:* Four views of a reconstructed point cloud (with approximately 75000 points), computed with 24 silhouettes using the procedure described in the text. *Bottom:* Four views of a mesh obtained using "ball-pivot" algorithm [14] from a dense point cloud.

4.6.3 A hardness function for points on a cone strip

We pointed out in Section 4.5 that the reconstruction of a visual hull is essentially *always* different from the actual object being observed (Figure 4.19). In particular, even in an ideal noiseless setting, when a point x belongs to the reconstructed visual hull H, it may not belong to the actual object R that is being observed. However, there exist special points in H, sometimes known as *hard points* [107], which are guaranteed to belong to R: these are points for which (at least) one of the visual rays emanating from the camera centers intersects H only at x. This characterization follows from the fact that every

[11] Presumably, Boyer and Franco considered only endpoints of the intersection intervals because they exploited only image contours, while, for robustly computing the intersection intervals, it is convenient to have access to the silhouette as a *region* (*i.e.*, as a binary mask). Our approach, on the other hand, uses contours *and* masks. This means that a basic pre-processing step might be required to obtain both representations (*e.g.*, if a continuous contour is given, a mask can be obtained using a "flood-fill" algorithm), however this is not an issue in practice. This aspect also points out how our viewpoint is somewhat in between "volumetric" and "surface-based" methods.

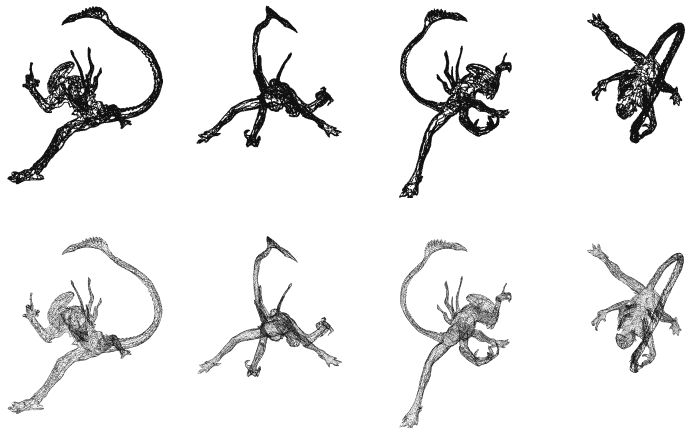


Figure 4.25: "Alien" dataset: 24 silhouettes and projection matrices. *Top:* Four views of the intersection curves. *Bottom:* Four views of a point cloud on the visual hull surface (approximately 70000 points).

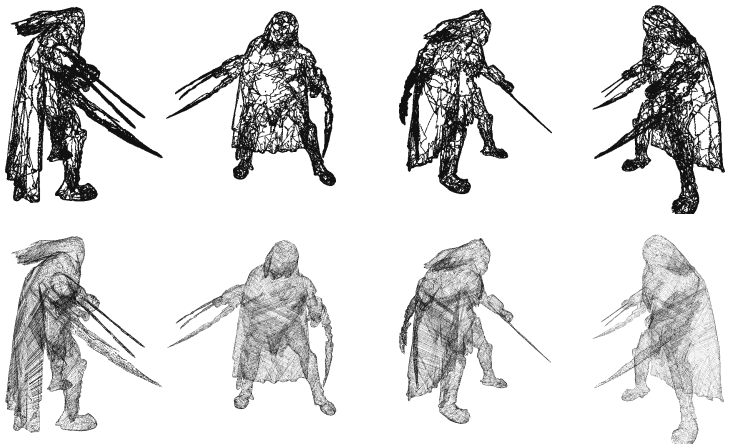


Figure 4.26: "Predator" dataset: 24 silhouettes and projection matrices. *Top:* Four views of the intersection curves. *Bottom:* Four views of a point cloud on the visual hull surface (approximately 130000 points).

Figure 4.27: "Roman" dataset: 48 silhouettes and projection matrices. *Top:* Four views of the intersection curves. *Bottom:* Four views of a point cloud on the visual hull surface (approximately 240000 points).

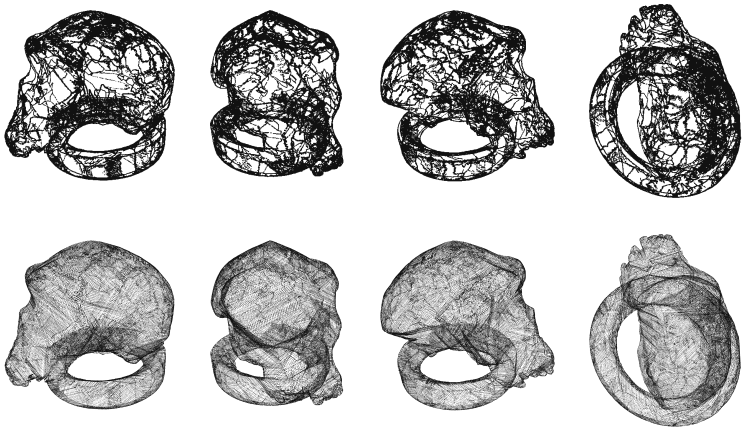


Figure 4.28: "Skull" dataset: 24 silhouettes and projection matrices. *Top:* Four views of the intersection curves. *Bottom:* Four views of a point cloud on the visual hull surface (approximately 140000 points).

visual ray intersecting H must also intersect R (and that R is necessarily contained in H). In principle, we can be certain that the surface ∂H of a reconstructed visual hull coincides with the surface ∂R of the object being observed only if all points on ∂H are hard. On the other hand, it is easy to see that for general surfaces observed from a finite number of viewpoints this is never the case. For this reason, we introduce a simple generalization of the concept of hardness in the form of a function:

Definition 4.6.1. Let $x = (u_1, \ldots, u_n)$ be a point belonging to a cone strip ∂H_i of $H = (T_1, \ldots, T_n)$ (so that $u_i \in \partial S_i$). The *hardness* of the point x (relative to $\partial_i H$), or of the image point u_i, is defined as:

$$h(x) = h(u_i) = \text{len}(CH(\ell_i \cap H)) \tag{4.21}$$

where CH denotes the convex hull, ℓ_i is the visual ray associated to u_i and len is the length measured using the Euclidean distance in $\mathbb{R}^3$.

In other words, the hardness is simply the total length of the intersection intervals along the viewing ray associated to u_i. It is easy to realize that a point x has hardness zero (relative to some strip) if and only if it is a hard point, and that more generally the hardness $h(x)$ provides an upper bound on the distance between x and a point belonging to observed object R: indeed, as mentioned above, at least one point on the intersection intervals in a viewing ray must belong to R.[12] See Figure 4.29.

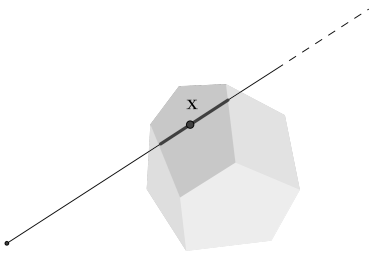


Figure 4.29: The hardness of a point on the surface of the visual hull is essentially the width of the cone strip at the point, *i.e.*, the length of the intersection interval.

In practice, the hardness of a point $u_i \in \partial S_i$ can be computed directly from image data, by recovering the intersection intervals as described in Section 4.1. Moreover, the "average hardness", for a cloud of points on the surface of the visual hull may provide a meaningful measure for the quality of the visual hull approximation. In particular,

[12] This upper bound is often not very tight, and, in fact, one could introduce variations of our definition which take into account the position of x relative to the intersection interval, or the intersection intervals associated with all of the views (rather than just the one associated to the cone strip). We chose this definition because it is much simpler and easily interpretable, and also because it provides a notion that is defined for image contour points as well.

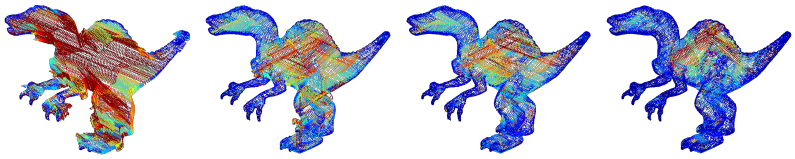


Figure 4.30: Colored point clouds, using 4,8, 12, and 24 silhouettes respectively, where the color of each point represents the hardness. In these four cases, the average hardness is 149.82, 60.26, 46.88, 31.06 respectively, compared with the dimensions of object that are approximately $250 \times 450 \times 550$. These quantitative measurements confirm the visible fact that there is a significant improvement in the quality of the visual hull reconstruction when passing from 4 to 8 silhouettes. Note also how, even in the case of 4 views, points with low hardness correctly belong to the actual object.

it can be used to estimate whether the current images provide a sufficiently accurate reconstruction. Figure 4.30 shows some examples.

4.7 Conclusions

In this chapter, we have studied geometric properties of *image consistency* and of *visual hulls*. These notions are in some sense complementary, since consistent image sets are projections in images of the same object in space, while visual hulls are 3D shapes corresponding to families of images. We have pointed out that image consistency can be seen as an extension of traditional multi-view geometry, which is devoted to consistency of points and lines. We also presented a "dual" interpretation of consistency, expressing conditions for planar sets to be sections, rather than projections, of a single object. Similarly to convex hulls, we argued that visual hulls are arguably easier to understand for arbitrary (projective) sets, rather than for smooth or polyhedral surfaces. By spelling out the topological relationship between image sets and 3D shapes, we presented a simple image-based method for obtaining a representation of the boundary of the visual hull.

Our discussion was focused almost entirely on pinhole projections, however our definitions can be easily adapted to more general imaging models. For example, if $C_1, \dots, C_n$ are *geometric cameras* (see Chapter 2), so that $C_i : \mathbb{P}^3 \dashrightarrow \mathrm{Gr}(1, \mathbb{P}^3)$, then the finite visual hull of a set S for these cameras can be defined as

$$FVH(S \,|\, C_1, \dots, C_n) = \{x \in \mathbb{P}^3 \,|\, C_i(x) \cap S \neq \emptyset, i = 1, \dots, n\} \subset \mathbb{P}^3. \tag{4.22}$$

This is the largest set whose whose images under $C_1, \dots, C_n$ are the same as for S. On the other hand, our discussion on discrete sets of lines in Section 4.4.1 points to some differences between consistency for line bundles more and general sets of lines (compare

Theorems 4.4.4 and 4.4.6). It would be interesting to see whether these differences have any impact on reconstruction methods.

We conclude with a brief discussion on two related problems which we believe could be topics for future work.

Silhouette compatibility. In our study of consistency, we considered families of image sets $T_1, \dots, T_n$ together with *known* camera projections $P_1, \dots, P_n$. On the other hand, an arguably more interesting question is to understand, given a family of sets $T_1, \dots, T_n$ in $\mathbb{P}^2$, what are the possible cameras $P_1, \dots, P_n$ that make these sets consistent. For this, we can revisit the notion of consistency from Definition 4.2.1 as a condition on both image sets as well as cameras, characterizing when $T_1, \dots, T_n$ and $P_1, \dots, P_n$ are "compatible". This idea is also important in multi-view geometry, where correspondence constraints can be viewed as conditions on image coordinates as well as on camera parameters. More generally, we can say that image sets $T_1, \dots, T_n$ are compatible if there *exist* cameras $P_1, \dots, P_n$ for which they are consistent according to our usual notion from Section 4.2.

The question of whether silhouettes can be projections of the same object for unknown cameras was discussed by Bottino and Laurentini in [19]. In that paper, the authors assume the extrinsic parameters of the cameras to be unknown, and analyze in detail the geometric constraints for compatibility in a particular case (orthographic viewing directions parallel to the same plane). Bottino and Laurentini also make an interesting observation, namely that if one assumes projective cameras (unknown intrinsic and extrinsic parameters), then *any* family of silhouettes can be projections of a single object: they observe in fact that considering a convex object, and applying local protrusions with appropriate shapes, one is able to produce arbitrary silhouettes by placing cameras near the surface. On the other hand, this property would seem to imply that SfM methods can never exploit the geometry of the silhouettes in order to recover projective camera parameters (contrary for example to the results in [81]): indeed, if $T_1, \dots, T_n$ are image sets that uniquely determine a configuration of cameras, then it is clear that another set T_{n+1} in a new image must be constrained in order for it to be consistent with $T_1, \dots, T_n$. We note however that the construction proposed in [19] and sketched above violates our assumption (A) for *all* pairs of cameras and silhouettes (viewing cones must be extremely "wide", and thus contain all other centers which lie near the surface of the convex object). We believe instead that if one excludes certain regions in the space of parameters, then silhouettes will not always be compatible.

As an example, consider two pictures with two arbitrary silhouettes T_1, T_2, one in each image. In this case, the camera configurations such that T_1, T_2 are consistent *and* such that each epipole lies outside the corresponding silhouette (as in condition (A)), have 5 degrees of freedom, compared to 7 in the unconstrained case. Indeed, if we fix the epipoles e_1, e_2 outside of T_1, T_2, then the image sets impose two constraints on camera

geometry, namely correspondence of two the epipolar tangents. Fixing instead e_1, e_2 in the interiors of T_1 and T_2, the homography of epipolar lines is unconstrained. Compatibility constraints for silhouettes in two views have actually been used recover camera parameters (directly) when the ambiguity of camera motion has only two degrees of freedom, namely for pure translation motion [160], or when a visible plane provides a homography between the views [42]. For general camera motions, however, these conditions are difficult to use because they cannot be written in closed form for arbitrary image sets T_1, T_2. Explicit constraints can in principle be formulated when the image sets admit an algebraic description: see for example [96], which essentially addresses compatibility for regions bounded by conics in $\mathbb{P}^2$.

In general, a better understanding of silhouette compatibility can be important for developing some theoretical foundations for silhouette-based SfM. Moreover, it would be interesting to understand whether a family of silhouettes alone (without camera parameters) can be used provide a unique representation of an object.

Visual hull approximations. As pointed out several times throughout the chapter, the finite visual hull is the largest set that projects onto a given family of silhouettes, but is usually different from the actual object that generated the silhouettes. In particular, especially for a small number of views, the difference between the (unknown) shape that is being observed and the visual hull can be significant (see Figure 4.4). Understanding the relationship between an object and its reconstructed visual hull, as a function of the location of the viewpoints, is an interesting and, to the best of our knowledge, largely unexplored problem. This kind of study could be important for applications, for example in viewpoint planning strategies [179].

The simple hardness function we proposed in Section 4.6.3 can be seen as a step in this direction, but it does not provide any kind of guarantee regarding the topological correctness of the reconstruction. In this sense, the kind of analysis described for two views in Section 4.2.4 might be useful, since it relates the topological features of the silhouettes with those of the visual hull, and also of all 3D shapes with the correct projections. An important tool in the study of these issues could be the "line-dual" representation S^λ in $\mathrm{Gr}(1, \mathbb{P}^3)$ of an object S (see Section 4.5). This set represents the family of all lines that meet S, and captures visual features of S in a very direct way: for example, images of S are (appropriate) planar sections of S^λ and, as we will see in Chapter 5, special singularities on the boundary S^λ correspond to "visual events". On the other hand, to our knowledge there is not much literature on data structures for representing line-geometric data (see [150]). Developing effective representations S^λ for real objects or 3D models will first require work on these topics.

Chapter 5

Changing Views on Curves and Surfaces

In this chapter we study *visual events* from an algebraic viewpoint. Given a sufficiently general curve or surface in 3-space, we consider the contour curve associated with a moving viewpoint. Qualitative changes in the contour occur when the viewpoint crosses the *visual event surface*. Our algebraic framework allows us to give a new characterization of the components of this ruled surface as the iterated singular loci of the higher *associated hypersurfaces* of the original curve or surface: these associated hypersurfaces include the *dual surface* in $(\mathbb{P}^3)^*$, and the *Chow* and *Hurwitz* hypersurfaces in $\mathrm{Gr}(1, \mathbb{P}^3)$. We also present formulas, first derived by Salmon and Petitjean, for the degrees of these surfaces, and show how to compute exact representations for all visual events using algebraic methods.

The material of this chapter is based on the following publication[1]:

– Kathlén Kohn, Bernd Sturmfels, and Matthew Trager. "Changing Views on Curves and Surfaces". In: *Acta Mathematica Vietnamica* 43.1 (2018), pp. 1–29.

5.1 Introduction

In Chapter 4 we studied projections of arbitrary sets in projective space. We now take a different look at projections of *smooth algebraic curves* and *surfaces*. In this setting, we always represent the image of an object using a plane curve. If the observed object is a curve, then the image is simply its point-wise projection. For a surface, the image is described by the *outline curve*, or *image contour*, generated by the projection. The

[1] Another PhD student, Kathlén Kohn, worked on this paper as well. This chapter does not include a technical section on intersection theory which was mainly her contribution.

outline is the natural sketch one might use to depict the surface (with transparencies), and it is the projection of the critical points where viewing lines are tangent to the surface. In both cases, the image curve will present *singularities* that arise from the projection, even if the original curve or surface was smooth. If we now let the camera travel along a path in 3-space, this path naturally breaks up into segments according to how the image looks like: within each segment, the topology and singularities of the image curve do not change. The sharp transitions in the appearance of the image curve are known as *visual events*, and play an important role in the study of visual perception. For example, any object has only a finite number of "stable appearances", or *aspects*, which correspond to regions of viewpoints in the complement of the events. The overall structure of aspects and events can also be encoded in an *aspect graph* (or *view graph*) [22], in which vertices correspond to aspects, and edges correspond to visual events between stable views. Hence, the aspect graph describes all possible evolutions of appearances of a given surface.

Traditionally, visual events have been defined mathematically using *catastrophe theory* [182, 26], a branch of *singularity theory* [193, 27]. Roughly speaking, singularity theory tries to classify *stable germs* (*i.e.*, local equivalence classes) of differential maps, exhibiting whenever possible "normal forms" that can be used to represent them. Catastrophe theory, in particular, studies stable and non-stable critical points of "potential functions". This theory can be used to describe sudden qualitative changes of behaviors of very general dynamical systems, with applications ranging from biology to behavioral sciences [198]. On the other hand, the classification of germs is based on analytical tools, which do not seem to capture directly the special geometry involved in visual problems. Moreover, the theory is most well-suited for dealing with "local" instabilities, while "multi-local" events are also important in vision.

In this chapter, we revisit visual events from a more geometric (and projective) viewpoint. In the language of algebraic geometry, any curve or surface in projective space has various "associated hypersurfaces". For example, if X is a smooth surface, its *dual surface* $X^\vee$ in $(\mathbb{P}^3)^*$ contains all tangent planes to X; its associated *Hurwitz hypersurface* in $\mathrm{Gr}(1, \mathbb{P}^3)$ contains all tangent lines to X. It is intuitively clear that these objects determine the evolution of the contours generated by X. We make this connection precise, explaining how all visual events can be obtained from the iterated singular loci of associated hypersurfaces. We also discuss effective ways of computing event surfaces using algebraic methods.

5.1.1 Previous work

The appearance of a solid object under a continuously varying viewpoint was studied in the 1970's by Koenderink and van Doorn [101, 100], in the context of visual perception in psychology and artificial intelligence. A detailed discussion can be found in

Koenderink's book *Solid Shape* [99]. On the mathematical side, the topic was studied in singularity theory by Arnol'd and others [5, 98, 142]. In that setting, the transitions between locally stable views are the non-generic singularities from *catastrophe theory*. These catastrophes have been classified for "projection-generic" smooth surfaces, and the catalogue consists of the following six *visual events*. The first three names are due to René Thom [182]:

(L) Local events: *lip*, *beak-to-beak*, and *swallowtail*.

(M) Multi-local events: *tangent crossing*, *cusp crossing*, and *triple point*.

In the 80's and 90's, visual events became a popular research topic in computer vision [22, 139, 147]. We refer to Chapter 13 in the textbook by Forsyth and Ponce [57] for a general introduction. The interest in these topics was due to the fact that aspect graphs and visual events can in principle be used as compact representations for *all* qualitative appearances of objects. These ideas actually never found much practical use, mainly because of the computational complexity of aspect graphs, and the difficulties of dealing with real-life objects. Nevertheless, several algorithms for computing aspect graphs of algebraic surfaces were proposed, with test implementations involving both numerical and symbolic methods. For example, Ponce and Kriegman [147] and Rieger [155] studied the case of orthographic projections of parametric algebraic surfaces; methods for implicit algebraic surfaces were introduced by Petitjean *et al.* [139] for orthographic projections, and by Rieger [154] for perspective projections. All the examples shown in these articles are very special low-degree surfaces: we revisit this literature, now 25 years old, and develop it further from the viewpoint of today's applied algebraic geometry.

5.1.2 Main contributions

Our main contributions can be summarized as follows.

- We revisit visual events for curves and surfaces in a projective and algebraic setting.[2] The different visual events are *ruled surfaces* associated with sets of lines and planes that meet the original curve or surface in exceptional ways. In particular, we can characterize the events in terms of "associated hypersurfaces" for the observed object (the dual surface in $(\mathbb{P}^3)^*$, or the Chow or Hurwitz three-folds in $\mathrm{Gr}(1, \mathbb{P}^3)$). This viewpoint is more geometric, and arguably simpler, compared to the traditional approach for defining visual events based on singularity theory.

- We explain how to manipulate algebraically ruled and developable surfaces, that can be effectively represented as curves in $\mathrm{Gr}(1, \mathbb{P}^3)$ or $(\mathbb{P}^3)^*$. We then discuss

[2] In the computer vision literature, visual events are typically studied only for surfaces (mainly because surfaces are more suited for representing real-life objects). From a theoretical standpoint, however, it is quite natural to consider first the simpler case of projections of curves.

practical methods for computing the visual events associated with a curve or surface X in $\mathbb{P}^3$. Particularly when X is a surface, this is a non-trivial matter because the degrees of the ruled surfaces in the output are very high, as seen in Table 5.2. This challenge can be addressed by pre-computing equations for *multiple-root loci*. For example, when X is a quintic surface, this allows us to compute a representation for its flecnondal surface that has degree 260. We present case studies and examples showing the computation of different visual events.

Conventions. In this chapter, we write $\mathbb{P}^n = \mathbb{P}(\mathbb{C}^{n+1})$, and our model for the object to be viewed is a variety X in $\mathbb{P}^3 = \mathbb{P}(\mathbb{C}^4)$. We will however usually assume that our varieties are defined by equations with real coefficients, and we will often be interested in the associated locus $X_{\mathbb{R}}$ of real points of X. We will not distinguish between projective points and their vector representatives, and both will be written in regular font.

5.2 Preliminaries

We present some background and tools that will be used in our discussion on visual events. In particular, Section 5.2.1 is devoted to *ruled surfaces* in $\mathbb{P}^3$ and its subclass of *developable surfaces*. We introduce effective representations of ruled surfaces, and show how to compute with these. As we will see, this is relevant because all visual event surfaces are ruled. Section 5.2.2 then focuses on ruled surfaces arising from "associated hypersurfaces", *i.e.*, as iterated singular loci of Chow and Hurwitz threefolds in $\mathrm{Gr}(1, \mathbb{P}^3)$, and of dual varieties in $(\mathbb{P}^3)^*$.

5.2.1 Ruled surfaces and developable surfaces

An irreducible surface in $\mathbb{P}^3$ is *ruled* if it is a union of straight lines (see Figure 5.1, *left*). These lines are parameterized by some curve C, and they are known as the *generators* of the surface. A first example are smooth quadratic surfaces in $\mathbb{P}^3$: over $\mathbb{C}$, these always possess two rulings of lines. Many classical results on ruled surfaces can be found in the book by Edge [50]. In this section we develop tools for computing and representing ruled surfaces algebraically. This will be useful because all of the visual event surfaces associated with an object are ruled.

A key player in our study is the Grassmannian of lines in space $\mathrm{Gr}(1, \mathbb{P}^3)$. We will write $p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}$ for the Plücker coordinates representing lines in $\mathbb{P}^3$. We refer to Section 2.2 or to Appendix A for an introduction to the Grassmannian $\mathrm{Gr}(1, \mathbb{P}^3)$ and to Plücker coordinates. Throughout the chapter we will also use the fact that any line

satisfying $p_{34} \neq 0$ has the parametric representation

$$z(t) \;=\; \begin{bmatrix} p_{12} \\ tp_{12} \\ tp_{13} - p_{23} \\ tp_{14} - p_{24} \end{bmatrix} \in \mathbb{P}^3. \tag{5.1}$$

Consider now an irreducible curve C in $\mathrm{Gr}(1, \mathbb{P}^3)$ that has degree d in $\mathbb{P}^5$. We write I_C for its prime ideal in the $\mathbb{R}[p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}]$. The union of all lines on C forms a ruled surface $\mathcal{S}_C$ in $\mathbb{P}^3$.

Lemma 5.2.1. *The ruled surface $\mathcal{S}_C$ is irreducible and it has degree d in $\mathbb{P}^3$. Conversely, every irreducible ruled surface in $\mathbb{P}^3$ arises in this way from some irreducible curve C in $\mathrm{Gr}(1, \mathbb{P}^3)$.*

Proof. This is one of the basic facts derived in Edge's book [50, Chapter I, §26]. □

In practice, the defining polynomial of the surface $\mathcal{S}_C \subset \mathbb{P}^3$ can be computed from C using the incidence conditions

$$\begin{bmatrix} 0 & p_{34} & -p_{24} & p_{23} \\ -p_{34} & 0 & p_{14} & -p_{13} \\ p_{24} & -p_{14} & 0 & p_{12} \\ -p_{23} & p_{13} & -p_{12} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \;=\; \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{5.2}$$

We add these four bilinear forms to the ideal I_C, and we then saturate with respect to the irrelevant ideal $\langle p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34} \rangle$ of $\mathbb{P}^5$.[3] The resulting ideal describes the incidence correspondence of points on lines that are in the curve C. Now, by eliminating the variables p_{ij}, we obtain an ideal in $\mathbb{R}[x_1, x_2, x_3, x_4]$ defined by a single polynomial of degree d, which defines the desired surface. This computation can also be reversed: given a surface $\mathcal{S}$ in $\mathbb{P}^3$, we can compute the *Fano scheme* of all lines on $\mathcal{S}$, that lives in $\mathrm{Gr}(1, \mathbb{P}^3)$. Its ideal in $\mathbb{R}[p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}]$ is obtained by substituting (5.1) into the equation of $\mathcal{S}$, extracting the coefficients of the resulting polynomial in t, and saturating their ideal by $\langle p_{34} \rangle$. The Fano scheme is usually empty or consists of points. However, if it is a curve C, then the surface is ruled and $\mathcal{S} = \mathcal{S}_C$. In summary, we can describe developable surfaces in $\mathbb{P}^3$ using curves in $\mathrm{Gr}(1, \mathbb{P}^3)$ (or $\mathbb{P}^5$), and it is possible to convert each representation into the other algebraically.

We now turn to *developable surfaces* (see Figure 5.1, *right*). Given an arbitrary subvariety X in $\mathbb{P}^3$, its *dual* $X^\vee \subset (\mathbb{P}^3)^*$ is the Zariski closure of the set of all planes that are tangent at a smooth point of X (see Appendix B). Typically, the dual $\mathcal{S}^\vee$ of a surface of degree at

[3] Saturation is an algebraic operation that removes components from an algebraic set. In our presentation, it is often used to remove components associated with the "irrelevant" ideal, that is generated by the coordinate variables. See Appendix B

least two will be a surface in $(\mathbb{P}^3)^*$. However, it can happen that $\mathcal{S}^\vee$ is a curve, in which case we say that $\mathcal{S}$ is a *developable surface*. In other words, $\mathcal{S}$ is developable if it has a one-dimensional family of tangent planes: this coincides with the informal definition that it can be "rolled" on a plane. It is clear that each developable surface $\mathcal{S}$ in $\mathbb{P}^3$ is encoded by its dual curve $\mathcal{S}^\vee$ in $(\mathbb{P}^3)^*$ since we can recover the surface by the biduality relation $\mathcal{S} = (\mathcal{S}^\vee)^\vee$. The following result shows that developable surfaces in $\mathbb{P}^3$ are a special subfamily of ruled surfaces.

Theorem 5.2.2. *Every developable surface $\mathcal{S}$ in $\mathbb{P}^3$ is a ruled surface,* i.e., *it satisfies $\mathcal{S} = \mathcal{S}_C$ for some curve C in $\mathrm{Gr}(1, \mathbb{P}^3)$. For a curve C in $\mathrm{Gr}(1, \mathbb{P}^3)$, the corresponding ruled surface $\mathcal{S}_C$ is developable if and only if all tangent lines of C viewed as a curve in $\mathbb{P}^5$ are contained in the Plücker quadric.*

Proof. For the first statement see [50, Chapter V, §344]. The second is [13, Prop. 12.4.1]. □

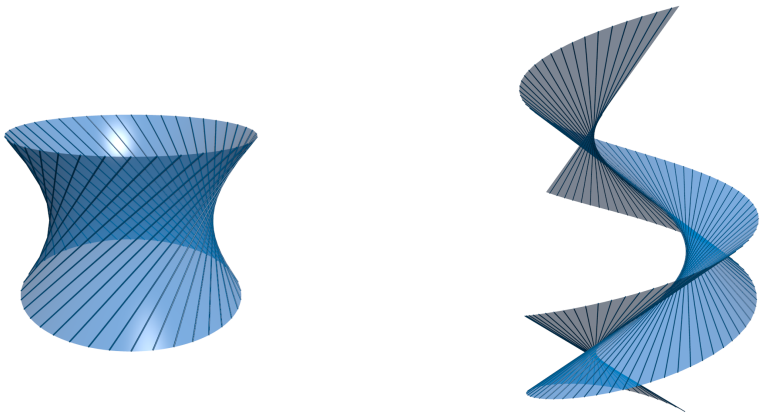


Figure 5.1: A ruled surface (*left*) and a developable surface (*right*). Note that the developable surface is the union of lines tangent to a space curve. This curve is the *edge of regression*.

A developable surface $\mathcal{S} = \mathcal{S}_C$ that is not a cone has three distinct encodings as a curve. First, there is the curve C in the Grassmannian $\mathrm{Gr}(1, \mathbb{P}^3)$. Second, there is the dual curve $\mathcal{S}^\vee$ in $(\mathbb{P}^3)^*$. We saw how to recover $\mathcal{S}$ from these encodings. Finally, there is the *edge of regression* $E(\mathcal{S})$ which lies on the surface $\mathcal{S}$ in $\mathbb{P}^3$. Points in $E(\mathcal{S})$ correspond to planes in $(\mathbb{P}^3)^*$ that intersect the curve $\mathcal{S}^\vee$ with multiplicity three ("osculating planes"). The surface $\mathcal{S}$ is the tangential surface of $E(\mathcal{S})$, *i.e.*, it is the union of lines that are tangent to $E(\mathcal{S})$ (see [140, page 111]). This also verifies that $\mathcal{S}$ is indeed a ruled surface. The curves $E(\mathcal{S})$ and $\mathcal{S}^\vee$ are also related by a biduality relation; namely, $\mathcal{S}^\vee$ is the set of planes in $\mathbb{P}^3$ that intersect $E(\mathcal{S})$ with multiplicity three [141, Thm. 5.1]. Moreover, the tangent lines of $E(\mathcal{S})$ and $\mathcal{S}^\vee$ are dual to each other. This situation degenerates when the surface $\mathcal{S}$ is a cone, which means that its dual $\mathcal{S}^\vee$ is a plane curve: in that special

case, the edge of regression $E(\mathcal{S})$ is the vertex of the cone $\mathcal{S}$. We illustrate the three curve encodings of a developable surface with a simple example.

Example 5.2.3. Let $\mathcal{S}$ be the surface of degree six in $\mathbb{P}^3$ that is defined by the polynomial

$$f \;=\; 16x_2^3x_3^3 - 27x_1^2x_3^4 + 6x_1x_2^2x_3^2x_4 - 27x_2^4x_4^2 + 48x_1^2x_2x_3x_4^2 - 16x_1^3x_4^3.$$

This is the surface in [162, §3, eq. (9)]. We verify that $\mathcal{S}$ is developable by computing the ideal of its dual variety $\mathcal{S}^\vee \subset (\mathbb{P}^3)^*$. This shows that $\mathcal{S}^\vee$ is a smooth rational quartic curve:

$$\langle\, y_2y_3 - 4y_1y_4\,,\; y_3^3 + 4y_2y_4^2\,,\; y_1y_3^2 + y_2^2y_4\,,\; y_2^3 + 4y_1^2y_3 \,\rangle. \tag{5.3}$$

The curve C in the Grassmannian $\mathrm{Gr}(1,\mathbb{P}^3)$ that encodes the ruling of $\mathcal{S} = \mathcal{S}_C$ has the ideal

$$I_C \;=\; \langle\, 2p_{14} - p_{23}\,,\; p_{24}^2 + 3p_{13}p_{34}\,,\; p_{13}p_{24} - 9p_{12}p_{34}\,,\; p_{23}^2 - 16p_{12}p_{34}\,,\; p_{13}^2 + 3p_{12}p_{24} \,\rangle.$$

This ideal defines the Fano scheme of $\mathcal{S}$ in $\mathbb{P}^5$. Finally, the edge of regression $E(\mathcal{S})$ is the rational quartic curve $\{(s^4 : s^3t : st^3 : t^4)\}$ in $\mathbb{P}^3$. The ideal of this curve equals

$$\langle\, x_2x_3 - x_1x_4\,,\; x_3^3 - x_2x_4^2\,,\; x_1x_3^2 - x_2^2x_4\,,\; x_2^3 - x_1^2x_3 \,\rangle. \tag{5.4}$$

This curve has $\mathcal{S}$ as its tangential surface. All of these computations can also be reversed: this shows how various objects can serve as a representation of the surface $\mathcal{S}$. ♢

Many of the ruled surfaces $\mathcal{S}_C$ we shall encounter in later sections have the property that their defining polynomial f is extremely large and impossible to compute symbolically. In such cases, the curve C in $\mathrm{Gr}(1,\mathbb{P}^3) \subset \mathbb{P}^5$ is more manageable, and we can often compute generators for its ideal I_C. This encoding of the curve also enables us to carry out directly computations that involve the surface $\mathcal{S}_C$. For example, we may wish to compute the points of intersection of the ruled surface $\mathcal{S}_C$ with a general line L: indeed, if a camera travels along L, the real intersection points with the visual event surfaces are precisely the visual events that can be observed from that camera motion.

To perform this computation, we fix two points $[a_1,a_2,a_3,a_4]$ and $[b_1,b_2,b_3,b_4]$ on L, and parameterize L by

$$x_i \;=\; sa_i + tb_i \quad \text{for } \; i = 1,2,3,4. \tag{5.5}$$

To compute $\mathcal{S}_C \cap L$ from I_C, we substitute (5.5) into (5.2), we add the resulting four bilinear forms to I_C, we saturate with respect to $\langle p_{12},\ldots,p_{34}\rangle$, and we then eliminate the six Plücker coordinates. The result is the principal ideal in $\mathbb{R}[s,t]$ that is generated by the binary form

$$f\big(\,sa_1 + tb_1,\; sa_2 + tb_2,\; sa_3 + tb_3,\; sa_4 + tb_4\,\big). \tag{5.6}$$

Thus, even when f is unknown, we can compute its specialization (5.6) directly from I_C.

When $\mathcal{S}$ is developable, the specialization (5.6) can also be obtained from the ideal $I(\mathcal{S}^\vee)$. Let J be a Jacobian matrix for the ideal $I(\mathcal{S}^\vee)$ in $\mathbb{R}[y_1, y_2, y_3, y_4]$. This matrix has four columns. Let J_x be the matrix obtained from J by adding one more row, namely the vector (x_1, x_2, x_3, x_4) in (5.5). We now add the 3×3-minors of J_x to the ideal $I(\mathcal{S}^\vee)$, we saturate with respect to the ideal of 2×2-minors of J, and then we eliminate the unknowns y_1, y_2, y_3, y_4. The result is the desired principal ideal generated by (5.6) in $\mathbb{R}[s, t]$. See Example 5.4.4 for an application of this method.

These strategies can be adapted to compute the curve that is obtained as the intersection of a ruled or developable surface $\mathcal{S}$ with fixed *plane* H in $\mathbb{P}^3$. For event surfaces, this corresponds to restricting the camera movement to a plane, or to assuming that all projections are orthographic (which implies that the viewpoint lies on the plane at infinity). It is sufficient to parameterize the points on H by substituting $x_i = sa_i + tb_i + uc_i$ in (5.6).

5.2.2 Associated ruled surfaces

The ruled surfaces of interest to us arise from an arbitrary curve or surface X in $\mathbb{P}^3$, and represent families of planes and lines that intersect X with prescribed multiplicities. These correspond to the bottom rows of the diagrams in Figures 5.2 and 5.3, that can be interpreted as follows. For a general curve or surface X in $\mathbb{P}^3$, the rows of these diagrams correspond to *codimension* in $(\mathbb{P}^3)^*$ or $\mathrm{Gr}(1, \mathbb{P}^3)$ (starting from zero on the top). The shown subvarieties consist of lines and planes that intersect X with various multiplicities m. A solid edge from Y_1 to Y_2 means that Y_2 is an irreducible component of the singular locus of Y_1; a dashed edge means that Y_2 is simply contained in Y_1. Below the ambient spaces $(\mathbb{P}^3)^*$ and $\mathrm{Gr}(1, \mathbb{P}^3)$ we see the so-called *coisotropic hypersurfaces* studied by Gel'fand, Kapranov and Zelevinsky in [64, Sec. 4.3.B] (see also [102, 176]). These codimension one loci are:

- the *dual surface* $X^\vee$ in $(\mathbb{P}^3)^*$,
- the *Chow threefold* $\mathrm{Ch}(X)$ in $\mathrm{Gr}(1, \mathbb{P}^3)$, consisting of lines that meet the curve X,
- the *Hurwitz threefold* $\mathrm{Hur}(X)$ in $\mathrm{Gr}(1, \mathbb{P}^3)$, of lines that are tangent to the surface X.

Every irreducible hypersurface in $\mathrm{Gr}(1, \mathbb{P}^3)$ is defined by one equation in Plücker coordinates, which is unique up to scaling and modulo the Plücker quadric. In the two cases above, this equation is called *Chow* or *Hurwitz form* of X. As we will see in Section 5.3, the irreducible components of the visual event surface of X are (iterated) singular loci of the coisotropic hypersurfaces associated to X. The developable components are dual to the singular curves in the dual surface $X^\vee$; the non-developable components are parameterized by the singular curves in the singular locus of the Chow or Hurwitz threefold of X.

Let us first consider a general smooth curve X in $\mathbb{P}^3$. The left diagram in Figure 5.2 depicts the landscape in $(\mathbb{P}^3)^*$. The dual surface $X^\vee$ consists of planes that meet X with multiplicity 2. The singular locus of $X^\vee$ is the union of two irreducible curves, whose points are osculating planes ($m = 3$) and bitangent planes ($m = 2 + 2$). The symbols that denote our loci, like $\mathcal{T}^p(X)$ and $\mathcal{E}^p(X)$, will be explained in Section 5.3. The right diagram in Figure 5.2 shows the landscape in the Grassmannian $\mathrm{Gr}(1, \mathbb{P}^3)$. We refer to [103, Theorem 1.1] for precise statements and their proofs, also for the right diagram in Figure 5.3. The singular locus of the Chow threefold $\mathrm{Ch}(X)$ is the surface (or *congruence*, cf. Chapter 2) $\mathrm{Sec}(X)$ in $\mathrm{Gr}(1, \mathbb{P}^3)$ of secant lines, *i.e.*, lines that meet X twice. The singular locus of $\mathrm{Sec}(X)$ is the curve $\mathcal{D}^\ell(X)$ of all trisecant lines. The curve $\mathcal{T}^\ell(X)$ of tangent lines is contained in $\mathrm{Sec}(X)$ but it does not belong to the singular locus.

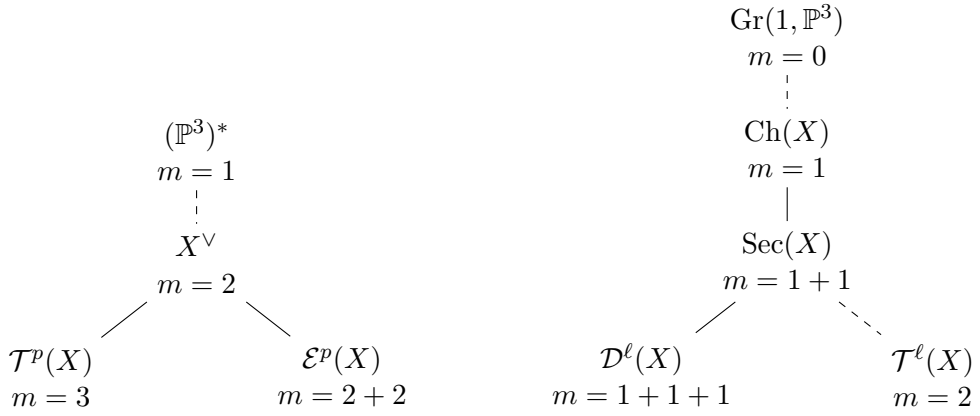


Figure 5.2: Loci of planes and lines that meet a curve X with assigned multiplicities.

In Figure 5.3, we consider various loci associated with a general smooth surface X in $\mathbb{P}^3$. The dual surface $X^\vee$ is singular along two irreducible curves. The nodal component $\mathcal{E}^p(X)$ of its singular locus is the set of all bitangent planes, and the cuspidal component $\mathcal{P}^p(X)$ is the set of all planes that intersect X with multiplicity 3 at a point (see also Figure 5.4). The Hurwitz threefold $\mathrm{Hur}(X)$ is singular along two irreducible surfaces. Its nodal component $\mathrm{Bit}(X)$ is the congruence of all bitangent lines, and its cuspidal component $\mathrm{PT}(X)$ is the congruence of principal tangents, *i.e.*, lines that meet X with multiplicity 3. The latter was denoted $\mathrm{Infl}(X)$ in [103]. These surfaces contain three special curves $\mathcal{F}^\ell(X)$, $\mathcal{C}^\ell(X)$ and $\mathcal{T}^\ell(X)$, indicating lines that meet X with multiplicity 4, or 3+2, or 2+2+2. For instance, 2+2+2 refers to tritangent lines. Sections 5.3.2 and 5.4 are devoted to the ruled surfaces in $\mathbb{P}^3$ that are represented by these curves.

For each of the five curves at the bottom of Figure 5.3, there is also an associated curve on X, which consists of the points on X where the special intersection occurs. For example, the curve associated with $\mathcal{E}^p(X)$ is the locus of points on X that lie on bitangent planes. These are the contact points on a curved object when it rolls on a table. Our favorite

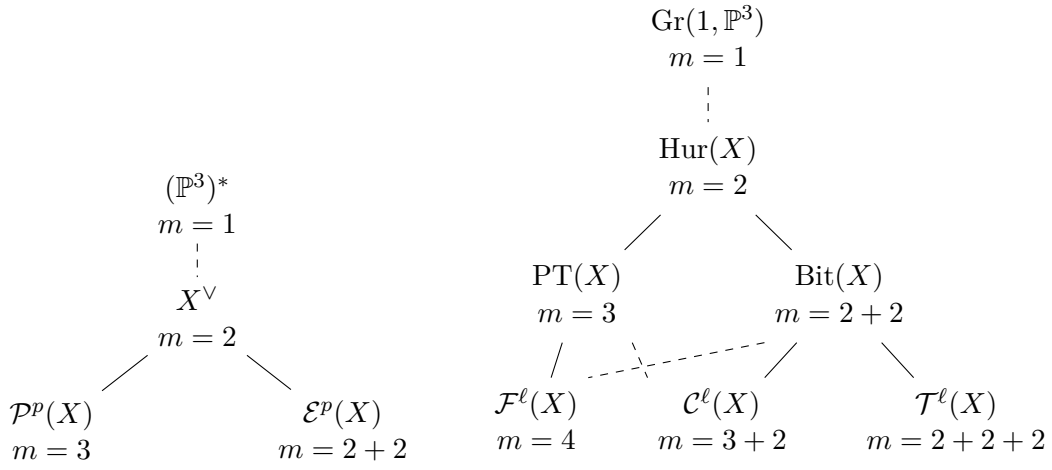


Figure 5.3: Loci of planes and lines that meet a surface X with assigned multiplicities.

terminology for this curve is due to Cayley: he calls it the *node-couple curve*. For $\mathcal{P}^p(X)$ and $\mathcal{F}^\ell(X)$, the special contact occurs at a single point, and we can give a more detailed description. At a general point x, the surface has two *principal tangents*, that are the tangent lines of the nodal curve obtained by intersecting X with its tangent plane at x. The same lines are the intersection of the tangent plane with the Hessian quadric at x defined by (here we assume that X is defined by a polynomial f in $\mathbb{R}[x_1,x_2,x_3,x_4]$)

$$y\ H_f(x)\,y^T\ =\ 0, \quad \text{where } y=[y_1,y_2,y_3,y_4] \ \text{ and } \ H_f\ =\ \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right).$$

Exceptional situations occur at *flecnodal* and *parabolic* points x. At a flecnodal point, one of the two principal tangents has intersection multiplicity 4. At a parabolic point, the Hessian matrix $H_f(x)$ drops rank, and the two principal tangents degenerate to a double line. At these points, the intersection of X with its tangent plane has a cusp at x. The locus of all parabolic points is the curve given by the intersection of X with the *Hessian surface* $\{\det(H_f)=0\}$. Over the real numbers, the parabolic curve is the boundary between the *elliptic* and *hyperbolic* regions on X, where the two principal tangents are respctively both complex or both real. The curve $\mathcal{P}^p(X)$ is the set of tangent planes at parabolic points, and the curve $\mathcal{F}^\ell(X)$ is the set of higher order principal tangents at flecnodal points. The parabolic and flecnodal curves always intersect tangentially, at special points known as *godrons* (or *cusps of the Gauss map*). Interestingly, the node-couple curve also passes through the godrons, and has the same tangent as the parabolic and flecnodal curves [159].

Remark 5.2.4. Questions regarding the singular loci of the families of lines and planes described in this section are often delicate to answer formally. This is the topic of [103]. For example, in the course of examining parabolic surfaces, we discovered a small error

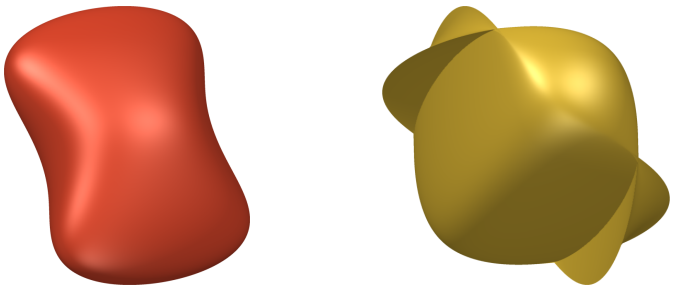


Figure 5.4: A simple quartic surface X and its dual $X^\vee$. Note that the dual presents singular curves: the nodal curve $\mathcal{E}^p(X)$ corresponds to bitangent planes for the original surface, while the cuspidal curve $\mathcal{P}^p(X)$ represents osculating planes.

in [7, Section 4], where Arrondo *et al.* consider the incidence variety for inflectional lines of X:

$$Y_2 = \{\, (x, L) \in X \times \mathrm{Gr}(1, \mathbb{P}^3) \,:\, L \text{ intersects } X \text{ at } x \text{ with multiplicity at least } 3 \,\}.$$

Lemma 4.1 (b) in [7] states that the surface Y_2 is singular at points (x, L) for which x is parabolic. This is incorrect. A general cubic surface X has a degree 12 curve of parabolic points. However, the incidence variety Y_2 is smooth. This is shown by direct computation.

5.3 Event Surfaces

In this section we describe the geometry of visual events for curves (Section 5.3.1) and surfaces (Section 5.3.2). In the case of curves we also present strategies for recovering the events computationally. For surfaces this task is more challenging, and will be addressed in Section 5.4.

Definition of events. It is convenient introduce some precise terminology and notation for the rest of the chapter. As explained in the introduction, our model for the object to be viewed is a smooth variety X of in (complex) projective space $\mathbb{P}^3$. Taking a picture of X is modeled by a linear projection $\pi : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ with pinhole z (for *zentrum*). This defines a curve $C_z(X)$ in the image plane $\mathbb{P}^2$. If X is a surface, then $C_z(X)$ is the *branch locus* of π restricted to X: this is the closure of the set of points in the image plane whose corresponding viewing lines are tangent to X at some smooth point. If X is a curve, then $C_z(X)$ is simply the closure of the image of X under π. Even if X is smooth in $\mathbb{P}^3$, the image curve $C_z(X)$ has in general many singular points in $\mathbb{P}^2$. For a surface viewed from a general viewpoint z, the only singularities in the contour are *nodes* and *cusps*; for a space curve, the image curve has only *nodes* (see Appendix B for a

discussion on nodes and cusps of plane curves). As the center z changes, the structure of its singularities is locally constant, however at some points a transition occurs, and the singularity structure changes: the *visual event surface* $\mathcal{V}(X)$ is the Zariski closure in $\mathbb{P}^3$ of the set of these transition points. The visual event surface $\mathcal{V}(X)$ usually has multiple components, corresponding to the different types of *visual events*. We will explain the geometry of these irreducible components in the rest of this section.

5.3.1 Views of curves

We begin by studying the visual events for a general curve X in $\mathbb{P}^3$ of degree d and genus g.[4] This automatically implies that X is smooth and irreducible. There are three possible visual events, corresponding to the three *Reidemeister moves* that are well-known in classical knot theory [43]. These are shown in Figure 5.5.

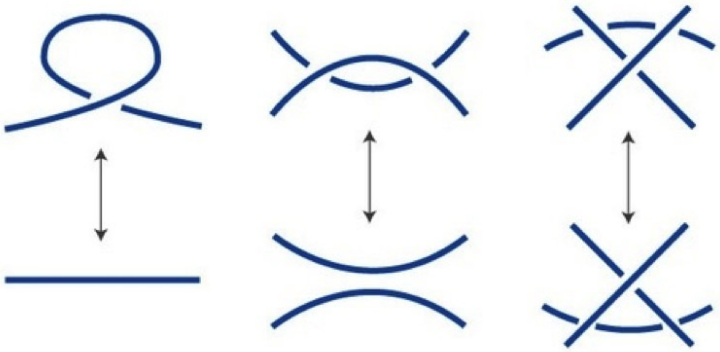


Figure 5.5: Changing views of a curve correspond to Reidemeister moves. The viewpoint z crosses the tangential surface (left), edge surface (middle), or trisecant surface (right).

The three events correspond to three components of the visual event surface $\mathcal{V}(X)$ associated with the curve X. These components can be characterized as follows:

1. The *tangential surface* $\mathcal{T}(X)$, also known as the *tangent developable*, is the union of all tangent lines to X. It represents viewpoints z such that the plane curve $C_z(X)$ has a *cusp*. When z crosses $\mathcal{T}(X)$, a node on $C_z(X)$ transitions from being real to complex.

2. The *edge surface* $\mathcal{E}(X)$ is the union of all secant lines that are edges. An *edge* is the line spanned by two points on X whose tangent lines lie in a common plane. This

[4] The *genus* is an important birational invariant of algebraic curves (and more general algebraic varieties). For smooth complex curves, it coincides with the topological notion of genus for the associated Riemann surface. Here we will not need the actual definition of the genus, but we will use the fact that space curves of fixed degree and genus admit a uniform algebraic parameterization, known as a "Hilbert scheme" [6]. In particular, this allows us to speak about the "general behavior" of space curves with a given degree and genus.

surface represents viewpoints z such that the plane curve $C_z(X)$ has a *tacnode*. When z crosses $\mathcal{E}(X)$, a pair of nodes transitions between being real and being complex.

3. The *trisecant surface* $\mathcal{D}(X)$ is the union of all lines that are spanned by triples of collinear points on X (the symbol $\mathcal{D}$ stands for *drei*). This represents viewpoints z such that $C_z(X)$ has a *triple point*. When z crosses $\mathcal{D}(X)$, the real curve $C_z(X)$ experiences a triangle crossing, but the number of real singularities is unchanged.

This classification helps explain the notation used in Section 5.2. The symbols $\mathcal{T}^\ell(X)$, $\mathcal{E}^\ell(X)$ and $\mathcal{D}^\ell(X)$ denote the curves in the Grassmannian $\mathrm{Gr}(1,\mathbb{P}^3)$ that represent the surfaces $\mathcal{T}(X)$, $\mathcal{E}(X)$ and $\mathcal{D}(X)$. The surfaces $\mathcal{T}(X)$ and $\mathcal{E}(X)$ are developable, so they can also be represented by their dual curves in $(\mathbb{P}^3)^*$:

$$\mathcal{T}^p(X) := \mathcal{T}(X)^\vee \qquad \text{and} \qquad \mathcal{E}^p(X) := \mathcal{E}(X)^\vee.$$

Here the index "p" stands for planes. The earlier used upper index "ℓ" stands for lines. The trisecant surface $\mathcal{D}(X)$ is ruled but not developable, so it has no associated curve in $(\mathbb{P}^3)^*$.

The following classical theorem characterizes the expected *degrees* of the ruled surfaces $\mathcal{T}(X)$, $\mathcal{E}(X)$ and $\mathcal{D}(X)$. These coincide with the degrees of the curves $\mathcal{T}^\ell(X)$, $\mathcal{E}^\ell(X)$ and $\mathcal{D}^\ell(X)$ inside $\mathbb{P}^5$.

Theorem 5.3.1. *For a general space curve X of degree d and genus g, the degrees of the tangential surface $\mathcal{T}(X)$, the edge surface $\mathcal{E}(X)$ and the trisecant surface $\mathcal{D}(X)$ are as follows:*

$$\begin{array}{rcl} \deg(\mathcal{T}(X)) & = & 2(d+g-1) \\ \deg(\mathcal{E}(X)) & = & 2(d-3)(d+g-1) \\ \deg(\mathcal{D}(X)) & = & \frac{(d-1)(d-2)(d-3)}{3} - (d-2)g \end{array}$$

Proof. The degree of the tangential surface $\mathcal{T}(X)$ is the Riemann-Hurwitz number $2d+2g-2$. This coincides with the degree of the dual surface $X^\vee$. See [140, page 111] for a geometric derivation and [93] for computational examples. The formula for the degree of the edge surface $\mathcal{E}(X)$ appears in [153, Theorem 2.1]. The proof given there is based on *De Jonquiéres' Formula*. The degree of the trisecant surface $\mathcal{D}(X)$ is due to Berzolari who first found it in 1895. One can find Berzolari's formula in Bertin's article [15] on the geometry of $\mathcal{D}(X)$. □

Table 5.1 summarizes the conclusion of Theorem 5.3.1 for space curves of degree d at most 6. The genus g ranges from 0 to *Castelnuovo's bound* (which describes the maximal possible genus among irreducible non-degenerate curves of given degree). Note that, for fixed d and increasing g, the degree of $\mathcal{D}(X)$ decreases while that of the others increases. In particular, there is no trisecant surface for twisted cubic curves ($d=3, g=0$) and elliptic quartic curves ($d=4, g=1$) (cf. [15, Proposition 1]). We also point out that the

d	g	$\deg(\mathcal{T}(X))$	$\deg(\mathcal{E}(X))$	$\deg(\mathcal{D}(X))$
3	0	4	0	0
4	0	6	6	2
4	1	8	8	0
5	0	8	16	8
5	1	10	20	5
5	2	12	24	2
6	0	10	30	20
6	1	12	36	16
6	2	14	42	12
6	3	16	48	8
6	4	18	54	4

Table 5.1: Degrees of the components of the visual event surface of a space curve.

edge surface $\mathcal{E}(X)$ is of importance in convex geometry because the algebraic boundary of the convex hull of a real affine curve X consists of $\mathcal{E}(X)$ and the tritangent planes of X [153].

Consider now the projection $\pi_z : X \subset \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ from a center $z \in \mathbb{P}^3 \setminus X$. The image $C_z(X)$ of X is equivalent to appropriate planar sections of the associated Chow hypersurface $\mathrm{Ch}(X)$ in $\mathrm{Gr}(1, \mathbb{P}^3)$ and of the dual hypersurface $X^\vee$ in $(\mathbb{P}^3)^*$. More precisely, writing $\alpha(z)$ the set of all lines in $\mathbb{P}^3$ that contain z and $z^\vee$ for the set of all planes in $\mathbb{P}^3$ that contain z, we have the following result.

Proposition 5.3.2. *The image $C_z(X)$ of the curve X is projectively equivalent to the curve $\alpha(z) \cap \mathrm{Ch}(X)$ in the Grassmannian $\mathrm{Gr}(1, \mathbb{P}^3)$. The curve $(C_z(X))^\vee \subset (\mathbb{P}^2)^*$ that is dual to the contour is projectively equivalent to the curve $z^\vee \cap X^\vee$ in the dual projective space $(\mathbb{P}^3)^*$.*

These facts follow from basic projective geometry (compare this with Proposition 4.3.4). The two planar curves $\alpha(z) \cap \mathrm{Ch}(X)$ and $z^\vee \cap X^\vee$ can be viewed as "intrinsic realizations" of the image curve $C_z(X)$ (*i.e.*, they are geometric objects that do not depend on choices of coordinates). Moreover, the *visual cone* associated with X (the union of all lines in the pre-image of $C_z(X)$) is a particular developable surface: the two curves from Proposition 5.3.2 can be seen as representatives for this surface in $\mathrm{Gr}(1, \mathbb{P}^3)$ and $(\mathbb{P}^3)^*$, as discussed in Section 5.2.1.

Remark 5.3.3. Inside $\mathrm{Gr}(1, \mathbb{P}^3)$ and $(\mathbb{P}^3)^*$, visual events correspond to planes that intersect $\mathrm{Ch}(X)$ and $X^\vee$ in non-generic ways. In particular, let us consider $\mathrm{Ch}(X)$. As illustrated in Figure 5.2 (right), the singular locus of $\mathrm{Ch}(X)$ is the *secant congruence* $\mathrm{Sec}(X)$ of lines meeting X twice. It is shown in [103] that this congruence has *order* $\nu = \left(\frac{1}{2}(d-1)(d-2) - g\right)$, which means that a general α-plane in $\mathrm{Gr}(1, \mathbb{P}^3)$ will always meet $\mathrm{Sec}(X)$ in exactly ν distinct points (see Chapter 2). These points correspond to

the nodes that are generically visible in the image curve $C_z(X)$. On the other hand, the number of distinct nodes will be different from ν if only if $\alpha(z)$ intersects $\mathrm{Sec}(X)$ at a singular point, or at a smooth point non-transversely. According to Figure 5.2, singular points of $\mathrm{Sec}(X)$ are trisecant lines to X. Moreover, one can show that non-transversal of a plane $\alpha(x)$ at smooth points of $\mathrm{Sec}(X)$ can occur at two types of points in $\mathrm{Gr}(1,\mathbb{P}^3)$ (or lines in $\mathbb{P}^3$): lines that are tangent to X, and lines that are "edges" for X. This gives a geometric interpretation of the visual events for curves.

Computing events for curves. We now address the task of computing three event surfaces $\mathcal{T}(X)$, $\mathcal{E}(X)$ and $\mathcal{D}(X)$ for a given space curve X. In our paper [104], we first focused on a few distinguished classes of curves, such as curves that lie on a quadric surface, or rational curves ($g = 0$) of arbitrary degree, and described some specialized computational techniques that are effective for these particular cases. Here we only present "general purpose" methods. The interested reader is referred to our paper for [104] for a more detailed discussion on special curves.

We assume that the curve X is given by its ideal $I = \langle f_1, f_2, \ldots, f_k \rangle$ in $\mathbb{R}[x_1, x_2, x_3, x_4]$. Our goal is to compute the defining polynomial F of each ruled event surface $\mathcal{S}_C$ in $\mathbb{P}^3$. If the polynomial F is too large, then we compute the ideal of the curve C in $\mathrm{Gr}(1,\mathbb{P}^3)$ or when $\mathcal{S}_C$ is developable, the ideal of the dual $\mathcal{S}_C^\vee$ in $(\mathbb{P}^3)^*$.

The edge surface $\mathcal{E}(X)$ was already discussed in [153, 162]. We therefore focus on the other two ruled surfaces in Theorem 5.3.1. The easier among them is the tangential surface $\mathcal{T}(X)$. At any given point p on the curve X, the tangent line is defined by the linear equations

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1}(p) & \frac{\partial f_1}{\partial x_2}(p) & \frac{\partial f_1}{\partial x_3}(p) & \frac{\partial f_1}{\partial x_4}(p) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_k}{\partial x_1}(p) & \frac{\partial f_k}{\partial x_2}(p) & \frac{\partial f_k}{\partial x_3}(p) & \frac{\partial f_k}{\partial x_4}(p) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{5.7}$$

To find the polynomial F defining $\mathcal{T}(X)$, we take a variable point $p = [y_1, y_2, y_3, y_4]$, and augment the ideal I with the constraints from (5.7). This gives an ideal inside $\mathbb{R}[x_1, x_2, x_3, x_4, y_1, y_2, y_3, y_4]$. From that ideal, we saturate and eliminate the variables y_1, y_2, y_3, y_4. The resulting ideal is $\langle F \rangle$.

The trisecant surface $\mathcal{D}(X)$ will be represented by its curve $\mathcal{D}^\ell(X)$ in $\mathrm{Gr}(1,\mathbb{P}^3)$. To compute this curve, we parameterize lines in $\mathbb{P}^3$ using (5.1). Suppose for now that our curve is defined by the vanishing of two forms (*i.e.*, it is a complete intersection) $X = V(f_1, f_2)$. We want the univariate polynomials $f_1(z(t))$ and $f_2(z(t))$ to have three common roots, *i.e.*, their greatest common divisor (GCD) must have degree ≥ 3. This can be expressed using *subresultants* [1]. The vanishing of all subresultants of order $i = 0, \ldots, r-1$ for two polynomials in t means that their GCD has degree at least r. In our case, we form the ideal given by the subresultant coefficients of $f_1(z(t))$ and $f_2(z(t))$

of order $0, 1$ and 2 (together with the Plücker relation). The ideal of the trisecant curve $\mathcal{D}^\ell(X)$ is obtained by saturating by the ideal of the leading coeffcients of $f_1(z(t))$ and $f_2(z(t))$. This approach also generalizes to the case when X is not a complete intersection. Indeed, if X is defined by $f_1, \ldots, f_k$, then we can use the same strategy to impose that $s_1 f_1(z(t)) + \cdots + s_{k-1} f_{k-1}(z(t))$ and $f_k(z(t))$ have three roots in common for any choice of $s_1, \ldots, s_{k-1}$.

We conclude this section with an example that illustrates the last row of Table 5.1.

Example 5.3.4. Let X be the smooth curve of degree 6 and genus 4 in $\mathbb{P}^3$ defined by

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 \;=\; x_1^3 + x_2^3 + x_3^3 + x_4^3 \;=\; 0.$$

The above method easily yields the equation of degree 18 for the tangential surface $\mathcal{T}(X)$:

$$4x_1^{12}x_2^6 - 12x_1^{12}x_2^5x_3 - 12x_1^{12}x_2^5x_4 + 21x_1^{12}x_2^4x_3^2 + \cdots + \underline{13770}x_1^6x_2^4x_3^4x_4^4 + \cdots + 24x_3^7x_4^{11} + 4x_3^6x_4^{12}.$$

This polynomial has 1094 terms. Its largest coefficient is underlined. A compact encoding is given by the 11 quadratic generators of the ideal of $\mathcal{T}^\ell(X)$. It is also easy to compute the quartic surface $\mathcal{D}(X)$, and it takes a little longer to compute the degree 54 curve $\mathcal{E}^\ell(X)$. ♢

5.3.2 Views of surfaces

We now turn to the visual events for a general surface X in projective 3-space $\mathbb{P}^3$. The six visual events associated with X were mentioned in Section 5.1.1 in items (L) and (M). These correspond to the following five irreducible components of $\mathcal{V}(X)$:

1. The *flecnodal surface* $\mathcal{F}(X)$ is the union of all lines L with contact of order 4 at a point of X. In other words, the equation of X restricted to L has a root of multiplicity 4.

2. The *cusp crossing surface* $\mathcal{C}(X)$ is the union of all lines L with contact of order $3+2$ at two points of X, *i.e.*, the equation for $X \cap L$ on L has a triple root and a double root.

3. The *tritangent surface* $\mathcal{T}(X)$ is the union of all lines L with contact of order $2{+}2{+}2$ at three points of X, *i.e.*, the equation for $X \cap L$ on L has three double roots.

4. The *edge surface* $\mathcal{E}(X)$ is the envelope of the bitangent planes of X. It is the union of all bitangent lines arising from these planes. This surface was denoted $(X^{[2]})^\vee$ in [153].

5. The *parabolic surface* $\mathcal{P}(X)$ is the envelope of all tangent planes that have contact of order 3 with X. It is the union of all principal tangents at parabolic points [139, §A.1.2].

The following theorem characterizes the expected degrees of these five surfaces.

Theorem 5.3.5. *For a general surface X of degree d in $\mathbb{P}^3$, the visual event surface $\mathcal{V}(X)$ decomposes into the five components listed above. The degrees of these surfaces are:*

$\deg(\mathcal{F}(X))$	$=$	$2d(d-3)(3d-2)$,	[159, §597],	[138, Prop. 4.5]
$\deg(\mathcal{C}(X))$	$=$	$d(d-3)(d-4)(d^2+6d-4)$,	[159, §598],	[138, Prop. 4.12]
$\deg(\mathcal{T}(X))$	$=$	$\frac{1}{3}d(d-3)(d-4)(d-5)(d^2+3d-2)$,	[159, §599],	[138, Prop. 4.10]
$\deg(\mathcal{E}(X))$	$=$	$d(d-2)(d-3)(d^2+2d-4)$,	[159, §613],	[138, Prop. 4.16]
$\deg(\mathcal{P}(X))$	$=$	$2d(d-2)(3d-4)$.	[159, §608],	[138, Prop. 4.3]

We first learned these degree formulas from Petitjean's article [138]. Only much later did we notice that all five formulas already appeared in Salmon's 1882 book [159]. The precise pointers to both sources are given above.

Remark 5.3.6. In Section 5 of our paper [104] we present new proofs for (some of) the degree formulas shown in Theorem 5.3.5. Our exposition there makes use of modern *intersection theory* [51, 61]. Our arguments are different from those of Petitjean [138], that require instead a full understanding of Colley's multiple point theory [37, 38], and also leave out several steps. On the other hand, the derivations in Salmon's book [159] are inspiring, but lack the rigor of 20th century intersection theory. Unfortunately, although techniques we used are rather well-known in algebraic geometry (they can be learned from the textbook by Eisenbud and Harris [51]), they also involve a lot of technical machinery, and are beyond the scope of this thesis. The interested reader is referred to [104] for these proofs.

In this section we explain the geometry of the five irreducible event surfaces. We begin with possible formal argument which shows that the list of visual events in Theorem 5.3.5 is indeed exhaustive.

Proof sketch of the first part of Theorem 5.3.5. Let u be an image point in $C_z(X)$ and let $L_u := \pi_z^{-1}(u)$ be the the pre-image line associated with u. Platonova [142, Main Theorem] characterizes all possible *local* singularities in the contour curve. In particular, she shows in the case $L_u \in \mathrm{PT}(X)$ that u is not a simple cusp on its branch of the contour curve if and only if L_u is either a flecnodal line or the unique principal tangent at a parabolic point. Hence, we only have to characterize the possible multi-local events. We noted above that tritangent lines and principal bitangent lines do not yield simple nodes in the contour. The final observation is that a line L_u which has contact order of exactly 2 at exactly two distinct points of the surface X projects to a simple node u if and only if L_u is not contained in a bitangent plane to X which is tangent at the same two points as L_u. □

The five ruled surfaces in Theorem 5.3.5 are encoded by the curves shown in the last row in Figure 5.3. The surfaces $\mathcal{E}(X)$ and $\mathcal{P}(X)$ are developable, and are the duals of

the singular loci shown on the left in Figure 5.3. The remaining three surfaces $\mathcal{T}(X)$, $\mathcal{C}(X)$ and $\mathcal{F}(X)$ arise from the curves in the Grassmannian $\mathrm{Gr}(1,\mathbb{P}^3)$ seen on the right of that diagram.

Remark 5.3.7. The curves of lines and planes from Figure 5.3 capture both the local and multi-local features of the surface X. This is an advantage compared to the traditional approach for studying the appearance of surfaces based on differential geometry and singularity theory. In the computer vision literature [22, 139, 146, 147, 155], prominent local features of a surface were defined in terms of the *euclidean Gauss map* and the *asymptotic spherical image*. These are maps from the surface to the unit sphere $\mathbb{S}^2$, taking a point on $X_{\mathbb{R}}$ to its normal direction, or to the direction of one of its principal tangents. In our algebro-geometric setting, the role of $\mathbb{S}^2$ is played by the dual surface $X^\vee \subset (\mathbb{P}^3)^*$ and the principal tangent surface $PT(X) \subset \mathrm{Gr}(1,\mathbb{P}^3)$. These surfaces carry much more information than the unit sphere $\mathbb{S}^2$.

Consider now the projection $\pi_z : X \subset \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ from a center $z \in \mathbb{P}^3 \backslash X$. The follwing result, analogous to Proposition 5.3.2, describes "intrinsic realizations" of the contour $C_z(X)$.

Proposition 5.3.8. *The contour $C_z(X)$ of our surface X is projectively equivalent to the curve $\alpha(z) \cap \mathrm{Hur}(X)$ in the Grassmannian $\mathrm{Gr}(1,\mathbb{P}^3)$. The curve $(C_z(X))^\vee \subset (\mathbb{P}^2)^*$ that is dual to the contour is projectively equivalent to the curve $z^\vee \cap X^\vee$ in the dual projective space $(\mathbb{P}^3)^*$.*

As noted for curves in Remark 5.3.3, visual events are associated with planes $\alpha(z)$ that intersect $\mathrm{Hur}(X)$ and $X^\vee$ non-generic ways. In particular, let u be a point in the image curve $C_z(X)$, and let $L_u = \pi_z^{-1}(u)$ be the associated pre-image line in $\mathrm{Gr}(1,\mathbb{P}^3)$. This line is a point of contact between $\alpha(z)$ and $\mathrm{Hur}(X)$ inside $\mathrm{Gr}(1,\mathbb{P}^3)$. Assuming that the intersection of $\alpha(z)$ and $\mathrm{Hur}(X)$ is transversal at L_u, then we have that:

- If L_u is a general point of $\mathrm{Hur}(X)$, then u is a *smooth point* of the curve $C_z(X)$.
- If L_u is a general point of $\mathrm{Bit}(X) \subset \mathrm{Hur}(X)$, then u is a *simple node* of $C_z(X)$.
- If L_u is a general point of $\mathrm{PT}(X) \subset \mathrm{Hur}(X)$, then u is a *simple cusp* of $C_z(X)$.

Another interesting fact (shown *e.g.* in [146]) is that if L_u is a (non-principal) tangent line at a parabolic point, then u is a flex point of $C_z(X)$.

Cuspidal and nodal singularities exist for any viewpoint z, since $\alpha(z)$ always intersects the congruences $\mathrm{Bit}(X)$ and $\mathrm{PT}(X)$ in $\mathrm{Gr}(1,\mathbb{P}^3)$. In fact, it is shown in [103] that the orders of these congruences are respectively $\nu = \frac{1}{2}d(d-1)(d-2)(d-3)$ and $\kappa = d(d-1)(d-2)$, which means that a general image curve will have exactly ν distinct nodes and κ distinct cusps (over $\mathbb{C}$). On the other hand, for non generic planes $\alpha(z)$ this will not be the case. First, $\alpha(z)$ can intersect the singular loci of $\mathrm{Bit}(X)$ and $\mathrm{PT}(X)$. According to Figure 5.3, these singular loci consist of $\mathcal{T}^\ell$, $\mathcal{C}^\ell$ and $\mathcal{F}^\ell$. These sets of lines correspond to special singularities in the image curve $C_z(X)$.

($\mathcal{T}$) If L_u is a tritangent line, then u is a triple point. This is a *triple point* event.

($\mathcal{C}$) If L_u is a principal bitangent, then $C_z(X)$ has a smooth branch and a cuspidal branch that meet at u. This is a *cusp crossing* event.

($\mathcal{F}$) If L_u is a flecnodal line, then u is the limit of two cusps and a node, *i.e.* an infinitesimal change of the viewpoint produces two cusps and a node. This is a *swallowtail* event.

Alternatively, $\alpha(z)$ can meet $\mathrm{Bit}(X)$ or $\mathrm{PT}(X)$ at a smooth point but not transversely. One can show that this occurs when $\alpha(z)$ contains a line L_u that is either a bitangent line on a bitangent plane, or a principal tangent at a parabolic point. This corresponds to the singular loci $\mathcal{E}^p(X)$ and $\mathcal{P}^p(X)$ of the dual surface $X^\vee$. Once again this gives rise to special singularities in $C_z(X)$.

($\mathcal{E}$) If L_u is a bitangent line on a bitangent plane, then the associated image point u is a tacnode. It is obtained as the limit of two smooth branches coming together at u. This is a *tangent crossing* event.

($\mathcal{P}$) If L_u is the principal tangent at a parabolic point p, then, over the real numbers, two behaviors are possible: either u is an isolated node, which corresponds a *lip* event, or u is a tacnode, obtained as the limit of two cusps, which is a *beak-to-beak* event. See below for a discussion.

The six visual events are shown in Figure 5.6. Similar pictures are quite common in the computer vision literature: for instance, see [139, Figures 5 and 6], and Figures 13.20 through 13.25 in the textbook [57]. In Figure 5.7, we show some actual examples of parabolic and flecnodal surfaces.

Parabolic events. We now briefly explain how to distinguish two possible local behaviors ("lip" vs. "beak-to-beak") of the contour when the viewpoint z belongs to the parabolic surface $\mathcal{P}(X)$. As argued in Section 5.2.1, the parabolic surface $\mathcal{P}(X)$ is a *developable surface*, since it is dual to the curve $\mathcal{P}^p(X)$ in $(\mathbb{P}^3)^*$. In particular, all principal tangents at parabolic points are the tangents of the edge of regression curve, denoted by $E(\mathcal{P}(X))$. This curve is different from the parabolic curve X, and in fact we can associate each parabolic point x with another point e_x, where the principal tangent from x meets (and in fact is tangent to) $E(\mathcal{P}(X))$. In real projective space, the complement of $\{x, e_x\}$ in that line has two connected components. The distinction between lip and beak-to-beak is made by which of these two components the viewpoint z belongs to. This fact is shown in [142, Theorem 4.10].

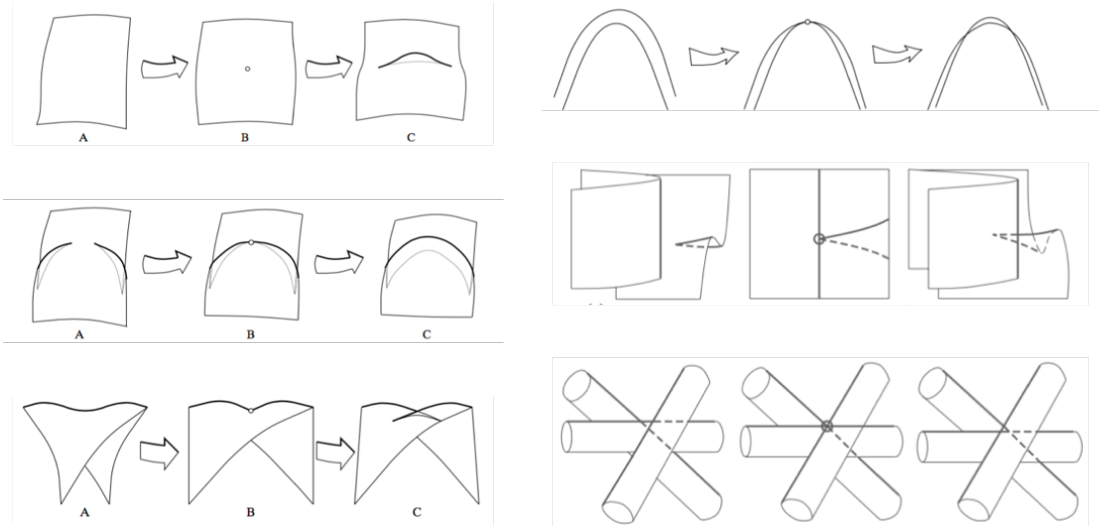


Figure 5.6: The catalogue of visual events for the projections of a smooth surface from a viewpoint that moves. The local events (*left*, from top to bottom) are lip, beak-to-beak, swallowtail. The multi-local events (*right*, from top to bottom) are tangent crossing, cusp crossing, triple point. Reprinted from [134] with permission of Springer.

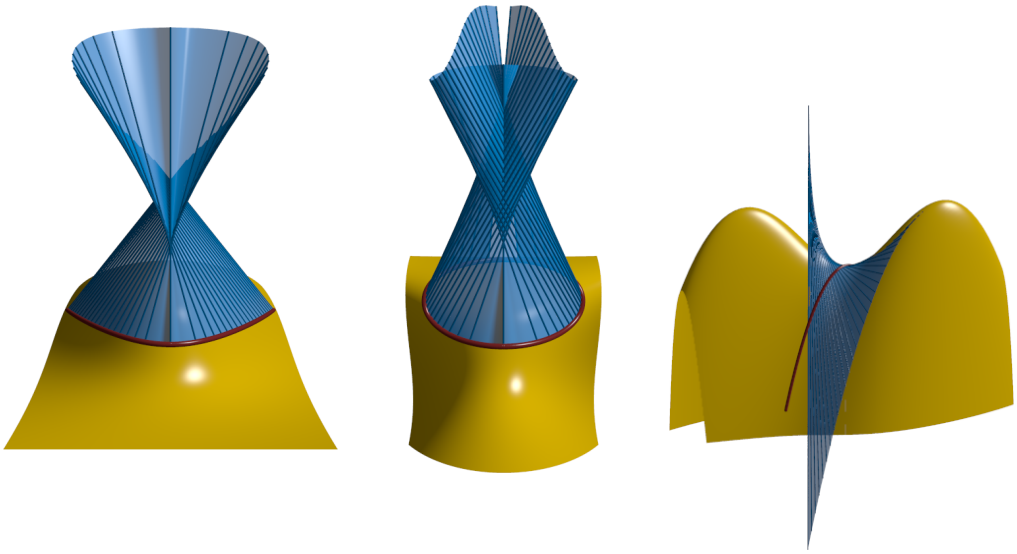


Figure 5.7: Visual event surfaces. The first two figures show two parabolic surfaces, corresponding to a lip event (*left*) and a beak-to-beak event (*center*). The parabolic curves on the surface are drawn in red. The last figure shows a flecnodal curve and a flecnodal surface.

Real-life visual events. We conclude this section with an informal discussion that should provide an intuitive understanding of our five event surfaces. The following are some real life situations where these events can actually be observed.

We first note that cuspidal and nodal singularities of image contours are stable features, which are visible in most surfaces. Nodes occur whenever occlusions create discontinuities in the contour. Cusps only appear for non-convex objects. For instance, they can be observed on the folds of a piece of cloth. From an exceptional viewpoint, it is possible that several of these singularities occur along the same visual ray. This gives rise to a multi-local visual event (cusp crossing, tangent crossing, or triple points). One can spot these events with a napkin or a towel.

The three local events on the left in Figure 5.6 are more complicated. It takes some practice to discover them in the real world. Here are some concrete examples:

- *Lip event:* If we observe a small hill from a high aereal viewpoint (say, from a hot air balloon), then all points on the ground are visible. The hill does not generate an image contour. However, as we descend closer to the ground, the profile of the hill suddenly becomes visible in the contour. This qualitative change of appearance is a lip event.

- *Beak-to-beak.* Observe a glass bottle from the bottom, with your eye close to the base. You see a part of the contour generated by the convex region where the sectional diameter of the bottle decreases. Now, tilt the bottle slowly towards its upright position. At some point, you see a complete path from the base to the top of the bottle. Previously your view had been blocked. This is a beak-to-beak event. Contrary to the lip event, the contour does not disappear at the transition point, but it breaks into two pieces.

- *Swallowtail.* The traditional drawing of a (transparent) torus presents two swallowtails. We see both cuspidal and nodal singularities in the contour [155, Fig. 2]. As we rotate the torus, a visual event occurs, and these singularities disappear.

5.4 Computing Visual Events

We now return to the computation of visual event surfaces. Theorem 5.3.5 gives the degrees of the irreducible components of the visual event surface when X is a general surface of degree d in $\mathbb{P}^3$. Table 5.2 summarizes these degrees for $d \leq 7$. One notices that the degrees are now much larger than those for curves in Table 5.1.

d	$\deg(\mathcal{F}(X))$	$\deg(\mathcal{C}(X))$	$\deg(\mathcal{T}(X))$	$\deg(\mathcal{E}(X))$	$\deg(\mathcal{P}(X))$
3	0	0	0	0	30
4	80	0	0	160	128
5	260	510	0	930	330
6	576	2448	624	3168	672
7	1064	7308	3808	8260	1190

Table 5.2: Degrees of the components of the visual event surface of a general surface.

The degrees in Table 5.2 pose a challenge because a homogeneous polynomial in four unknowns of degree δ can have as many as $\binom{\delta+3}{3}$ terms. For instance, if X is a quintic surface then its flecnodal surface $\mathcal{F}(X)$ has degree $\delta = 260$, so the expected number of terms is $\binom{\delta+3}{3} = 2997411$. In this section we address this challenge.

5.4.1 Multiple-root loci

Throughout this section, we make use of *multiple root loci* for binary forms. These are varieties that identify the coefficients of polynomials with multiple roots. More precisely, we consider the coefficients $c_0, c_1, \dots c_d$ of

$$c_0 t^d + c_1 t^{d-1} + c_2 t^{d-2} + \cdots + c_{d-1} t + c_d. \tag{5.8}$$

For a partition $\lambda = (\lambda_1, \dots, \lambda_k) \in \mathbb{N}^k$ with $\sum_{i=1}^k \lambda_i \leq d$, we write $\Delta_\lambda(d)$ for the ideal in $\mathbb{R}[c_0, \dots, c_d]$ representing polynomials (5.8) that have k complex roots with multiplicities $\lambda_1, \dots, \lambda_k$. The varieties are called multiple root loci in [112]. For example, if $d \geq 4$, then $\Delta_{(4)}(d)$ is the ideal of coefficients of polynomials of degree d with one quadruple root.

We refer to [112] for details on the ideals $\Delta_\lambda(d)$. Some relevant instances are listed in Table 5.3. Its entries are copied from [112, Table 1]. For instance, the entry "$6^{10}, 8^{38}$" in the last column means that $\Delta_{(3,2)}(7)$ is minimally generated by 10 polynomials of degree six and 38 polynomials of degree eight.

The ideals $\Delta_\lambda(d)$ can be computed explicitly using *subresultants* [1]. The i-th subresultant $S_i(h_1, h_2)$ of two polynomials $h_1(t), h_2(t)$ is a polynomial of degree at most i whose coefficients are the determinants of particular minors of the Sylvester matrix of h_1 and h_2. The vanishing of $S_i(h_1, h_2)$ for $0 \leq i \leq d-1$ means that the greatest common divisor (GCD) of h_1 and h_2 has degree at least d. Moreover, if $S_d(h_1, h_2)$ not zero, it is exactly this GCD. If we let h_d be the polynomial (5.8) and h'_d be its derivative with respect to t, then the condition that h_d has roots with multiplicity $\lambda = (\lambda_1, \dots, \lambda_k)$ is equivalent to the fact that the GCD of h_d and h'_d has degree $\sum_{i=1}^k (\lambda_i - 1)$ and has roots with multiplicities $\lambda' = (\lambda_1 - 1, \dots, \lambda_k - 1)$.

This allows us to compute ideal $\Delta_\lambda(d)$ recursively. For example, let us assume that we have recovered the conditions for a polynomial to have a double root ($\lambda = (2)$) and a triple root ($\lambda = (3)$) for all degrees up to $d-1$. We consider the ideal I defined by $S_0(h_d, h'_d)$, $S_1(h_d, h'_d)$ and $S_2(h_d, h'_d) = 0$, saturated by the leading coefficient c_0. This ideal decomposes into three components corresponding to $\Delta_{(4)}(d), \Delta_{(3,2)}(d)$ and $\Delta_{(2,2,2)}(d)$. The component $\Delta_{(4)}(d)$ can be recovered by adding to I the conditions for $S_3(h_d, h'_d)$ to have a triple root. The component $\Delta_{(3,2)}(d)$, is obtained by adding I the conditions for $S_3(h_d, h'_d)$ to have a double root, and saturating by the condition for it to have a triple root. The component $\Delta_{(2,2,2)}(d)$ is obtained by saturating I for the condition that $S_3(h_d, h'_d)$ has a double root.

Ruled surface	Partition	$d=4$	$d=5$	$d=6$	$d=7$
$\mathcal{F}(X)$	$\lambda=(4)$	2^6	2^3	$2^1, 3^3, 4^1$	4^{20}
$\mathcal{C}(X)$	$\lambda=(3,2)$		4^{28}	$4^1, 5^3, 6^{31}$	$6^{10}, 8^{38}$
$\mathcal{T}(X)$	$\lambda=(2,2,2)$			4^{45}	6^{78}

Table 5.3: The ideals $\Delta_{(\lambda)}(d)$ of multiple root loci relevant for visual events of surfaces.

5.4.2 Events for surfaces

In what follows we assume that the ideals $\Delta_\lambda(d)$ have been pre-computed for $d \leq 7$. We use these ideals to compute the curves $\mathcal{F}^\ell(X)$, $\mathcal{C}^\ell(X)$ and $\mathcal{T}^\ell(X)$ in the Grassmannian $\mathrm{Gr}(1, \mathbb{P}^3)$. The correspondence between the three multi-local events $\mathcal{F}, \mathcal{C}, \mathcal{T}$ and the three special partitions λ was seen on the right side in Figure 5.3, where λ was denoted by m.

Let $f = f(x_1, x_2, x_3, x_4)$ be the polynomial of degree d that defines the surface X. We parameterize points $z(t)$ on the line in $\mathbb{P}^3$ with Plücker coordinates p using a parameter t as in (5.1). We substitute $z(t)$ into the polynomial f, and we regard $f(z(t))$ as a univariate polynomial in t, written as in (5.8). The coefficients c_i are now homogeneous expressions of degree d in the Plücker coordinates q. At this point, we substitute these expressions $c_i(q)$ into the generators of $\Delta_\lambda(d)$. The result is an ideal in the Plücker coordinates q that defines the desired curve set-theoretically.

Example 5.4.1. Let $d = 5$ and consider the smooth quintic surface X defined by

$$f \;=\; x_1^5 + x_2^5 + x_3^5 + x_4^5 + (x_1 + x_2 + x_3 + x_4)^5 + x_1x_2x_3x_4(x_1 + x_2 + x_3 + x_4).$$

We compute the curve $\mathcal{F}^\ell(X)$ in $\mathrm{Gr}(1, \mathbb{P}^3)$ that represents the flecnodal surface. Its prime ideal has degree 260 and is generated by 10 sextics plus the Plücker quadric. This computation was done with the method above, starting from the ideal $\Delta_{(4)}(5)$. ♢

Let us shift gears and focus on the local events $\mathcal{P}$ and $\mathcal{E}$, seen on the left in Figure 5.3. We start with the parabolic surface $\mathcal{P}(X)$. Let X be defined by a polynomial $f \in \mathbb{R}[x_1, x_2, x_3, x_4]$. The ideal $I(P)$ of the parabolic curve P is defined by f and the determinant of the Hessian matrix H_f. Consider the incidence variety of the parabolic curve and its tangent planes, that is $\{(x, T_x(X)) \mid x \in P\} \subset \mathbb{P}^3 \times (\mathbb{P}^3)^*$. We compute the ideal of the incidence variety by adding the 2×2-minors of the matrix $\begin{pmatrix} \partial f/\partial x_1 & \partial f/\partial x_2 & \partial f/\partial x_3 & \partial f/\partial x_4 \\ y_1 & y_2 & y_3 & y_4 \end{pmatrix}$ to $I(P)$. We then saturate the resulting ideal by $\langle x_1, x_2, x_3, x_4 \rangle$ and afterwards eliminate x_1, x_2, x_3, x_4. This furnishes the ideal of the dual curve $\mathcal{P}^p(X)$ in $(\mathbb{P}^3)^*$, which encodes the developable surface $\mathcal{P}(X)$.

For a general parabolic point x of X, the Hessian matrix $H_f(x)$ has rank three, so its nullspace represents a unique point p_x in $\mathbb{P}^3$. We use the following simple fact to compute $\mathcal{P}^\ell(X)$.

Lemma 5.4.2. *For $x \in P$, the point p_x lies on the unique principal tangent of X at x.*

Proof. The relation $x\, H_f(x)\, p_x^T = 0$ holds. Euler's relation shows that $x\, H_f(x)$ is the gradient vector of f at x. Hence p_x lies on the tangent plane to X at x. Furthermore, p_x belongs to the principal tangent since $p_x\, H_f(x)\, p_x^T$ is zero. Hence x and p_x span the principal tangent. □

The curve $\mathcal{P}^\ell(X)$ in $\mathrm{Gr}(1, \mathbb{P}^3)$ is precisely the collection of the lines spanned by a general parabolic point x and the corresponding point p_x from Lemma 5.4.2. This allows us to compute the ideal of $\mathcal{P}^\ell(X)$ in Plücker coordinates $p_{12}, p_{13}, \ldots, p_{34}$. First, we recover the ideal I of the incidence variety $\{(x, y) \,|\, x \in P,\, y \in \ker H_f(x)\}$ by adding the four entries of the column vector $H_f(x) \cdot y$ to the ideal $I(P) = \langle f, \det H_f(x) \rangle$. Secondly, we consider the map from the coordinate ring of the Grassmannian to the quotient ring of I that maps Plücker coordinates q_{ij} to the 2×2-minors of $\left(\begin{smallmatrix} x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \end{smallmatrix}\right)$. The kernel of this ring map is the ideal of the curve $\mathcal{P}^\ell(X)$ in $\mathrm{Gr}(1, \mathbb{P}^3)$.

Example 5.4.3. Let $d = 3$ and consider the *Fermat cubic* X defined by $f = x_1^3 + x_2^3 + x_3^3 + x_4^3$. We can easily compute the ideal of the curve $\mathcal{P}^\ell(X)$ as described above, and from this we find the parabolic surface $\mathcal{P}(X)$. It decomposes into irreducible components of low degree:

$$\begin{aligned}
&(x_0 + x_1)(x_0 + x_2)(x_0 + x_3)(x_1 + x_2)(x_1 + x_3)(x_2 + x_3) \\
&\cdot (x_0^2 - x_0x_1 + x_1^2)(x_0^2 - x_0x_2 + x_2^2)(x_0^2 - x_0x_3 + x_3^2) \\
&\cdot (x_1^2 - x_1x_2 + x_2^2)(x_1^2 - x_1x_3 + x_3^2)(x_2^2 - x_2x_3 + x_3^2) \\
&\cdot (x_1^3 + x_2^3 + x_3^3)(x_0^3 + x_2^3 + x_3^3)(x_0^3 + x_1^3 + x_3^3)(x_0^3 + x_1^3 + x_2^3).
\end{aligned}$$

This is one of the few cases where symbolic computation of the equation of $\mathcal{P}(X)$ is easy. ♢

Example 5.4.4. Let $d = 3$ and let X be the cubic surface defined by $f = x_1^3 + x_2^3 + x_3^3 + x_4^3 + (x_1 + 2x_2 + 3x_3 + 4x_4)^3$. Using the method above, we easily compute the ideal of $\mathcal{P}^\ell(X)$. We next demonstrate how to find the visual events of type $\mathcal{P}$ as the camera moves along a line.

Consider the line with parametric representation $z(t) = [t, 1, t-1, t+1]$ in $\mathbb{P}^3$. Let Q be the skew-symmetric 4×4 matrix obtained from (5.2) by substituting to Plücker coordinates. We add the four coordinates of $z(t) \cdot Q$ to the ideal of $\mathcal{P}^\ell$, we then saturate with respect to $\langle p_{12}, \ldots, p_{34} \rangle$, and thereafter we eliminate the unknowns p_{ij}. The result is

$$\begin{aligned} &4954039466358213551576831457 28t^{30} + 434950525322602430919258122035 2t^{29} \\ &+1843773930667965426193833894643 2t^{28} + 505623210540135536148084632789 12t^{27} \\ &+ \cdots \cdots - 8150915394320070700 8t^2 - 188527342464707308 8t - 19650742648215232. \end{aligned}$$

This polynomial has 30 distinct complex roots. Precisely 8 of them are real, namely

$$\begin{aligned} \{&-1.0135860298526, -1.01135228952, -0.60097492358065, -0.3501467610081, \\ &-0.266855069244, -0.19167605663, -0.081116156693251, 0.37894374773077\}. \end{aligned}$$

These 8 roots mark the visual events of type $\mathcal{P}$ as the viewpoint travels along the line $z(t)$. ♢

We found that the computation the edge surface $\mathcal{E}(X)$ is more challenging than that of the parabolic surface $\mathcal{P}(X)$. Consider the case when X is a general quartic: the surface $\mathcal{E}(X)$ has in this case degree 160, and hence so does the curve $\mathcal{E}^\ell(X)$ in $\mathrm{Gr}(1, \mathbb{P}^3)$. We succeeded in computing the ideal of this curve only for quartics X that are singular or very special. For instance, if X is the Fermat quartic then $\mathcal{E}(X)$ a surface of degree 80, with multiplicity 2. Since $\mathcal{E}(X)$ is developable, we could also try to use $\mathcal{E}^p(X)$ as an encoding; unfortunately, the degree is then even higher. By Proposition 5.1 in [104], in fact, the dual curve $\mathcal{E}^p(X)$ has degree 480 in $(\mathbb{P}^3)^*$. The computation of edge surfaces $\mathcal{E}(X)$ definitely requires further research.

5.5 Conclusions

We have studied visual events of general algebraic curves and surfaces in projective space. All the events can be characterized geometrically in terms of "associated hypersurfaces" in $(\mathbb{P}^3)^*$ and $\mathrm{Gr}(1, \mathbb{P}^3)$, which contain planes and lines that intersect the observed object in exceptional ways. Computing the visual events for general algebraic surfaces is challenging, because the degrees of the event surfaces increase very rapidly (see Table 5.2). We have discussed some strategies for addressing this issue: in particular, it is convenient to represent ruled and developable surfaces as curves in $\mathrm{Gr}(1, \mathbb{P}^3)$ of $(\mathbb{P}^3)^*$, and to pre-compute ideals of multiple-point loci. Further work will be necessary for efficiently computing the edge surface. Moreover, we did not discuss here the computation of the

aspect graph [22]. This graph describes the topology of the complement of the event surface inside real projective space, so its computation requires tools from real algebraic geometry (such as the tracing procedure described in [139]).

In this chapter we have considered only pinhole projections, however it is possible to define visual events for more general imaging models as well. More precisely, if C is a line congruence in $\mathrm{Gr}(1,\mathbb{P}^3)$ representing a geometric camera (cf. Chapter 2), and if X is the curve or surface that is being observed, then a visual event will occur when C contains a line that is a generator for one of the ruled surfaces described in Section 5.3, that are associated with X. Equivalently, C must intersect certain special curves inside $\mathrm{Gr}(1,\mathbb{P}^3)$. A difficulty here is that we cannot treat the event surfaces as actual "boundaries", since a camera is not identified with a single viewpoint. On the other hand, we can apply different motions to a camera, and ask when the resulting congruence intersects the aforementioned curves in $\mathrm{Gr}(1,\mathbb{P}^3)$. For example, if we apply a translational (1D) motion to the congruence, then by dimensionality arguments there will be a finite number of points where these visual events take place. We expect the nature (and appearance) of the "generic" visual events to be the same as in the traditional pinhole case. However, allowing for a higher dimensional motion space, there will be more complex behaviors for exceptional (higher codimension) camera locations.

Although algebraic curves and surfaces are usually not well suited for modeling real-life objects, we believe that our idealized approach provides useful tools for understanding visual events. In particular, restricting ourselves to an algebraic framework does not affect the overall geometric picture, but avoids some technical difficulties that arise when using the traditional differential geometric viewpoint. More precisely, we can use the fact that in an algebraic setting, and particularly in projective space and over the complex numbers, parameterized geometric systems typically have a unique "general" behavior. This allows us to easily characterize exceptional projections, without having to use notions of stability of germs for differential functions. On a more practical side, describing visual events in terms of the incidences of a variety with lines and planes is useful for capturing simultaneously *local* and *multilocal* instabilities. As noted in Remark 5.3.7, previous geometric descriptions of visual events made use of the *Euclidean Gauss map* and the *asymptotic spherical image* [56]. These maps can be used to understand the three local visual events, *i.e.*, those corresponding to the parabolic surface (lip and beak-to-beak events) and to the flecnodal surface (swallowtail events). On the other hand, in our study we considered analogous objects in $(\mathbb{P}^3)^*$ and $\mathrm{Gr}(1,\mathbb{P}^3)$: the *dual surface* $X^\vee$ and the *congruence of principal tangents* $\mathrm{PT}(X)$ (which respectively correspond to the two maps mentioned above), as well as the *congruence of bitangents* $\mathrm{Bit}(X)$ (which was not previously represented as a map into $\mathbb{S}^2$). We have seen that these sets of planes and lines can be used to characterize *all* visual events.

Conclusions

This thesis was devoted to mathematical models for describing the geometry of imaging processes. The first three chapters focused on models for cameras: Chapter 1 investigated topics on the multi-view geometry of pinhole cameras; Chapter 2 studied general imaging systems as mappings from points to lines; Chapter 3 extended this last geometric model with an analytical map associating viewing lines to coordinates. The final two chapters discussed, from different perspectives, the relationship between 3D objects and their images. Chapter 4 considered arbitrary sets in $\mathbb{P}^3$, and the consistency conditions that are naturally satisfied by their projections in a finite set of images. Chapter 5 focused on smooth curves and surfaces, and described the corresponding visual events by studying sets of lines in $\mathrm{Gr}(1, \mathbb{P}^3)$ with special incidence properties with the observed curve or surface.

Possible extensions of our work were discussed in detail in the conclusions of the chapters, but we recall here some general directions that we believe are worth pursuing. For example, it could be useful to develop more concrete models for representing the space of configurations of pinhole cameras (or, equivalently, for parameterizing joint images). In some mathematical papers, the space of camera configurations has been identified with a component the of a "Hilbert scheme" parameterizing varieties with an assigned Hilbert function [4, 116]. On the other hand, by using so-called "reduced" cameras, one immediately sees that the configuration space of n cameras is intimately related with the space of configurations of $n + 4$ points in $\mathbb{P}^3$. We are currently working on spelling out this relationship, using *Carlsson-Weinshall duality* [31] and some classical works on point configurations [36].

Turning to the non central imaging systems from Chapters 2 and 3, we hope to gain a better understanding of the action of *refractions* and *reflections* on line congruences, and use this to develop better models for representing catadioptric cameras. We also find it amusing to *build* physical cameras associated with general congruences. For example, we have assembled a prototype of a (1,2)-camera: this device consists of a box with two slits, one straight and one circular, that generates an image corresponding to the set of light rays passing through both slits. In the future, we would like to construct more general systems as well, possibly using programmable directional sensors similar to the

flexible cameras described in [130].

Finally, we believe that the study of contact properties between viewing rays and objects, along the lines of our discussion in Chapters 4 and 5, is a topic of great importance for both computer vision and computer graphics. The *plenoptic function* [2] and the *light field* [114], for example, are fundamental representations of a 3D scene, that encode photometric information for producing images from all viewpoints. These functions are complex (they have 5D and 4D parameter space), but the representations also have some redundancies, since the maps must satisfy certain conditions in order to describe an actual scene. This is in some sense a "continuous" version of our notion of multi-view consistency from Chapter 4. Mathematically, this also seems related to the study of varieties in $\mathrm{Gr}(1, \mathbb{P}^3)$ that are defined by objects in $\mathbb{P}^3$. These are essentially the "coisotropic" varieties described in [64, 102] and considered in Chapter 5. Developing more concrete models for these varieties, perhaps in terms of *discrete differential geometry* [17] relative to 3D meshes, could be useful to apply many theoretical ideas from this thesis in practical situations.

Appendix A

Projective Geometry and Plücker Coordinates

This Appendix contains a brief introduction to projective geometry (Section A.1), and then collects some basics on Grassmannians and line geometry (Section A.2).

A.1 Projective Geometry

The following is a very brief introduction to projective geometry. This material (including proofs) can be found in any standard book on the subject, for example [40, 163]. On the other hand, many classical references such as these use a different style and older language. Our organization follows the first part of the exercise book [58], which is modern and concise.

Basic definitions. Let $\mathbb{K}$ be a field, and let V be a finite-dimensional vector space over $\mathbb{K}$. The *projective space* associated with V is defined as

$$\mathbb{P}(V) = (V \setminus \{0\}) / \sim,$$

where two vectors $\mathbf{v}$ and $\mathbf{w}$ satisfy the relation $\mathbf{v} \sim \mathbf{w}$ if and only if $\mathbf{v} = \lambda\mathbf{w}$ for some λ in $\mathbb{K} \setminus \{0\}$. A *projective point* is an equivalence class $[\mathbf{v}] = \{\lambda v \,|\, \lambda \in \mathbb{K}\}$. The *dimension* of $\mathbb{P}(V)$ is defined to be $\dim V - 1$. Note that if $V = \{0\}$ then $\mathbb{P}(V)$ has dimension -1.

If $V = \mathbb{K}^{n+1}$, then $\mathbb{P}^n = \mathbb{P}(\mathbb{K}^{n+1})$ is the *standard projective space* of dimension n over $\mathbb{K}$. We will write $[\mathbf{x}] = [x_1, \ldots, x_{n+1}]$ for the class of the vector $\mathbf{x} = (x_1, \ldots, x_{n+1})$ in $\mathbb{K}^{n+1}$.

If W is a linear subspace of V, then $\mathbb{P}(W)$ is a *projective subspace* (or *linear subspace*) of $\mathbb{P}(V)$. Concretely, a projective subspace of dimension d corresponds to the set of vector lines within a linear vector space of dimension $d+1$.

Incidence between subspaces. The intersection of two (or more) projective subspaces is a always a projective subspace. Indeed, if $S_1 = \mathbb{P}(W_1)$, $S_2 = \mathbb{P}(W_2)$ are projective subspaces of $\mathbb{P}(V)$, then $\mathbb{P}(W_1) \cap \mathbb{P}(W_2) = \mathbb{P}(W_1 \cap W_2)$. We will often write the intersection of two projective subspaces using the *meet* ($\wedge$) operator, so that $S_1 \wedge S_2 = S_1 \cap S_2$. We say that S_1, S_2 are *incident* if $S_1 \wedge S_2 \neq \emptyset$, otherwise they are *skew*.

The union of two projective subspaces is not a projective subspace. However, we use the *join* ($\vee$) operator, and write $S_1 \vee S_2$ for the smallest projective subspace containing $S_1 \cup S_2$. For example, given two points P, Q in $\mathbb{P}^3$, we write $P \vee Q$ for the line they span.

Proposition A.1.1. *Let $S_1 = \mathbb{P}(W_1)$, $S_2 = \mathbb{P}(W_2)$ be projective subspaces of $\mathbb{P}(V)$. Then*

$$\dim(S_1 \vee S_2) = \dim(S_1) + \dim(S_2) - \dim(S_1 \wedge S_2).$$

In particular, if $\dim S_1 + \dim S_2 \geq \dim \mathbb{P}(V)$, then S_1 and S_2 cannot be skew.

Projective Transformations. Linear maps between vector spaces induce transformations between the corresponding projective spaces. More precisely, if V and W are vector spaces, and $\varphi : V \to W$ is a linear map with nullspace K, then the associated *projective map* $\bar{\varphi} : \mathbb{P}(V) \setminus \mathbb{P}(K) \to \mathbb{P}(W)$ is defined by $\bar{\varphi}([\mathbf{v}]) = [\varphi(\mathbf{v})]$ for all $\mathbf{v}$ not in K. In particular, if φ is a linear isomorphism (so $K = \{0\}$), then $\bar{\varphi} : \mathbb{P}(V) \to \mathbb{P}(W)$ is called a *projective isomorphism*, or *homography*.

Homogeneous coordinates. Points $P_1 = [\mathbf{v}_1], \ldots, P_k = [\mathbf{v}_k]$ inside a projective space $\mathbb{P}(V)$ of dimension n are said to be *linearly independent* if $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent vectors in V. Clearly, if $\mathbb{P}(V)$ has dimension n, at most $n+1$ points can be linearly independent. More generally, $P_1 = [\mathbf{v}_1], \ldots, P_k = [\mathbf{v}_k]$ with $k \geq n+1$ are in *general position* if any subset of at most $n+1$ points among them is linearly independent.

An ordered set $\mathcal{R} = \{P_1, \ldots, P_{n+2}\}$ of $n+2$ points in general position in $\mathbb{P}(V)$ is called a *projective reference frame*. The points $P_1, \ldots, P_{n+1}$ are called *fundamental points*, while P_{n+2} is the *unit point*. A projective reference frame induces a system of *homogeneous coordinates* on $\mathbb{P}(V)$ as follows. For any $\mathbf{u} \in V$ such that $[\mathbf{u}] = P_{n+2}$, there exists a unique vector basis $\mathcal{B}_u = \{\mathbf{v}_1, \ldots, \mathbf{v}_{n+1}\}$ of V such that $P_i = [\mathbf{v}_i]$ and $\mathbf{v}_1 + \ldots + \mathbf{v}_{n+1} = \mathbf{u}$. Now if $P = [\mathbf{v}]$ and $(x_1, \ldots, x_{n+1})$ are the coordinates of $\mathbf{v}$ with respect to the basis $\mathcal{B}_u$, then we take $[x_1, \ldots, x_{n+1}]$ to be the *homogeneous coordinates* of P with respect to $\mathcal{R}$. It is easy to verify that these coordinates are determined by P and $\mathcal{R}$ up to a non-zero multiplicative coefficient.

Note that fixing a projective reference frame on $\mathbb{P}(V)$ is the same as fixing a projective isomorphism between $\mathbb{P}(V)$ and $\mathbb{P}^n$. More generally:

Theorem A.1.2. *If $\mathbb{P}(V)$ and $\mathbb{P}(W)$ are two projective spaces of dimension n, then given two projective reference frames $P_1, \ldots, P_{n+2}$ and $Q_1, \ldots, Q_{n+2}$ on $\mathbb{P}(V)$ and $\mathbb{P}(W)$ respectively, there exists a* unique *projective isomorphism $\varphi : \mathbb{P}(V) \to \mathbb{P}(W)$ such that $\varphi(P_i) = Q_i$ for $i = 1, \ldots, n+2$.*

Analytic representation of projective maps. Let us consider two projective spaces $\mathbb{P}(V_1)$ and $\mathbb{P}(V_2)$ of dimension n and m, equipped with fixed projective reference frames $\mathcal{R}_1$, $\mathcal{R}_2$ respectively. We also let $\mathcal{B}_1, \mathcal{B}_2$ be two normalized vector bases on V_1, V_2 associated with the reference frames $\mathcal{R}_1$, $\mathcal{R}_2$.

If $f : \mathbb{P}(V_1) \setminus K \to \mathbb{P}(V_2)$ is a projective map induced by a linear map $\varphi : V_1 \to V_2$, with $K = \mathbb{P}(\operatorname{Ker} \varphi)$, then we associate with f an $(m+1) \times (n+1)$ matrix $\mathbf{A}$ describing φ with respect to the bases $\mathcal{B}_1, \mathcal{B}_2$. The matrix $\mathbf{A}$ is determined by f up to a non-zero multiplicative coefficient. If a point P in $\mathbb{P}(V_1) \setminus K$ has homogeneous coordinates $P = [\mathbf{v}]$ with $\mathbf{v}$ in $\mathbb{K}^{n+1}$, then $f(P)$ will have homogeneous coordinates $[\mathbf{Av}]$. In the case $V_1 = V_2$, invertible square matrices represent *changes of projective coordinates*.

Affine Geometry. For many concrete applications of projective geometry, including computer vision, it is often convenient to view projective geometry as an extension of affine geometry. Inside $\mathbb{P}^n$, we fix $H_\infty = \{x_{n+1} = 0\}$ as the "hyperplane at infinity". The complement $U = \mathbb{P}^n \setminus H_\infty$ is the set of "finite points". There is a natural bijection $j : \mathbb{K}^n \to U$ given by

$$j : (u_1, \ldots, u_n) \mapsto [u_1, \ldots, u_n, 1]$$

$$j^{-1} : [x_1, \ldots, x_{n+1}] \mapsto \left(\frac{x_1}{x_{n+1}}, \ldots, \frac{x_n}{x_{n+1}} \right).$$

The map j induces an immersion of $\mathbb{K}^n$ in $\mathbb{P}^n$. We can thus interpret $\mathbb{P}^n$ as an *extension* of $\mathbb{K}^n$ with the hyperplane at infinity H_∞.

The projective transformations f of $\mathbb{P}^n$ such that $f(H_\infty) = H_\infty$ correspond to matrices of the form.

$$\mathbf{A} = \left[\begin{array}{c|c} \mathbf{B} & t \\ \hline \mathbf{0} & 1 \end{array} \right].$$

With the identification j introduced above, these describe classical affine affine maps.

More generally, an affine structure can be defined for any set $U_H = \mathbb{P}^n \setminus H$ where H is an arbitrary hyperplane of $\mathbb{P}^n$. The set U_H is called an *affine chart*.

Duality. Let V be a vector space and let V^* be its *dual*, *i.e.,* the vector space of all linear functionals $\psi : V \to \mathbb{K}$. The projective space $\mathbb{P}(V^*)$ is known as the *dual projective space* of $\mathbb{P}(V)$, and is denoted by $\mathbb{P}(V)^*$. The spaces $\mathbb{P}(V)$ and $\mathbb{P}(V)^*$ are projectively isomorphic. If ψ is an element of V^*, the corresponding projective point $[\psi]$ in $\mathbb{P}(V)^*$ uniquely represents the hyperplane $\{[\mathbf{v}] \,|\, \psi(\mathbf{v}) = 0\}$ in $\mathbb{P}(V)$. Hence, points in $\mathbb{P}(V)^*$ can be naturally identified with *hyperplanes* in $\mathbb{P}(V)$.

More concretely, let us consider the standard projective space $\mathbb{P}^n$, so $V = \mathbb{K}^{n+1}$. In this case, it customary to use the usual scalar product to identify $\mathbb{K}$ and $\mathbb{K}^*$. With this identification, a hyperplane $\{a_1x_1 + \ldots + a_{n+1}x_{n+1} = 0\}$ in $\mathbb{P}^n$ is the "dual point" $[a_1, \ldots, a_{n+1}]$ in $(\mathbb{P}^n)^*$. Equivalently, the hyperplane $\{a_1x_1 + \ldots + a_{n+1}x_{n+1} = 0\}$ is represented by $[a_1, \ldots, a_{n+1}]$ using "dual homogeneous coordinates" on $(\mathbb{P}^n)^*$: these are induced by the normalized basis on $(\mathbb{K}^{n+1})^*$ that is dual to the standard vector basis on $\mathbb{K}^{n+1}$.

If W is a vector subspace of $\mathbb{K}^n$, we write $W^\perp$ for the orthogonal complement of W in $(\mathbb{K}^{n+1})^*$:

$$W^\perp = \{(a_1, \ldots, a_n) \in (\mathbb{K}^n)^* \,|\, a_1x_1 + \ldots + a_{n+1}x_{n+1} = 0,\ \forall (x_1, \ldots, x_n) \in W\} \subset (\mathbb{K}^{n+1})^*.$$

If $S = \mathbb{P}(W)$ is a projective subspace of $\mathbb{P}^n$, then we define the associated *dual space* $S^\vee = \mathbb{P}(W^\perp)$ in $(\mathbb{P}^n)^*$ as the projectivization of the orthogonal complement $W^\perp$. This represents the set of hyperplanes which contain all points in S. If S has dimension k in $\mathbb{P}^n$, then $S^\vee$ has dimension $n - k - 1$ in $(\mathbb{P}^n)^*$. The association $S \mapsto S^\vee$ is known as the *duality correspondence*, and it satisfies $S^{\vee\vee} = S$.[1] Moreover, duality reverses containments and the operations of join and meet:

$$(S_1 \wedge S_2)^\vee = S_1^\vee \vee S_2^\vee, \qquad (S_1 \vee S_2)^\vee = S_1^\vee \wedge S_2^\vee.$$

A.2 Grassmannians and Line Geometry

In this section we introduce Grassmannians and Plücker coordinates, with particular focus on the Grassmannian of lines $\mathrm{Gr}(1, \mathbb{P}^3)$. We should mention that a popular formalism for expressing geometric properties in projecive spaces is the *Grassmann-Cayley algebra* (see for example [174]). We will not introduce this theory here, since it is not directly used in the thesis. We define Plücker coordinates for arbitrary Grassmannians, but describe the operations of "join" and "meet" only for points, lines, and planes in $\mathbb{P}^3$.

[1] Note that if $S = \mathbb{P}(W)$, then $S^\vee$ is *not* the same as $\mathbb{P}(W^*)$. For example, $(\mathbb{P}^n)^\vee = \emptyset \neq (\mathbb{P}^n)^*$.

A.2.1 Grassmannians and Plücker coordinates

Let $\mathbb{P}^n = \mathbb{P}(\mathbb{K}^{n+1})$ denote the n dimensional projective space over a field $\mathbb{K}$ inside $\mathbb{C}$. For any $0 \leq k \leq n$, the *Grassmann variety* (or *Grassmannian*) $\mathrm{Gr}(k, \mathbb{P}^n)$ is the set of all projective subspaces of $\mathbb{P}^n$ that have dimension k. It is possible to equip any Grassmannian with a set of homogeneous coordinates as follows.

Let $H = \mathbb{P}(W)$ be a projective subspace of $\mathbb{P}^n$ of dimension k, spanned by the independent points $P_1 = [\mathbf{w}_1], \ldots, P_{k+1} = [\mathbf{w}_{k+1}]$. If $\mathbf{W}$ is the $(n+1) \times (k+1)$ matrix $\mathbf{W}$ whose columns are the vectors $\mathbf{w}_1, \ldots, \mathbf{w}_{k+1}$, we associate with the subspace H the point $[\mathbf{p}_H]$ in $\mathbb{P}^{\binom{n+1}{k+1}-1}$ whose homogeneous coordinates are the maximal minors of $\mathbf{M}$, ordered lexicographically. It is customary to write $p_{i_1 \ldots, i_{k+1}}$ (with $0 \leq i_1 < \ldots < i_{k+1} \leq n$) for the coordinate corresponding to the determinant of the $(k+1) \times (k+1)$ matrix defined by the columns $i_1, \ldots, i_{k+1}$ of $\mathbf{W}$. It easy to see that the point $[\mathbf{p}_H]$ in $\mathbb{P}^{\binom{n+1}{k+1}-1}$ is independent of the choice of $P_1, \ldots, P_{k+1}$ in H. The association $\mathrm{Gr}(k, \mathbb{P}^n) \to \mathbb{P}^{\binom{n+1}{k+1}-1}$ is known as the *Plücker embedding*, and the homogeneous coordinates in $\mathbb{P}^{\binom{n+1}{k+1}-1}$ of a subspace H are known as its *Plücker coordinates*.

It is also possible to represent H using *dual Plücker coordinates*. If H is the intersection of $n-k$ hyperplanes $H_1, \ldots, H_{n-k}$ represented by points $[\mathbf{u}_1], \ldots, [\mathbf{u}_{n-k}]$ in $(\mathbb{P}^n)^*$, we consider the $(n-k) \times (n+1)$ matrix $\mathbf{U}$ whose rows are $\mathbf{u}_1, \ldots, \mathbf{u}_{k+1}$. We then associate with H the point $[\mathbf{q}_H]$ in $\mathbb{P}^{\binom{n+1}{n-k}-1} = \mathbb{P}^{\binom{n+1}{k+1}-1}$ whose coordinates are the maximal minors of $\mathbf{U}$, ordered lexicographically. Note that this point also represents the primal Plücker coordinates of the *dual space* $H^\vee$ of H inside $\mathrm{Gr}(n-k-1, \mathbb{P}^n)$. For dual Plücker coordinates it is customary to write $q_{j_1 \ldots, j_{n-k}}$ (with $0 \leq j_1 < \ldots < j_{n-k} \leq n$) for the coordinate corresponding to the determinant of the matrix defined by the columns $j_1, \ldots, j_{n-k}$ of $\mathbf{U}$. Primal and dual plucker coordinates are related by

$$q_{j_1 \ldots, j_{n-k}} = (-1)^{\sigma(i_1, \ldots, i_{k+1})} p_{i_1 \ldots, i_{k+1}}, \tag{A.1}$$

where $j_1, \ldots, j_{n-k}$ are complementary indices of $i_1, \ldots, i_{k+1}$ (ordered appropriately), and $\sigma(i_1, \ldots, i_{k+1})$ is the sign of the permutation $(i_1, \ldots, i_{k+1}, j_1, \ldots, j_{n-k})$.

Finally, we mention that it is sometimes convenient to allow for indices of (primal or dual) Plücker to be any sequence of distinct numbers $i_1, \ldots, i_{k+1}$ or $j_1, \ldots, j_{n-k}$ between 1 and n. In this case, the transposition of two indices is simply a change of sign of $p_{i_1, \ldots, i_{k+1}}$ or $q_{j_1, \ldots, j_{n-k}}$.

Plücker relations. The image of Plücker embedding of $\mathrm{Gr}(k, \mathbb{P}^n)$ is a subvariety of $\mathbb{P}^{\binom{n+1}{k+1}-1}$. Indeed, Plücker coordinates always satisfy certain quadratic relations, known as *Plücker relations*. These are completely characterized by the following result (see for example [174] for a proof).

Theorem A.2.1. *The Plücker coordinates of any subspace in* $\mathrm{Gr}(k, \mathbb{P}^n)$ *always satisfy quadratic relations of the form*

$$\sum_{a=1}^{k+2} p_{i_1,\ldots,i_k,l_a} p_{l_1,\ldots,\hat{l_a},\ldots,l_{k+2}} = 0, \tag{A.2}$$

where $1 \leq i_1 < \ldots < i_k \leq n$ *and* $1 \leq l_1 < \ldots < l_{k+2} \leq n$*, and the symbol* $\hat{l_a}$ *indicates that the index* l_a *is missing.*

Conversely, any vector $\mathbf{p} = (p_{i_1,\ldots,i_{k+1}})$ *satisfying (A.2) represents the Plücker coordinates of a subspace in* $\mathrm{Gr}(k, \mathbb{P}^n)$*. In fact, all such relations generate the ideal associated with (the image of)* $\mathrm{Gr}(k, \mathbb{P}^n)$ *inside* $\mathbb{P}^{\binom{n+1}{k+1}-1}$*.*

A.2.2 Lines in projective space

In this thesis we are mostly interested in the Grassmannian $\mathrm{Gr}(1, \mathbb{P}^3)$ of lines in $\mathbb{P}^3$. According to the general discussion above, this set can be embedded in $\mathbb{P}^5$. In particular, the line ℓ through the points $x = [x_1, x_2, x_3, x_4]$ and $y = [y_1, y_2, y_3, y_4]$ can be identified with $[p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}]$ where $p_{ij} = x_i y_j - x_j y_i$. These are the (primal) Plücker coordinates of ℓ. The Plücker relations reduce in this case to a single quadratic constraint, namely

$$p_{12}p_{34} - p_{13}p_{24} + p_{14}p_{23} = 0. \tag{A.3}$$

Hence, $\mathrm{Gr}(1, \mathbb{P}^3)$ can be viewed as a quadric hypersurface in $\mathbb{P}^5$.

Similarly, if a line ℓ is the intersection of two planes represented as $a = [a_1, a_2, a_3, a_4]$ and $b = [b_1, b_2, b_3, b_4]$ in $(\mathbb{P}^3)^*$ (which means that, for example, a represents the plane $\{a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 = 0\}$ in $\mathbb{P}^3$), then the dual Plücker coordinates of ℓ are $[q_{12}, q_{13}, q_{14}, q_{23}, q_{24}, q_{34}]$ where $q_{ij} = a_i b_j - a_j b_i$. As explained above, primal and dual coordinates satisfy $p_{ij} = \sigma(ijkl) q_{kl}$, where i, j, k, l are distinct indices and $\sigma(ijkl)$ is the sign of the permutation $(ijkl)$.

Note that if we identify $\mathbb{P}^3$ and $(\mathbb{P}^3)^*$ using the usual dot product, then the dual $\ell^\vee$ of a line ℓ with Plücker coordinates $[\mathbf{p}] = [p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}]$ will have coordinates $[\mathbf{p}^*] = [p_{34}, -p_{24}, p_{23}, p_{14}, -p_{13}, p_{12}]$. In other words, the (primal) Plücker coordinates of $\ell^\vee$ are the dual Plücker coordinates of ℓ.

Remark A.2.2. Throughout the thesis, we identify the Grassmannian $\mathrm{Gr}(1, \mathbb{P}^3)$ with its embedding in $\mathbb{P}^5$. Thus, we sometimes abuse notation and write $\ell = [\mathbf{p}]$, where ℓ is a line $\mathbb{P}^3$ and $[\mathbf{p}]$ is a point in $\mathbb{P}^5$.

Plücker matrices and incidence conditions. To express incidences of lines with points and planes, it is convenient to write the Plücker coordinates of a line $[\mathbf{p}] =$

$[p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}]$ and its dual as the entries of two skew-symmetric 4×4-matrices:

$$[\mathbf{P}] = \begin{bmatrix} 0 & p_{34} & -p_{24} & p_{23} \\ -p_{34} & 0 & p_{14} & -p_{13} \\ p_{24} & -p_{14} & 0 & p_{12} \\ -p_{23} & p_{13} & -p_{12} & 0 \end{bmatrix} \quad \text{and} \quad [\mathbf{P}^*] = \begin{bmatrix} 0 & p_{12} & p_{13} & p_{14} \\ -p_{12} & 0 & p_{23} & p_{24} \\ -p_{13} & -p_{23} & 0 & p_{34} \\ -p_{14} & -p_{24} & -p_{34} & 0 \end{bmatrix}. \tag{A.4}$$

If $\mathbf{x}$ and $\mathbf{y}$ are column vectors representing points on the line, then our definition for the associated matrix $\mathbf{P}^*$ is $\mathbf{x}\mathbf{y}^T - \mathbf{y}\mathbf{x}^T$. Similarly, if $\mathbf{u}$ and $\mathbf{v}$ are column vectors representing planes containing the line, then the matrix $\mathbf{P}$ is $\mathbf{u}\mathbf{v}^T - \mathbf{v}\mathbf{u}^T$. The conditions $\operatorname{rank}(\mathbf{P}) = 2$, $\operatorname{rank}(\mathbf{P}^*) = 2$, and $\operatorname{trace}(\mathbf{P}\mathbf{P}^*) = 0$ are all equivalent to the Plücker relation (A.3).

Two lines $[\mathbf{p}] = [p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}]$ and $[\mathbf{q}] = [q_{12}, q_{13}, q_{14}, q_{23}, q_{24}, q_{34}]$ are *concurrent* at a point in $\mathbb{P}^3$ if and only if

$$\mathbf{p}^* \cdot \mathbf{q} = \mathbf{q}^* \cdot \mathbf{p} = p_{12}q_{34} - p_{13}q_{24} + p_{14}q_{23} + p_{23}q_{14} - p_{24}q_{13} + p_{34}q_{12} = 0. \tag{A.5}$$

If $\mathbf{P}, \mathbf{P}^*, \mathbf{Q}, \mathbf{Q}^*$ are the matrices associated with $[\mathbf{p}]$ and $[\mathbf{q}]$ as above, then (A.5) is equivalent to

$$\operatorname{trace}(\mathbf{P}\mathbf{Q}^*) \;=\; \operatorname{trace}(\mathbf{P}^*\mathbf{Q}) \;= 0 \tag{A.6}$$

Note that the concurrence condition is *bilinear* in Plücker coordinates. In particular, all lines that intersect a fixed line $[\mathbf{p}]$ form a threefold in $\mathbb{P}^5$, obtained by intersecting $\operatorname{Gr}(1, \mathbb{P}^3)$ with a hyperplane.

If $x = [\mathbf{x}]$ is a point and $\ell = [\mathbf{p}]$ is a line in $\mathbb{P}^3$, then the join $x \vee \ell$ is the plane with dual homogeneous coordinates $[\mathbf{u}]$ given by $\mathbf{u} = \mathbf{P}\mathbf{x}$. This plane is not defined if and only if x belongs to ℓ, in which case

$$\mathbf{P}\mathbf{x} \;=\; \mathbf{0}. \tag{A.7}$$

For a fixed point $x = [\mathbf{x}]$, this yields three independent linear equations in the entries of $\mathbf{P}$. These equations define a plane in $\mathbb{P}^5$ that is actually completely contained in the Grassmannian $\operatorname{Gr}(1, \mathbb{P}^3)$. This is known as the *$\alpha$-plane* of x.

Similarly, if $u = [\mathbf{u}]$ is a dual point representing a plane in $\mathbb{P}^3$ and $\ell = [\mathbf{p}]$ is a line, then the meet $u \wedge \ell$ is the point $[\mathbf{x}]$ with $\mathbf{x} = \mathbf{P}^*\mathbf{u}$. This point is not defined if and only if u contains ℓ, that is if

$$\mathbf{P}^*\mathbf{u} \;=\; \mathbf{0}. \tag{A.8}$$

For a fixed $u = [\mathbf{u}]$, this defines a plane in $\mathbb{P}^5$ contained in $\operatorname{Gr}(1, \mathbb{P}^3)$. This is known as the *$\beta$-plane* of u.

Remark A.2.3. The families of α and β-planes form two disjoint *rulings* on the Plücker quadric. Two different planes in the same family (α or β) always intersect in exactly one point in $\operatorname{Gr}(1, \mathbb{P}^3)$: this reflects the fact that in $\mathbb{P}^3$ there is exactly one line through two generic points, and exactly one line contained in two generic planes. On the other hand, the α-plane of a point x and the β-plane of a plane u do not meet unless x lies on u. In other words, there is no line contained in a given plane and passing through a given point, unless the point lies on the plane.

Appendix B

Projective Varieties

This appendix is a very short (and incomplete) introduction to the algebraic geometry of projective varieties. Introductory courses on algebraic geometry usually first focus on affine varieties, however in this thesis we are only interested in varieties in projective space. We found that a useful reference for our purposes are the lecture notes by Arrondo [6], which present many results on projective varieties in a very concrete fashion. Another excellent and quite accessible reference is the book by Beltrametti, Carletti, Gallarati, and Monti Brigadin [13], that introduces algebraic geometry from a rigorous but very classical perspective ("classical" here is opposed to the abstract approach of modern algebraic geometry, usually based on the language of *schemes*). Another well-known and slightly more advanced introductory textbook is by Harris [73].

B.1 Ideals and Varieties

Classical algebraic geometry works over the field $\mathbb{C}$ of complex numbers. Throughout this chapter we let $\mathbb{P}^n = \mathbb{P}(\mathbb{C}^{n+1})$. We also write $\mathbb{C}[x_1, \ldots, x_{n+1}]$ for the ring of polynomials in $x_1, \ldots, x_{n+1}$ with complex coefficients.

We recall that a *homogeneous polynomial* F in $\mathbb{C}[x_1, \ldots, x_{n+1}]$ is a polynomial that has all terms of the same degree d. In that case, we have $F(\lambda a_1, \ldots, \lambda a_{n+1}) = \lambda^d F(a_0, \ldots, a_{n+1})$ for any $a_1, \ldots, a_{n+1}, \lambda$ in $\mathbb{C}$. Any polynomial F of degree d can be decomposed as $F = F_0 + \ldots + F_d$ where F_k is homogeneous of degree k. In the following, if $a = [\mathbf{a}]$ is a projective point in $\mathbb{P}^n$, and F is a homogeneous polynomial $\mathbb{C}[x_1, \ldots, x_{n+1}]$, we write $F(a) = 0$ if F evaluated at $\mathbf{a} = (a_1, \ldots, a_{n+1})$ is zero. This is well-defined because F is homogeneous. More generally, even if F is not homogeneous, we write $F(a) = 0$ to say that $F_k(a) = 0$ for all F_k in the homogeneous decomposition $F = F_0 + \ldots + F_d$.

Definition B.1.1. A *projective variety* (or simply *variety*) $X \subset \mathbb{P}^n$ is the zero-set of a family of *homogeneous* polynomials. More precisely, X is a projective variety if there exists a set $S \subset \mathbb{C}[x_1, \ldots, x_{n+1}]$ of homogeneous polynomials such that $X = V(S)$ where

$$V(S) = \{a \in \mathbb{P}^n \mid F(a) = 0, \ \forall F \in S\}. \tag{B.1}$$

When $X = V(S)$, we say that X is the variety *defined* by S.[1]

A *(polynomial) ideal* is a subset I of $\mathbb{C}[x_1, \ldots, x_{n+1}]$ that forms an additive group ($F_1, F_2 \in I$ implies $F_1 + F_2 \in I$) and that is closed under multiplication with any other polynomial ($F \in I$ then $GF \in I$ for all $G \in \mathbb{C}[x_1, \ldots, x_{n+1}]$). One can show that all polynomials ideals are finitely generated, *i.e.*, there always exist polynomials $F_1, \ldots, F_r$ such that

$$I = \{A_1T_1 + \ldots + A_rT_r \mid A_i \in \mathbb{C}[x_1, \ldots, x_{n+1}]\}.$$

In this case, we say that the polynomials $F_1, \ldots, F_k$ are *generators* of I, and we write $I = \langle F_1, \ldots, F_r \rangle$. The ideal I is a *homogeneous ideal* if $I = \langle F_1, \ldots, F_r \rangle$ where each F_i is a homogeneous polynomial.

Definition B.1.2. If $X \subset \mathbb{P}^n$ is a variety, we define

$$I(X) = \{F \in \mathbb{C}[x_1, \ldots, x_{n+1}] \mid F(a) = 0, \ \forall A \in X\}.$$

This set is always a homogeneous ideal in $\mathbb{C}[x_1, \ldots, x_{n+1}]$.

According to Definitions B.1.1 and B.1.2, a variety X has an associated ideal $I(X)$, and an ideal I has an associated variety $V(I)$. This association between varieties and polynomial ideals is the first brick in the foundation of algebraic geometry. The following are some basic properties of this correspondence.

i) If S is a general set of polynomials in $\mathbb{C}[x_1, \ldots, x_{n+1}]$, then $V(S) = V(I)$ where I is the smallest ideal containing S.

ii) The operators V and I reverse containments: if $S \subset S'$ in $\mathbb{C}[x_1, \ldots, x_{n+1}]$ then $V(S) \supset V(S')$ in $\mathbb{P}^n$, and if $X \subset X'$ in $\mathbb{P}^n$ then $I(X) \supset I(X')$ in $\mathbb{C}[x_1, \ldots, x_{n+1}]$.

iii) If X is a variety, then $V(I(X)) = X$.

iv) If I is a homogeneous ideal, then $I(V(I)) \supset I$.

The last point can be mode precise. We define the *radical* $\sqrt{I}$ of an ideal I as

$$\sqrt{I} = \{F \mid \exists\ m \geq 1 \ s.t.\ F^m \in I\}.$$

The radical $\sqrt{I}$ is an ideal that contains I. We say that I is a *radical ideal* if $I = \sqrt{I}$. The following important result, known as the *(projective) Nullstellensatz*, fully explains

[1] For some authors this is actually definition of an affine *projective set*, since they require projective varieties to be also irreducible (see Section B.2).

the correspondence of ideals and projective sets, showing that varieties are uniquely associated with radical ideals in $\mathbb{C}[x_1, \ldots, x_{n+1}]$. The proof can be found for example in [6].

Theorem B.1.3 (Projective Nullstellensatz)**.** *If I is contained in $\mathfrak{M} = \langle x_1, \ldots, x_{n+1} \rangle$, then $I(V(I)) = \langle 1 \rangle$. Otherwise $I(V(I)) = \sqrt{I}$.*

The ideal $\mathfrak{M} = \langle x_1, \ldots, x_{n+1} \rangle$ is called the "irrelevant ideal", and it plays a special role because it is the largest polynomial set such that $V(\mathfrak{M}) = \emptyset$ (any projective point has at least one non-zero coordinate). For this ideal we have $I(V(\mathfrak{M})) = \mathbb{C}[x_1, \ldots, x_{n+1}]$, even though $\mathfrak{M}$ is radical. Excluding the irrelevant ideal, there is a one-to-one correspondence between projective varieties in $\mathbb{P}^n$ and homogeneous radical ideals in $\mathbb{C}[x_1, \ldots, x_{n+1}]$:

projective varieties		homogeneous radical ideals
X	$\xrightarrow{I}$	$I(X)$
$V(I)$	$\xleftarrow{V}$	I

Finally, we note that varieties can be viewed as closed sets of a a topology on $\mathbb{P}^n$. This is known as the *Zariski topology*. In the context of varieties, we always consider this topology. For example, the closure of any set $Y \subset \mathbb{P}^n$ is the smallest variety containing Y.

B.2 Irreducible Components

A projective variety X in $\mathbb{P}^n$ is said to be *irreducible* if it is not the union $X = Z_1 \cup Z_2$ of two proper algebraic subsets $Z_1 \subsetneq X$ and $Z_2 \subsetneq X$. It is clear that inside an irreducible variety X, (Zariski) open sets $U \subset X$ are always *dense* (*i.e.*, their closure is the whole set $\overline{U} = X$).

If X is an irreducible projective variety, the corresponding ideal $I(X)$ is a *prime ideal*. Algebraically, an ideal I is prime if and only if it satisfies

$$FG \in I \Rightarrow F \in I \text{ or } G \in I.$$

If an ideal I is prime it is always radical ($I = \sqrt{I}$).

A projective variety X in $\mathbb{P}^n$ can always be decomposed as the union of a finite number of *irreducible components*

$$X = Z_1 \cup \ldots \cup Z_k, \tag{B.2}$$

and moreover this decomposition is unique up to order. The decomposition of X has an associated decomposition of ideals (notice the union now becomes an intersection):

$$I(X) = I(Z_1) \cap \ldots \cap I(Z_k). \tag{B.3}$$

In this decomposition $I(X)$ is radical, and $I(Z_1), \ldots, I(Z_k)$ are prime. We note that a non-radical ideal I cannot be expressed as an intersection of primes, but it admits a more general decomposition based on *primary ideals*, which is not completely unique. In this thesis we mostly deal with radical ideals associated with varieties, and that can be decomposed into primes. However, we briefly recall for completeness the more general notion of *primary decomposition*.

An ideal I is said to *primary* if $FG \in I$ and $G \notin I$ implies that there exists $m \geq 1$ such that $F^m \in I$. It is easy to see that if I is primary then $\sqrt{I}$ is prime. The following result now holds.

Theorem B.2.1. *Any ideal I can be decomposed as*

$$I = I_1 \cap \ldots \cap I_r$$

where each I_j is primary, $\cap_{j \neq k} I_j \not\subset I_k$ holds, and the radicals $\sqrt{I_1}, \ldots, \sqrt{I_r}$ are distinct prime ideals.

Moreover, the prime ideals $\sqrt{I_1}, \ldots, \sqrt{I_r}$ are uniquely determined by I, and any primary ideal I_j contained a minimal *prime (with respect to the inclusion) is also uniquely determined. However primary ideals associated with non-minimal prime ideals are in general not unique.*

Finally, we define here the operation of *ideal saturation*, which is used many times in the computations described in this thesis. The saturation of an ideal I with respect to another ideal J is the ideal $I : J^\infty$ defined as

$$I : J^\infty = \{F \mid \forall G \in J \; \exists m \geq 0 \; s.t. \; FG^m \in I\}.$$

This ideal has the property that

$$V(I : J^\infty) = \overline{V(I) \setminus V(J)},$$

where the bar denotes the (Zariski) closure (see for example [39]). In other words, saturating I with respect to J excludes from $V(I)$ all the irreducible components of $V(I)$ that are contained in $V(J)$. A particularly important case is when $J = \mathfrak{M} = \langle x_1, \ldots, x_{n+1} \rangle$ is the irrelevant ideal.

B.3 Dimension and Degree

Let X be a variety in $\mathbb{P}^n$. The *dimension* of X can be defined as the maximum integer $r = \dim X$ such that any linear subspace of $\mathbb{P}^n$ of codimension at most r meets X. Another equivalent definition is the maximum length r of a strictly increasing chain $Z_0 \subsetneq Z_1 \subsetneq \ldots \subsetneq Z_r$ of irreducible varieties contained in X. Many other characterizations of the dimension exist. When the dimension of X is $n-1$, X is called a *hypersurface*.

The *degree* of a variety X of dimension r is the maximum number $d = \deg X$ of points in the intersection of X with a linear space of codimension r (when this intersection is finite).

In general we expect the intersection of two varieties $X_1 \cap X_2$ in $\mathbb{P}^n$ to have dimension $\dim X_1 + \dim X_2 - n$ and degree $\deg X_1 \cdot \deg X_2$.

Theorem B.3.1 (Bezout's theorem)**.** *If X_1, X_2 are varieties in $\mathbb{P}^n$ that intersect* generically transversely*, then*

$$\deg(X_1 \cap X_2) = \deg X_1 \cdot \deg X_2.$$

In particular, if $\dim X_1 + \dim X_2 = n$, then $X_1 \cap X_2$ consists of $\dim X_1 \cdot \dim X_2$ points.

Here the intersection between X_1 and X_2 is said to be generically transverse if each of the components of $X_1 \cap X_2$ contains a point where the intersection is transverse (we refer to Section B.6 for a discussion of transversality).

B.4 Products of Projective Spaces

Many important algebraic varieties studied in this thesis are subsets of *products of projective spaces*. For dealing with these sets, it is convenient to introduce the *Segre map*. For a product $\mathbb{P}^{n_1} \times \mathbb{P}^{n_2}$, the Segre map σ is defined as

$$\sigma : \mathbb{P}^n \times \mathbb{P}^m \to \mathbb{P}^{(n+1)(m+1)-1}, \quad ([x_1, \ldots, x_{n+1}], [y_1, \ldots, y_{m+1}]) \mapsto [\ldots, x_i y_j, \ldots]. \quad \text{(B.4)}$$

In other words, the Segre map associates a pair $([\mathbf{x}], [\mathbf{y}])$ to a point in $\mathbb{P}^{(n+1)(m+1)-1}$ whose coordinates are the pairwise products of the coordinates of $[\mathbf{x}]$ and $[\mathbf{y}]$. This definition can be iterated for more general products $\mathbb{P}^{n_1} \times \ldots \times \mathbb{P}^{n_k}$ (and everything that follows can be extended to this case: here we restrict ourselves to the product of two projective spaces only for notational simplicity). The image of the Segre map $\sigma(\mathbb{P}^n \times \mathbb{P}^m)$ inside $\mathbb{P}^{(n+1)(m+1)-1}$ is known as the *Segre variety*.

A variety X in $\mathbb{P}^n \times \mathbb{P}^m$ is by definition a set such that $\sigma(X)$ is a variety in $\mathbb{P}^{(n+1)(m+1)-1}$. The following result characterizes varieties in $\mathbb{P}^n \times \mathbb{P}^m$ more directly. We recall that a polynomial $F(x_1, \ldots, x_{n+1}, y_1, \ldots, y_{m+1})$ is *bihomogeneous* if it is independently homogeneous in the variables x_i and in the variables y_i (in general, with different degrees).

Proposition B.4.1. *A set X in $\mathbb{P}^n \times \mathbb{P}^m$ is a variety if and only if it can be characterized as the zero-set of a family bihomogeneous polynomials.*

The dimension of a variety X in $\mathbb{P}^n \times \mathbb{P}^m$ is the dimension $\sigma(X)$ in $\mathbb{P}^{(n+1)(m+1)-1}$. The notion of "degree" is however more subtle, and is best expressed using a *multidegree polynomial*. If X has dimension r and codimension $p = n + m - r$, then the multidegree is a homogeneous polynomial in $\mathbb{Z}[T_1, T_2]$ of degree p

$$c_{p,0} T_1^p + c_{p-1,1} T_1^{p-1} T_2 + \ldots c_{1,p-1} T_1 T_2^{p-1} + c_{0,p} T_2^p$$

where the coefficient $c_{d,p-d}$ of the monomial $T_1^d T_2^{p-d}$ is the number of intersections of X with a product $L_1 \times L_2$ of general linear spaces L_1 in $\mathbb{P}^n$ and L_2 in $\mathbb{P}^m$ of dimensions d and $p-d$ respectively.

B.5 Morphisms and Rational Maps

We now introduce two types of maps between algebraic varieties. Roughly speaking, *morphisms* (or *regular maps*) are polynomial maps that can be evaluated at every point of the domain, while *rational maps* are polynomial maps that can only be evaluated on an open dense set of the domain. The exact definition requires some extra work because these maps may not have a single polynomial representation, but may need to be defined by "stitching" together different local representations.

More precisely, let $X \subset \mathbb{P}^n$ be a variety. A *morphism* $F : X \to \mathbb{P}^m$ is a map such that for every x in X, there exists an open neighborhood $x \in U \subset X$ and homogeneous polynomials $f_1, \ldots, f_{m+1}$ of the same degree so that $F(p) = [f_1(p), \ldots, f_{m+1}(p)]$ for all p in U. Note that the polynomials $f_1, \ldots, f_{m+1}$ cannot simultaneously vanish on U. The image under a morphism of a projective (irreducible) set is a projective (irreducible) irreducible set.

Example B.5.1. Let $X = V(x_1^2 + x_2^2 - x_3^3)$ in $\mathbb{P}^2$. The map

$$\varphi_1 : [x_1, x_2, x_3] \mapsto [x_1, x_2 - x_3]$$

is not defined at the point $p = [0, 1, 1]$ on X. However, φ_1 can be extended to a morphism $F : X \to \mathbb{P}^1$ by setting $F(x) = \varphi_1(x)$ for $x \neq p$ and $F(p) = [1, 0]$. Indeed, if we consider

$$\varphi_2 : [x_1, x_2, x_3] \mapsto [x_2 + x_3, -x_1]$$

then $\varphi_1(x) = \varphi_2(x)$ for all $x \in X$ excluding the points p and $q = [0, -1, 1]$, and $\varphi_2(p) = [1, 0]$. Hence, the morphism F is defined by stitching together the polynomial representations φ_1 and φ_2. ♢

A *rational map* F from an algebraic variety X to $\mathbb{P}^m$ is only defined on a open dense set W of X. It is usually written as $F : X \dashrightarrow \mathbb{P}^m$. The definition is analogous to that of a morphism but we restricted to points in W: for every x in W, there must exist an open neighborhood $x \in U \subset W$ and homogeneous polynomials $f_1, \ldots, f_{m+1}$ of the same degree so that $F(p) = [f_1(p), \ldots, f_{m+1}(p)]$ for all p in W. When $X = \mathbb{P}^n$, one can show that every rational map $F : \mathbb{P}^n \dashrightarrow \mathbb{P}^m$ admits a global representation, *i.e.*, we can write $F(p) = [f_1(p), \ldots, f_{m+1}(p)]$ whenever F is defined. An important example for this thesis is when $F : \mathbb{P}^n \dashrightarrow \mathbb{P}^m$ is a *linear projection* (see Chapter 1). A rational map $F : X \dashrightarrow Y$ is *dominant* if the closure of the image of F is Y. Two rational maps $F : X \dashrightarrow Y$ and $G : Y \dashrightarrow Z$ can be composed only if F is dominant.

Two projective varieties X in $\mathbb{P}^n$ and Y in $\mathbb{P}^m$ are said to be *isomorphic* when there exist morphisms $F : X \to Y$ and $G : Y \to X$ that are inverses of each others. If X and Y both live in $\mathbb{P}^n$, then X and Y are isomorphic if they are in particular *projectively equivalent*, *i.e.*, if there exists a projective transformation $T : \mathbb{P}^n \to \mathbb{P}^n$ such that $T(X) = Y$. The converse however is not true, and two isomorphic varieties need not be projectively equivalent: for example, a line and a conic in $\mathbb{P}^2$ are isomorphic (see Example B.5.1).

Two projective varieties X in $\mathbb{P}^n$ and Y in $\mathbb{P}^m$ are said to be *birational* or *birationally equivalent* when there exist dominant rational maps $F : X \dashrightarrow Y$ and $G : Y \dashrightarrow X$ that are inverses of each other. This is clearly a weaker condition than being isomorphic (or projectively equivalent). A variety X that is birationally equivalent to $\mathbb{P}^n$ for some n is said to be *rational*.

B.6 Tangent Spaces

Let X be a variety in $\mathbb{P}^n$ with $I(X) = \langle F_1, \ldots, F_r \rangle$. If p is a point on X, the *embedded tangent space* of X at p is the linear space $T_p(X) \subset \mathbb{P}^n$ defined by the equations

$$\frac{\partial F_i}{\partial x_1}(p)x_1 + \ldots \frac{\partial F_i}{\partial x_n}(p)x_n = 0, \quad i = 1, \ldots, r.$$

Geometrically, $T_p(X) \subset \mathbb{P}^n$ coincides with the union of all lines meeting X at p with multiplicity at least two.

If we define *local dimension* of X at p as the maximum dimension of a component of X through p, then one can show that $T_p(X)$ always has dimension at least equal to this local dimension. When the dimension $T_p(X)$ coincides with the local dimension, we say that p is *smooth*, otherwise it is *singular*. More concretely, if the local dimension of X at p is r (or in particular if X is irreducible of dimension r), then p is singular if and only if the *Jacobian matrix*

$$J(p) = \begin{bmatrix} \frac{\partial F_1}{\partial x_1}(p) & \ldots & \frac{\partial F_1}{\partial x_1}(p) \\ \vdots & & \vdots \\ \frac{\partial F_r}{\partial x_1}(p) & \ldots & \frac{\partial F_r}{\partial x_1}(p) \end{bmatrix}$$

has rank strictly smaller than $n - r$. We say that X is smooth if all points p on X are smooth. Two varieties X and Y in $\mathbb{P}^n$ intersect *transversely* at a point p if $T_p(X)$ and $T_p(Y)$ span $\mathbb{P}^n$. We mention in this context the following important result.

Theorem B.6.1 (Bertini's Theorem)**.** *If X is a variety in $\mathbb{P}^n$, then there exists an open subset U in $(\mathbb{P}^n)^*$ such that for any hyperplane H in U the set $X \cap H$ is smooth at points where X is smooth. In particular, if X is smooth, then its intersection with a general hyperplane is smooth.*

B.7 Multiple Points on Hypersurfaces

We now take a closer look at singular points of hypersurfaces in $\mathbb{P}^n$. This is the simplest case to describe, since hypersurfaces in projective space are always defined by a single polynomial.

We let $X = V(F)$ be a hypersurface in $\mathbb{P}^n$, where F is homogeneous polynomial in $\mathbb{C}[x_1, \ldots, x_{n+1}]$ of degree d. We also fix a point p on X, and another point q not necessarily on X (to simplify notation, we use p, q for the corresponding coordinates vectors in $\mathbb{C}^{n+1}$). We now write the Taylor expansion of F at p:

$$F(p + tq) = t\Delta_q F(p) + (t^2/2)\Delta_q^2 F(p) + \ldots + (t^d/d!)\Delta_q^d F(p),$$

where

$$\Delta_q^k F(p) = \sum_{i_1,\ldots,i_k} q_{i_1} \ldots q_{i_k} \left(\frac{\partial^k F}{\partial x_{i_1} \ldots \partial x_{i_k}} \right)(p).$$

The *intersection multiplicity* of the line $L = p \vee q$ with X at p is m if

$$\Delta_q F(p) = \ldots = \Delta_q^{m-1} F(p) = 0 \text{ and } \Delta_q^m F(p) \neq 0.$$

We say that p is an *m-fold point* for X (or has *multiplicity m* for X) if the intersection multiplicity with X is at least m for all lines through p. This is equivalent to the fact that all $(m-1)$-derivatives of F vanish at P (this automatically implies that all lower order derivatives vanish as well). Thus, 1-fold points are smooth points, and m-fold points for $m \geq 2$ are singular.

If p is a point of multiplicity m on X, we define the *tangent cone* $C_p(X) \subset \mathbb{P}^n$ at p as the hypersurface of degree m defined by

$$\Delta_x^m F(p) = \sum_{i_1,\ldots,i_m} x_{i_1} \ldots x_{i_m} \left(\frac{\partial^m F}{\partial x_{i_1} \ldots \partial x_{i_m}} \right)(p) = 0.$$

This represents the union of all lines that intersect X at p with multiplicity higher than m. Thus, the tangent cone generalizes the tangent space when p is a singular point.

Within the tangent cone $C_p(X)$, we can consider special lines that intersect X with ever higher multiplicity. For this, we write $C_p^h(X)$ for the union of all lines that intersect X at p with multiplicity at least $m + h$. This set is defined by

$$\Delta_x^m F(p) = \Delta_x^{m+1} F(p) = \ldots = \Delta_x^{m+h-1} F(p) = 0.$$

In general, $C_p^h(X)$ contains a line only for $h \leq n-1$. For $h = n-1$, we have that $C_p^{n-1}(X)$ contains in general a finite number of lines that have intersection multiplicity exactly $m + n - 1$ at p: these lines are known as the *principal tangents* of X at p. In

particular, if p is a smooth point X, we expect to find $(n-1)!$ principal tangents that have intersection multiplicity n with X.

For example, let us consider a smooth surface X in $\mathbb{P}^3$. For any point p on X, we have that $C_p(X)$ is simply the embedded tangent plane $T_p(X)$ at p, while the set $C_p^2(X)$ generally contains two distinct principal tangents that have intersection multiplicity 3 with X. From the perspective of real geometry, one distinguishes points p where these two tangent are both real (*hyperbolic points*), or complex (*elliptic points*). We also note that exceptional situations occur at *flecnodal* and *parabolic* points: at a flecnodal point, one of the two principal tangents has intersection multiplicity at least 4; at a parabolic point, $\Delta_x^2(p)$ is a degenerate quadric, and the two principal tangents coincide. See Figure B.1

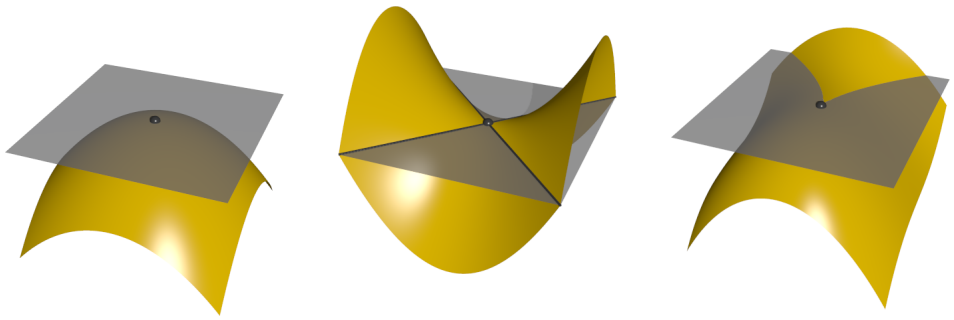


Figure B.1: Local shape of a surface: an elliptic point (*left*), a hyperbolic point (*center*), and a parabolic point (*right*).

Our analysis is also useful for describing singularities of plane curves X in $\mathbb{P}^2$. We first point out that if p is a smooth point on X, then there is only one principal tangent, which coincides with the tangent line $T_p(X)$, and usually has intersection multiplicity 2 with X. There are also special *flex points* that are smooth points p where the unique tangent line has intersection multiplicity 3 with X. On the other hand, if p is a 2-fold singular point on X (so that only the first order derivatives of F vanish at p) then the tangent cone $C_p(X)$ contains a pair of principal tangents through p that have multiplicity 3 with X. When these two lines are distinct, we say that p is a *nodal singularity* or simply a *node*; when the two lines coincide (which means that $\Delta_x^2(p)$ is the square of a linear form), we say that p is a *cuspidal singularity* or a *cusp*. See Figure B.2.

B.8 Dual Varieties

Let X be an irreducible projective variety in $\mathbb{P}^n$. We say that a hyperplane H in $\mathbb{P}^n$ is tangent to X if there exists a smooth point p on X such that H contains the tangent space $T_p(X)$. The closure $X^\vee$ in $(\mathbb{P}^n)^*$ of the set of all tangent hyperplanes to X is an

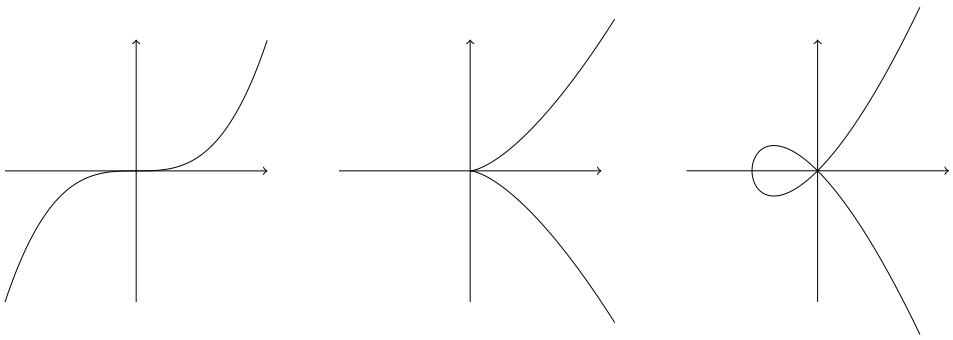


Figure B.2: A flex point (*left*), a cusp singularity (*center*), and a node singularity (*right*).

irreducible projective variety, known as the *dual variety* of X. This name comes from the following important result (see [64] for a proof).

Theorem B.8.1. *For any projective variety X in $\mathbb{P}^n$, we have $X^{\vee\vee} = X$. Moreover, if p is a smooth point of X, then H is tangent to X at p if and only if $p^\vee$, that is a hyperplane in $(\mathbb{P}^n)^*$, is tangent to $X^\vee$ at $H^\vee$.*

Independently of the dimension of X in $\mathbb{P}^n$, the dual variety is typically a hypersurface in $(\mathbb{P}^n)^*$. Indeed, one can show that if the dual variety $X^\vee$ has codimension $r+1$ then X is the union of r-dimensional linear projective spaces lying on X (*i.e.*, it is *ruled*).

Even if X is smooth, the dual variety $X^\vee$ usually has singularities. For example, if X is a smooth curve in $\mathbb{P}^2$, the dual curve $X^\vee$ in $(\mathbb{P}^2)^*$ will have *cusp points* that correspond to tangent lines at *flex points* of X, and it will have *nodes* that correspond to *bitangent lines* for X (*i.e.*, lines that are tangent at two distinct points on X). The expected number of these dual singularities is given by the classical *Plücker formulae*, which we state here in the case of smooth curves.

Theorem B.8.2 (Plücker formulae)**.** *If X is a smooth curve in $\mathbb{P}^2$ of degree d, the dual $X^\vee$ is a curve in $(\mathbb{P}^2)^*$ of degree $d(d-1)$. If X is* generic*, then $X^\vee$ contains $\kappa^\vee = 3(d-2)$ cusps and $\nu^\vee = 1/2d(d-1)(d^2-9)$ nodes. These correspond to $\kappa^\vee$ flex points $\nu^\vee$ bitangents for X.*

B.9 Computations with Sage

This thesis contains many algebraic results and examples that were derived using a computer algebra system. In particular, we used mainly Sage [158] or Macaulay2 [70]. These two softwares are different, in that Macaulay2 is intended exclusively for computations in algebraic geometry and commutative algebra, while Sage [158] is a "general purpose system", that can be used to study a wide variety of topics in mathematics, such as cal-

culus, number theory, cryptography, group theory, combinatorics, and many more. Both softwares are free and open-source. Sage is actually designed as a front-end interface to other existing open-source packages, and for algebraic computations it is mostly based on the system Singular [46]. In terms of functionality, Macaulay2 has more advanced tools that are intended for research in algebraic geometry. For the purposes of this thesis, however, the functions available in Sage are almost always sufficient (one exception being the Macaulay2 command `multidegree`, which is not available in Sage, and is used several times in the thesis). Sage is based on the Python programming language, and can be used in a Jupyter interactive notebook. It is also possible to call functions in Macaulay2 from within Sage. For these reasons, we describe here only some basic examples of commands that can be used for computations in Sage.

To define a polynomial ring over $\mathbb{Q}$ (which is normal in symbolic computation) one can use the command

```
R.<x1,x2,x3,x4> = PolynomialRing(QQ)
```

Here the ring `R` has variables `x1,x2,x3,x4`. We define an ideal in `R` as follows.

```
I_C = ideal(x1*x3 - x2^2, x1*x4 - x2*x3, x2*x4 - x3^2)
```

The ideal `I_C` represents a twisted cubic curve C in $\mathbb{P}^3$. The projection of this curve under the pinhole camera $[x_1, x_2, x_3, x_4] \mapsto [x_1, x_2, x_3]$ can be obtained simply be *eliminating* the variable `x4`:

```
J_C = I_X.elimination_ideal(x4); print(J)
```

The output is:

```
Ideal (x2^2 - x1*x3) of Multivariate Polynomial Ring
in x1, x2, x3, x4 over Rational Field
```

This is a conic in $\mathbb{P}^2$. Note that the pinhole of the projection $[0, 0, 0, 1]$ belonged to the original curve C, which explains is why the degree dropped.

In general, we can always compute projections for the camera $[x_1, x_2, x_3, x_4] \mapsto [x_1, x_2, x_3]$ by simply eliminating the variable `x4`. However, projecting with arbitrary cameras requires a different approach. A possible strategy is illustrated by following code (as in normal Python syntax, hash characters indicate comments in the code).

```
# xi are space variables, uj are image variables
R.<x1,x2,x3,x4,u1,u2,u3> = PolynomialRing(QQ)

# ideal of curve (can be arbitrary)
I_C = ideal(x1*x3 - x2^2, x1*x4 - x2*x3, x2*x4 - x3^2)

# vectors of variables:
X = matrix([[x1],[x2],[x3],[x4]])
```

```
U = matrix([[u1],[u2],[u3]])

# projection matrix (can be arbitrary):
M = matrix([[1,0,0,3],[0,1,0,2],[0,0,1,4]])

# Incidence conditions: U is the image of X
I_incidence = ideal(((M*X).augment(U)).minors(2))

# Full ideal
I_full = I_X + I_incidence
I_full = I_full.saturation(ideal((M*X).list()))[0]

# Image ideal (with optional change of ring)
J_C = I_full.elimination_ideal([x1,x2,x3,x4]).change_ring(QQ[u1,u2,u3])
print(J_C)
```

After defining the vectors `X` and `U` representing variable points in $\mathbb{P}^3$ and $\mathbb{P}^2$ respectively, as well as the projection matrix `M`, we impose that `U` is the image of `X` by requiring that the matrix `[M*X|U]` has rank one (by defining the ideal `I_incidence`). We then define the ideal `I_full` which collects the constraints that `X` belongs to the original curve, and that `U` is the image of `X`. Having written all the necessary conditions, we need to eliminate the variables `x1,x2,x3,x4`. However, we must first remove from `I_full` a spurious component, corresponding to the values of `X` for which `M*X` is zero. For this reason, we *saturate* the ideal `I_full` with respect to the ideal defined by the coordinates of `M*X`, using command `I_full.saturation(ideal((M*X).list()))[0]` (the last argument `0` is required since the Sage command also returns a "saturation exponent"). After this step we can finally eliminate the variables `x1,x2,x3,x4`, and obtain the ideal J_C that represents the projection of the given curve in $\mathbb{P}^2$.[2] For the curve and projection matrix used in the example above, the output is:

```
Ideal (18*u1*u2^2 + 61*u2^3 - 18*u1^2*u3 + 11*u1*u2*u3 -
132*u2^2*u3 + 16*u1*u3^2 + 66*u2*u3^2 - 17*u3^3)
of Multivariate Polynomial Ring in u1, u2, u3 over Rational Field
```

This describes a cubic curve in $\mathbb{P}^2$.

[2] We mention that it is also possible to perform the computation above without saturating, by defining the incidence ideal more simply as `I_incidence = ideal((M*X-U).list())`. However, this is a "trick", since the correct incidence condition is given by the minors of `[M*X|U]`.

Bibliography

[1] Jounaidi Abdeljaoued, Gema M Diaz-Toca, and Laureano Gonzalez-Vega. "Minors of Bezout matrices, subresultants and the parameterization of the degree of the polynomial greatest common divisor". In: *International Journal of Computer Mathematics* 81.10 (2004), pp. 1223–1238.

[2] Edward H Adelson and James R Bergen. *The plenoptic function and the elements of early vision*. Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology, 1991.

[3] Chris Aholt and Luke Oeding. "The ideal of the trifocal variety". In: *Mathematics of Computation* 83.289 (2014), pp. 2553–2574.

[4] Chris Aholt, Bernd Sturmfels, and Rekha Thomas. "A Hilbert Scheme in Computer Vision". In: *Canad. J. Math* 65.5 (2013), pp. 961–988.

[5] Vladimir I Arnold. "Singularities of smooth mappings". In: *Vladimir I. Arnold-Collected Works*. Springer, 1968, pp. 119–161.

[6] Enrique Arrondo. "Introduction to projective varieties". 2017.

[7] Enrique Arrondo, Marina Bertolini, and Cristina Turrini. "A focus on focal surfaces". In: *arXiv preprint math/0002109* (2000).

[8] Kalle Åström, Roberto Cipolla, and Peter J Giblin. "Generalised epipolar constraints". In: *Computer Vision—ECCV'96*. Springer, 1996, pp. 95–108.

[9] Simon Baker and Shree K Nayar. "A theory of single-viewpoint catadioptric image formation". In: *International Journal of Computer Vision* 35.2 (1999), pp. 175–196.

[10] Guillaume Batog. "Classical problems in computer vision and computational geometry revisited with line geometry". Theses. Université Nancy II, Dec. 2011. URL: https://tel.archives-ouvertes.fr/tel-00653043.

[11] Guillaume Batog, Xavier Goaoc, and Jean Ponce. "Admissible linear map models of linear cameras". In: *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE. 2010, pp. 1578–1585.

[12] Bruce Guenther Baumgart. *Geometric modeling for computer vision*. Tech. rep. DTIC Document, 1974.

[13] Mauro Beltrametti, Ettore Carletti, Dionisio Gallarati, and Giacomo Monti Bragadin. *Lectures on curves, surfaces and projective varieties: a classical view of algebraic geometry*. Vol. 9. European Mathematical Society, 2009.

[14] Fausto Bernardini, Joshua Mittleman, Holly Rushmeier, Cláudio Silva, and Gabriel Taubin. "The ball-pivoting algorithm for surface reconstruction". In: *Visualization and Computer Graphics, IEEE Transactions on* 5.4 (1999), pp. 349–359.

[15] Marie-Amélie Bertin. "On the singularities of the trisecant surface to a space curve". In: *Le Matematiche* 53.3 (1998), pp. 15–22.

[16] Silvia Biasotti, Daniela Giorgi, Michela Spagnuolo, and Bianca Falcidieno. "Reeb graphs for shape analysis and applications". In: *Theoretical Computer Science* 392.1 (2008), pp. 5–22.

[17] Alexander Bobenko. *Discrete differential geometry*. Springer, 2008.

[18] Adam Boocher. "Free resolutions and sparse determinantal ideals". In: *arXiv preprint arXiv:1111.0279* (2011).

[19] Andrea Bottino and Aldo Laurentini. "Introducing a new problem: Shape-from-silhouette when the relative positions of the viewpoints is unknown". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 25.11 (2003), pp. 1484–1493.

[20] Andrea Bottino and Aldo Laurentini. "The visual hull of piecewise smooth objects." In: *Computer Vision and Image Understanding ()* 110.1 (2008), pp. 7–18.

[21] Andrea Bottino and Aldo Laurentini. "The Visual Hull of Smooth Curved Objects." In: *IEEE Trans. Pattern Anal. Mach. Intell. ()* 26.12 (2004), pp. 1622–1632.

[22] Kevin W Bowyer and Charles R Dyer. "Aspect graphs: An introduction and survey of recent results". In: *International Journal of Imaging Systems and Technology* 2.4 (1990), pp. 315–328.

[23] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[24] Edmond Boyer. "On using silhouettes for camera calibration". In: *Computer Vision–ACCV 2006*. Springer, 2006, pp. 1–10.

[25] Edmond Boyer and Jean-Sébastien Franco. "A hybrid approach for computing visual hulls of complex objects". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society Press. 2003, pp. 695–701.

[26] Th Bröcker and L Lander. *Differentiable germs and catastrophes*. Vol. 17. Cambridge University Press, 1975.

[27] James William Bruce and Peter J Giblin. *Curves and Singularities: a geometrical introduction to singularity theory*. Cambridge university press, 1992.

[28] Boris Bukh, Xavier Goaoc, Alfredo Hubard, and Matthew Trager. "Consistent sets of lines with no colorful incidence". In: *International Symposium on Computational Geometry*. 2018.

[29] Neill DF Campbell, George Vogiatzis, Carlos Hernández, and Roberto Cipolla. "Automatic 3d object segmentation in multiple views using volumetric graph-cuts". In: *Image and Vision Computing* 28.1 (2010), pp. 14–25.

[30] S. Carlsson and D. Weinshall. "Dual computation of projective shape and camera positions from mutliple images". In: 27.3 (1998), pp. 227–241.

[31] Stefan Carlsson. "Duality of reconstruction and positioning from projective views". In: *Representation of Visual Scenes, 1995.(In Conjuction with ICCV'95), Proceedings IEEE Workshop on*. IEEE. 1995, pp. 85–92.

[32] Fabrizio Catanese and Cecilia Trifogli. "Focal loci of algebraic varieties I". In: *Communications in Algebra* 28.12 (2000), pp. 6017–6057.

[33] Kong Man German Cheung. "Visual hull construction, alignment and refinement for human kinematic modeling, motion tracking and rendering". PhD thesis. Citeseer, 2003.

[34] C H Chien and J K Aggarwal. "Volume/surface octrees for the representation of three-dimensional objects". In: *Computer Vision, Graphics, and Image Processing* 36.1 (Oct. 1986), pp. 100–113.

[35] David Claus and Andrew W Fitzgibbon. "A rational function lens distortion model for general cameras". In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 1. IEEE. 2005, pp. 213–219.

[36] Arthur B Coble. "Point sets and allied Cremona groups". In: *Proceedings of the National Academy of Sciences* 1.4 (1915), pp. 245–248.

[37] Susan Jane Colley. "Enumerating stationary multiple-points". In: *Advances in Mathematics* 66.2 (1987), pp. 149–170.

[38] Susan Jane Colley. "Lines having specified contact with projective varieties". In: *Proceedings of the 1984 Vancouver conference in algebraic geometry*. Vol. 6. 1986, pp. 47–70.

[39] David A Cox, John Little, and Donal O'Shea. "Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra". In: (2015).

[40] Harold Scott Macdonald Coxeter. *Projective geometry*. Springer Science & Business Media, 2003.

[41] Daniel Cremers and Kalin Kolev. "Multiview stereo and silhouette consistency via convex functionals over convex domains". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 33.6 (2011), pp. 1161–1174.

[42] Geoff Cross, Andrew W Fitzgibbon, and Andrew Zisserman. "Parallax geometry of smooth surfaces in multiple views". In: *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*. Vol. 1. IEEE. 1999, pp. 323–329.

[43] Richard H Crowell and Ralph Hartzler Fox. *Introduction to knot theory*. Vol. 57. Springer Science & Business Media, 2012.

[44] Pietro De Poi. "Congruences of lines with one-dimensional focal locus". In: *Portugaliae Mathematica* 61.3 (2004), pp. 329–338.

[45] Pietro De Poi, Emilia Mezzetti, et al. "On a class of first order congruences of lines". In: *Bulletin of the Belgian Mathematical Society-Simon Stevin* 16.5 (2009), pp. 805–821.

[46] Wolfram Decker, Gert-Martin Greuel, Gerhard Pfister, and Hans Schönemann. Singular *4-1-1 — A computer algebra system for polynomial computations*. http://www.singular.uni-kl.de. 2018.

[47] M. Demazure. *Sur deux problemes de reconstruction*. Tech. rep. RR-0882. INRIA, July 1988.

[48] Jan Draisma, Emil Horobeţ, Giorgio Ottaviani, Bernd Sturmfels, and Rekha R Thomas. "The Euclidean distance degree of an algebraic variety". In: *Foundations of Computational Mathematics* 16.1 (2016), pp. 99–149.

[49] Josef Maria Eder. *History of Photography*. Dover Publications, New York, 1945.

[50] William Leonard Edge. *The Theory of Ruled Surfaces*. Cambridge University Press, 1931.

[51] David Eisenbud and Joe Harris. *3264 and all that: A second course in algebraic geometry*. Cambridge University Press, 2016.

[52] Laura Escobar and Allen Knutson. "The multidegree of the multi-image variety". In: *arXiv preprint arXiv:1701.03852* (2017).

[53] Olivier Faugeras, Quang-Tuan Luong, and Theo Papadopoulo. *The geometry of multiple images: the laws that govern the formation of multiple images of a scene and some of their applications*. MIT press, 2004.

[54] Olivier Faugeras and Bernard Mourrain. "On the geometry and algebra of the point and line correspondences between n images". In: *Computer Vision, 1995. Proceedings., Fifth International Conference on*. IEEE. 1995, pp. 951–956.

[55] Doron Feldman, Tomas Pajdla, and Daphna Weinshall. "On the epipolar geometry of the crossed-slits projection". In: *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. IEEE. 2003, pp. 988–995.

[56] David A. Forsyth and Jean Ponce. *Computer Vision: A Modern Approach*. Prentice Hall Professional Technical Reference, 2002. ISBN: 0130851981.

[57] David A Forsyth and Jean Ponce. "Computer Vision - A Modern Approach, Second Edition." In: *Pitman 2012* (2012).

[58] Elisabetta Fortuna, Roberto Frigerio, and Rita Pardini. *Projective Geometry: Solved Problems and Theory Review*. Vol. 104. Springer, 2016.

[59] Jean-Sébastien Franco and Edmond Boyer. "Efficient polyhedral modeling from silhouettes". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31.3 (2009), pp. 414–427.

[60] Jean-Sébastien Franco and Edmond Boyer. "Exact polyhedral visual hulls". In: *British Machine Vision Conference (BMVC'03)*. Vol. 1. 2003, pp. 329–338.

[61] William Fulton. *Intersection theory*. Vol. 2. Springer Science & Business Media, 2013.

[62] Yasutaka Furukawa, Amit Sethi, Jean Ponce, and David Kriegman. "Structure and motion from images of smooth textureless objects". In: *Computer Vision-ECCV 2004*. Springer, 2004, pp. 287–298.

[63] Richard J Gardner. *Geometric tomography*. Vol. 6. Cambridge University Press Cambridge, 1995.

[64] Israel M Gelfand, Mikhail Kapranov, and Andrei Zelevinsky. *Discriminants, resultants, and multidimensional determinants*. Springer Science & Business Media, 2008.

[65] William Goldman. "Projective geometry on manifolds". In: *Lecture notes, University of Maryland* (1988).

[66] Gene H Golub and Charles F Van Loan. *Matrix computations*. Vol. 3. JHU Press, 2012.

[67] Vladimir Petrovich Golubyatnikov and V Yu Rovenski. "Some extensions of the class of k-convex bodies". In: *Siberian Mathematical Journal* 50.5 (2009), pp. 820–829.

[68] Jacob E Goodman. "When is a set of lines in space convex". In: *Notices of the AMS* 45.2 (1998), pp. 222–232.

[69] Jacob E Goodman and Richard Pollack. "Convexity on affine Grassmann manifolds". In: *International Mathematics Research Notices* 1994.3 (1994), pp. 141–150.

[70] Daniel R. Grayson and Michael E. Stillman. *Macaulay2, a software system for research in algebraic geometry*. Available at http://www.math.uiuc.edu/Macaulay2/.

[71] Michael D Grossberg and Shree K Nayar. "The raxel imaging model and ray-based calibration". In: *International Journal of Computer Vision* 61.2 (2005), pp. 119–137.

[72] Rajiv Gupta and Richard I Hartley. "Linear pushbroom cameras". In: *IEEE Transactions on pattern analysis and machine intelligence* 19.9 (1997), pp. 963–975.

[73] Joe Harris. *Algebraic geometry: a first course*. Vol. 133. Springer Science & Business Media, 2013.

[74] R. Hartley. "Computation of the quadrifocal tensor". In: 1998, pp. 20–35.

[75] Richard I Hartley. "Lines and points in three views and the trifocal tensor". In: *International Journal of Computer Vision* 22.2 (1997), pp. 125–140.

[76] Richard I Hartley and Tushar Saxena. "The cubic rational polynomial camera model". In: *Image Understanding Workshop*. Vol. 649. 1997, p. 653.

[77] Richard I Hartley and Frederik Schaffalitzky. "Reconstruction from projections using grassmann tensors". In: *International journal of computer vision* 83.3 (2009), pp. 274–293.

[78] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[79] Robin Hartshorne. *Algebraic geometry*. Vol. 52. Springer Science & Business Media, 2013.

[80] Herwig Hauser. "The Hironaka theorem on resolution of singularities (or: A proof we always wanted to understand)". In: *Bulletin of the American Mathematical Society* 40.3 (2003), pp. 323–403.

[81] Carlos Hernández, Francis Schmitt, and Roberto Cipolla. "Silhouette coherence for camera calibration under circular motion". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29.2 (2007), pp. 343–349.

[82] Jürgen Herzog, Takayuki Hibi, Freyja Hreinsdóttir, Thomas Kahle, and Johannes Rauh. "Binomial edge ideals and conditional independence statements". In: *Advances in Applied Mathematics* 45.3 (2010), pp. 317–333.

[83] Anders Heyden. "Tensorial properties of multiple view constraints". In: *Mathematical Methods in the Applied Sciences* 23.2 (2000), pp. 169–202.

[84] Anders Heyden and Kalle Åström. "Algebraic properties of multilinear constraints". In: *Mathematical Methods in the Applied Sciences* 20.13 (1997), pp. 1135–1162.

[85] D. Hilbert and S. Cohn-Vossen. *Geometry and the Imagination*. Chelsea Scientific Books. Chelsea Publishing Company, 1952.

[86] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 1990.

[87] Vyvyan Howard and Matthew Reed. *Unbiased stereology: three-dimensional measurement in microscopy*. Garland Science, 2004.

[88] Yong Hu, Vincent Tao, and Arie Croitoru. "Understanding the rational function model: methods and applications". In: *International Archives of Photogrammetry and Remote Sensing* 20.6 (2004).

[89] Hiroshi Ishiguro, Masashi Yamamoto, and Saburo Tsuji. "Omni-directional stereo". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14.2 (1992), pp. 257–262.

[90] Atsushi Ito, Makoto Miura, and Kazushi Ueda. "Projective reconstruction in algebraic vision". In: *arXiv preprint arXiv:1710.06205* (2017).

[91] David Jacobs, Peter Belhumeur, and Ian Jermyn. "Judging whether multiple silhouettes can come from the same object". In: *Visual Form 2001*. Springer, 2001, pp. 532–541.

[92] C M Jessop. *A Treatise on the Line Complex*. The University Press, 1903.

[93] Pal Hermunn Johansen. "The geometry of the tangent developable". In: *Computational methods for algebraic spline surfaces*. Springer, 2005, pp. 95–106.

[94] Tanuja Joshi, Narendra Ahuja, and Jean Ponce. "Structure and motion estimation from dynamic silhouettes under perspective projection". In: *Computer Vision, 1995. Proceedings., Fifth International Conference on*. IEEE. 1995, pp. 290–295.

[95] Alfrederic Josse and Francoise Pene. "On caustics by reflection of algebraic surfaces". In: *Advances in Geometry* 16.4 (2016), pp. 437–464.

[96] Fredrik Kahl and Anders Heyden. "Using conic correspondences in two images to estimate the epipolar geometry". In: *Computer Vision, 1998. Sixth International Conference on*. IEEE. 1998, pp. 761–766.

[97] Michael Kalkbrener. "On the stability of Gröbner bases under specializations". In: *Journal of Symbolic Computation* 24.1 (1997), pp. 51–58.

[98] Yannick Louis Kergosien. "La famille des projections orthogonales d'une surface et ses singularités". In: *CR Acad. Sc. Paris* 292 (1981), pp. 929–932.

[99] Jan J Koenderink. *Solid shape*. Vol. 2. Cambridge Univ Press, 1990.

[100] Jan J Koenderink and Andrea J van Doorn. "The internal representation of solid shape with respect to vision". In: *Biological cybernetics* 32.4 (1979), pp. 211–216.

[101] Jan J Koenderink and Andrea J Van Doorn. "The singularities of the visual mapping". In: *Biological cybernetics* 24.1 (1976), pp. 51–59.

[102] Kathlén Kohn. "Coisotropic hypersurfaces in the Grassmannian". In: *arXiv preprint arXiv:1607.05932* (2016).

[103] Kathlén Kohn, Bernt Ivar Utstøl Nødland, and Paolo Tripoli. "Secants, bitangents, and their congruences". In: *Combinatorial Algebraic Geometry*. Springer, 2017, pp. 87–112.

[104] Kathlén Kohn, Bernd Sturmfels, and Matthew Trager. "Changing Views on Curves and Surfaces". In: *Acta Mathematica Vietnamica* 43.1 (2018), pp. 1–29.

[105] E. Kummer. "Über die algebraischen Strahlensysteme, insbesondere über die der ersten und zweiten Ordnung". In: *Abh. K. Preuss. Akad. Wiss. Berlin* (1866), pp. 1–120.

[106] DG Larman, P Mani, et al. "On visual hulls". In: *Pacific J. Math* 32.1 (1970), pp. 157–171.

[107] A Laurentini. "How far 3D shapes can be understood from 2D silhouettes". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17.2 (1995), pp. 188–195.

[108] Aldo Laurentini. "The Visual Hull Concept for Silhouette-Based Image Understanding." In: *IEEE Trans. Pattern Anal. Mach. Intell.* 16.2 (1994), pp. 150–162.

[109] Svetlana Lazebnik, Edmond Boyer, and Jean Ponce. "On computing exact visual hulls of solids bounded by smooth surfaces". In: *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. Vol. 1. IEEE. 2001, pp. I–156.

[110] Svetlana Lazebnik, Yasutaka Furukawa, and Jean Ponce. "Projective visual hulls". In: *International Journal of Computer Vision* 74.2 (2007), pp. 137–165.

[111] Svetlana Lazebnik, Amit Sethi, Cordelia Schmid, David Kriegman, Jean Ponce, and Martial Hebert. "On pencils of tangent planes and the recognition of smooth 3D shapes from silhouettes". In: *Computer Vision—ECCV 2002*. Springer, 2002, pp. 651–665.

[112] Hwangrae Lee and Bernd Sturmfels. "Duality of multiple root loci". In: *Journal of Algebra* 446 (2016), pp. 499–526.

[113] Noam Levi and Michael Werman. "The viewing graph". In: *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*. Vol. 1. IEEE. 2003, pp. I–I.

[114] Marc Levoy and Pat Hanrahan. "Light field rendering". In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM. 1996, pp. 31–42.

[115] Binglin Li. "Images of rational maps of projective spaces". In: *International Mathematics Research Notices* (2017), rnx003.

[116] Max Lieblich and Lucas Van Meter. "Two Hilbert schemes in computer vision". In: *arXiv preprint arXiv:1707.09332* (2017).

[117] Shaowei Lin and Bernd Sturmfels. "Polynomial relations among principal minors of a 4× 4-matrix". In: *Journal of Algebra* 322.11 (2009), pp. 4121–4131.

[118] Qing Liu. *Algebraic geometry and arithmetic curves*. Vol. 6. Oxford University Press on Demand, 2002.

[119] H Christopher Longuet-Higgins. "A computer algorithm for reconstructing a scene from two projections". In: *Nature* 293.5828 (1981), p. 133.

[120] Q.-T. Luong and O.D. Faugeras. "The Fundamental matrix: theory, algorithms, and stability analysis". In: 17.1 (1996), pp. 43–76.

[121] Worthy N Martin and J K Aggarwal. "Volumetric Descriptions of Objects from Multiple Views". In: *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-5.2 (1983), pp. 150–158.

[122] Wojciech Matusik, Chris Buehler, and Leonard McMillan. "Polyhedral Visual Hulls for Real-Time Rendering". In: *Volume Graphics 2001*. Vienna: Springer Vienna, 2001, pp. 115–125.

[123] Wojciech Matusik, Chris Buehler, Ramesh Raskar, Steven J Gortler, and Leonard McMillan. "Image-based visual hulls". In: *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co. 2000, pp. 369–374.

[124] Stephen Maybank. *Theory of reconstruction from image motion*. Vol. 28. Springer Science & Business Media, 1993.

[125] Paulo RS Mendonça, Kwan-Yee K Wong, and R Cipolla. "Epipolar geometry from profiles under circular motion". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 23.6 (2001), pp. 604–616.

[126] Ezra Miller and Bernd Sturmfels. *Combinatorial Commutative Algebra*. Vol. 227. Springer Science & Business Media, 2006.

[127] Niloy J Mitra and Mark Pauly. "Shadow art". In: *ACM Transactions on Graphics*. Vol. 28. EPFL-CONF-149329. 2009, pp. 156–1.

[128] D. Mumford. *Abelian Varieties*. Studies in mathematics. Hindustan Book Agency, 2008. ISBN: 9788185931869.

[129] James R Munkres. "Topology: a first course". In: *Englewood Cliffs, New Jersey* (1975).

[130] Shree K Nayar, Vlad Branzoi, and Terry E Boult. "Programmable imaging: Towards a flexible camera". In: *International Journal of Computer Vision* 70.1 (2006), pp. 7–22.

[131] Brian Osserman and Matthew Trager. "Multigraded Cayley-Chow forms". In: *arXiv preprint arXiv:1708.03335* (2017). Submitted to *Advances in Mathematics*.

[132] Onur Ozyesil and Amit Singer. "Robust camera location estimation by convex programming". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, pp. 2674–2683.

[133] Onur Ozyesil, Vladislav Voroninski, Ronen Basri, and Amit Singer. "A survey of structure from motion." In: *Acta Numerica* 26 (2017), pp. 305–364.

[134] Sung-Il Pae and Jean Ponce. "On computing structural changes in evolving surfaces and their appearance". In: *International Journal of Computer Vision* 43.2 (2001), pp. 113–131.

[135] Tomáš Pajdla. "Geometry of two-slit camera". In: *Rapport Technique CTU-CMP-2002-02, Center for Machine Perception, Czech Technical University, Prague* (2002).

[136] Tomáš Pajdla. "Stereo with oblique cameras". In: *International Journal of Computer Vision* 47.1-3 (2002), pp. 161–170.

[137] Shmuel Peleg and Moshe Ben-Ezra. "Stereo panorama with a single camera". In: *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.* Vol. 1. IEEE. 1999, pp. 395–401.

[138] Sylvain Petitjean. "The complexity and enumerative geometry of aspect graphs of smooth surfaces". In: *Algorithms in algebraic geometry and applications*. Springer, 1996, pp. 317–352.

[139] Sylvain PetitJean, Jean Ponce, and David J Kriegman. "Computing exact aspect graphs of curved objects: Algebraic surfaces". In: *International Journal of Computer Vision* 9.3 (1992), pp. 231–255.

[140] Ragni Piene. "Cuspidal projections of space curves". In: *Mathematische Annalen* 256.1 (1981), pp. 95–119.

[141] Ragni Piene. "Numerical characters of a curve in projective n-space". In: *Real and complex singularities, Oslo* (1976), pp. 475–495.

[142] OA Platonova. "Projections of smooth surfaces". In: *Journal of Soviet Mathematics* 35.6 (1986), pp. 2796–2808.

[143] Robert Pless. "Using many cameras as one". In: *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on.* Vol. 2. IEEE. 2003, pp. II–587.

[144] J. Ponce, T. Papadopoulo, M. Teillaud, and B. Triggs. "The absolute quadratic complex and its application to camera self calibration". In: 2005.

[145] Jean Ponce. "What is a camera?" In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on.* IEEE. 2009, pp. 1526–1533.

[146] Jean Ponce and Martial Hebert. "On Image Contours of Projective Shapes". In: *European Conference on Computer Vision*. 2014.

[147] Jean Ponce and David J Kriegman. *Computing exact aspect graphs of curved objects: Parametric surfaces*. Department of Computer Science, University of Illinois at Urbana-Champaign, 1990.

[148] Jean Ponce, David H Marimont, and Todd A Cass. "Analytical methods for uncalibrated stereo and motion reconstruction". In: *European Conference on Computer Vision*. Springer. 1994, pp. 463–470.

[149] Jean Ponce, Bernd Sturmfels, and Matthew Trager. "Congruences and concurrent lines in multi-view geometry". In: *Advances in Applied Mathematics* 88 (2017), pp. 62–91.

[150] Helmut Pottmann and Johannes Wallner. *Computational line geometry*. Springer Science & Business Media, 2009.

[151] Long Quan. "Invariants of six points and projective reconstruction from three uncalibrated images". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17.1 (1995), pp. 34–46.

[152] Paul Rademacher and Gary Bishop. "Multiple-center-of-projection images". In: *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*. ACM. 1998, pp. 199–206.

[153] Kristian Ranestad and Bernd Sturmfels. "On the convex hull of a space curve". In: *arXiv preprint arXiv:0912.2986* (2009).

[154] JH Rieger. "Computing view graphs of algebraic surfaces". In: *Journal of Symbolic Computation* 16.3 (1993), pp. 259–272.

[155] JH Rieger. "Global bifurcation sets and stable projections of nonsingular algebraic surfaces". In: *International Journal of Computer Vision* 7.3 (1992), pp. 171–194.

[156] JH Rieger. "Three-dimensional motion from fixed points of a deforming profile curve". In: *Optics Letters* 11.3 (1986), pp. 123–125.

[157] Alessandro Rudi, Matia Pizzoli, and Fiora Pirri. "Linear solvability in the viewing graph". In: *Asian Conference on Computer Vision*. Springer. 2010, pp. 369–381.

[158] The Sage Developers. *SageMath, the Sage Mathematics Software System (Version 8.0.0)*. `http://www.sagemath.org`. 2017.

[159] George Salmon. *A treatise on the analytic geometry of three dimensions*. 4th. Hodges, Smith, and Company, 1882.

[160] Jun Sato and Roberto Cipolla. "Affine reconstruction of curved surfaces from uncalibrated views of apparent contours". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 21.11 (1999), pp. 1188–1198.

[161] Rolf Schneider. *Convex bodies: the Brunn–Minkowski theory*. 151. Cambridge University Press, 2013.

[162] Anna Seigal and Bernd Sturmfels. "Real rank two geometry". In: *Journal of Algebra* 484 (2017), pp. 310–333.

[163] John Greenlees Semple and Geoffrey Thomas Kneebone. *Algebraic projective geometry*. Oxford University Press, 1998.

[164] A. Shashua. "Algebraic functions for recognition". In: 17.8 (1995), pp. 779–789.

[165] Heung-Yeung Shum, Adam Kalai, and Steven M Seitz. "Omnivergent stereo". In: *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*. Vol. 1. IEEE. 1999, pp. 22–29.

[166] Sudipta N Sinha and Marc Pollefeys. "Camera network calibration and synchronization from silhouettes in archived video". In: *International journal of computer vision* 87.3 (2010), pp. 266–283.

[167] Sudipta N Sinha, Marc Pollefeys, and Leonard McMillan. "Camera network calibration from dynamic silhouettes". In: *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. Vol. 1. IEEE. 2004, pp. I–195.

[168] Greg Slabaugh, Ron Schafer, Tom Malzbender, and Bruce Culbertson. "A Survey of Methods for Volumetric Scene Reconstruction from Photographs". In: *Volume Graphics 2001*. Vienna: Springer Vienna, 2001, pp. 81–100.

[169] M.E. Spetsakis and Y. Aloimonos. "Structure from motion using line correspondences". In: 4.3 (1990), pp. 171–183.

[170] Jorge Stolfi. *Oriented projective geometry: A framework for geometric computations*. Academic Press, 2014.

[171] Peter Sturm. "A historical survey of geometric computer vision". In: *Computer Analysis of Images and Patterns*. Springer. 2011, pp. 1–8.

[172] Peter Sturm. "Multi-view geometry for general camera models". In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 1. IEEE. 2005, pp. 206–212.

[173] Peter Sturm, Srikumar Ramalingam, Jean-Philippe Tardif, Simone Gasparini, and Joao Barreto. "Camera models and fundamental concepts used in geometric computer vision". In: *Foundations and Trends in Computer Graphics and Vision* 6.1–2 (2011), pp. 1–183.

[174] Bernd Sturmfels. *Algorithms in invariant theory*. Springer Science & Business Media, 2008.

[175] Bernd Sturmfels. "Computational algebraic geometry of projective configurations". In: *Journal of symbolic computation* 11.5-6 (1991), pp. 595–618.

[176] Bernd Sturmfels. "The Hurwitz form of a projective variety". In: *Journal of Symbolic Computation* 79 (2017), pp. 186–196.

[177] Chris Sweeney, Torsten Sattler, Tobias Hollerer, Matthew Turk, and Marc Pollefeys. "Optimizing the viewing graph for structure-from-motion". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2015, pp. 801–809.

[178] C Vincent Tao and Yong Hu. "A comprehensive study of the rational function model for photogrammetric processing". In: *Photogrammetric engineering and remote sensing* 67.12 (2001), pp. 1347–1358.

[179] Konstantinos A Tarabanis, Peter K Allen, and Roger Y Tsai. "A survey of sensor planning in computer vision". In: *Robotics and Automation, IEEE Transactions on* 11.1 (1995), pp. 86–104.

[180] Evgueni A Tevelev. *Projective duality and homogeneous spaces*. Vol. 133. Springer Science & Business Media, 2006.

[181] SriRam Thirthala and Marc Pollefeys. "Radial multi-focal tensors". In: *International journal of computer vision* 96.2 (2012), pp. 195–211.

[182] René Thom. *Structural stability and morphogenesis*. CRC Press, 2018.

[183] M.M. Thompson, R.C. Eller, W.A. Radlinski, and J.L. Speert, eds. *Manual of Photogrammetry*. Third Edition. American Society of Photogrammetry, 1966.

[184] Csaba D Toth, Joseph O'Rourke, and Jacob E Goodman. *Handbook of discrete and computational geometry*. CRC press, 2004.

[185] Matthew Trager, Martial Hebert, and Jean Ponce. "Consistency of silhouettes and their duals". In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2016.

[186] Matthew Trager, Martial Hebert, and Jean Ponce. "The joint image handbook". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2015, pp. 909–917.

[187] Matthew Trager, Brian Osserman, and Jean Ponce. "On the Solvability of Viewing Graphs". In: *European Conference on Computer Vision*. 2018.

[188] Matthew Trager, Jean Ponce, and Martial Hebert. "Trinocular Geometry Revisited". In: *International Journal of Computer Vision* (2016), pp. 1–19.

[189] Matthew Trager, Bernd Sturmfels, John Canny, Martial Hebert, and Jean Ponce. "General models for rational cameras and the case of two-slit projections". In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2017.

[190] Bill Triggs. "Matching constraints and the joint image". In: *Computer Vision, 1995. Proceedings., Fifth International Conference on*. IEEE. 1995, pp. 338–343.

[191] Patrick L Van Hove. *Silhouette-Slice Theorems*. Tech. rep. DTIC Document, 1986.

[192] J. Weng, T.S. Huang, and N. Ahuja. "Motion and structure from line correspondences: closed-form solution, uniqueness, and optimization". In: 14.3 (1992), pp. 318–336.

[193] Hassler Whitney. "On singularities of mappings of Euclidean spaces. I. Mappings of the plane into the plane". In: *Annals of Mathematics* (1955), pp. 374–410.

[194] James V Whittaker. "A mountain-climbing problem". In: *Canad. J. Math* 18 (1966), pp. 873–882.

[195] Kwan-Yee Kenneth Wong and Roberto Cipolla. "Reconstruction of sculpture from its profiles with unknown camera positions". In: *Image Processing, IEEE Transactions on* 13.3 (2004), pp. 381–389.

[196] Li Yi and David W Jacobs. "Efficiently Determining Silhouette Consistency". In: *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE. 2007, pp. 1–8.

[197] Jingyi Yu and Leonard McMillan. "General linear cameras". In: *European Conference on Computer Vision*. Springer. 2004, pp. 14–27.

[198] E Christopher Zeeman. "Catastrophe theory". In: *Scientific American* 234.4 (1976), pp. 65–83.

[199] Enliang Zheng, Ke Wang, Enrique Dunn, and Jan-Michael Frahm. "Minimal solvers for 3d geometry from satellite imagery". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2015, pp. 738–746.

[200] Jiang Yu Zheng and Saburo Tsuji. "Panoramic representation of scenes for route understanding". In: *Pattern Recognition, 1990. Proceedings., 10th International Conference on*. Vol. 1. IEEE. 1990, pp. 161–167.

[201] Assaf Zomet, Doron Feldman, Shmuel Peleg, and Daphna Weinshall. "Mosaicing new views: The crossed-slits projection". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25.6 (2003), pp. 741–754.

Résumé

Cette thèse étudie les modèles mathématiques destinés à décrire la géométrie des processus d'imagerie en vision par ordinateur. Notre approche est enracinée dans le langage de la géométrie projective, qui fournit le cadre le plus général pour l'étude des propriétés des lignes et des incidences qui sont au cœur de la vision géométrique. Nous appliquons également des outils de la géométrie algébrique, car la plupart des objets que nous rencontrons sont décrits par des équations polynomiales. Par exemple, la géométrie de n caméras peut être encodée dans une variété algébrique en $(\mathbb{P}^2)^n$ formée par les correspondances de points. La Grassmannienne des lignes $\mathrm{Gr}(1, \mathbb{P}^3)$ joue également un rôle central dans notre étude. Les surfaces en $\mathrm{Gr}(1, \mathbb{P}^3)$ (ou "congruences") peuvent par exemple être utilisées pour représenter des caméras abstraites, qui associent des points à des lignes. Nous étudions aussi la relation entre les formes 3D et leurs images. En particulier, pour les ensembles arbitraires se projetant sur des silhouettes, l'image est déterminée par l'ensemble des lignes qui rencontrent l'objet observé ; pour les surfaces lisses, le "contour visuel" est déterminé par les lignes qui sont tangentes à la surface. Cette perspective est appliquée dans l'étude des "coques visuelles" et des "événements visuels".

Mots Clés

Modèles géométriques, géométrie projective, caméras, silhouettes, contours visuels.

Abstract

This thesis studies mathematical models for describing the geometry of imaging processes in computer vision. Our approach is rooted in the language of projective geometry, which provides the most general setting for studying properties of lines and incidences that are at the heart of geometric vision. We also apply some tools from algebraic geometry, since many of the objects that we encounter are described by polynomial equations. For example, the multiview geometry of n pinhole cameras (or in fact any type of cameras) can be encoded in the "joint image", that is an algebraic variety in $(\mathbb{P}^2)^n$ formed by all point correspondences. The Grassmannian of lines $\mathrm{Gr}(1, \mathbb{P}^3)$ also plays a central role in our study. In particular, surfaces in the Grassmannian (or "line congruences") can be used to represent abstract cameras, that are mappings from points to viewing lines. In addition to modeling cameras, we investigate the relationship between 3D shapes and their images. For arbitrary sets projecting onto opaque silhouettes, the image is determined by the set of viewing lines that meet the observed object; for smooth surfaces, the "image contour" is determined by the set viewing lines that are tangent to the surface. This perspective is applied to the study of "visual hulls" and "visual events".

Keywords

Geometric models, projective geometry, cameras, silhouettes, image contours.