A Functorial Approach to Algebraic Vision

Lucas Van Meter

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Max Lieblich, Chair

Bianca Viray

Sándor Kovács

Program Authorized to Offer Degree:
Mathematics

University of Washington

Abstract

A Functorial Approach to Algebraic Vision

Lucas Van Meter

Chair of the Supervisory Committee:
Professor Max Lieblich
Mathematics

We study multiview moduli problems that arise in computer vision. We show that these moduli spaces are always smooth and irreducible, in both the calibrated and uncalibrated cases, for any number of views. We also show that these moduli spaces always embed in suitable Hilbert schemes, and that these embeddings are open immersions for more than four views, extending and refining work of Aholt–Sturmfels–Thomas.

We also give a new construction of the space of essential matrices from first principles. This construction enables us to re-prove the fundamental results of Demazure and to re-prove the recent description of the essential variety due to Kileel–Fløystad–Ottaviani as well as extend the classical twisted pair covering of the essential variety.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

This thesis would not have been possible without the support of many people.

I want to thank my advisor, Max Lieblich, for his indispensable advice and guidance, his insights and wealth of mathematical ideas, and his wholehearted encouragement.

During the development of this thesis I also worked closely with Rekha Thomas, Bianca Viray, and Andrew Pryhuber, all of whom have provided essential aspects to this work. In addition I have had useful conversations with Sameer Agarwal, Dan Bragg, Kristin DeVleming, Gabriel Dorfsman-Hopkins, Charlie Godfrey, Siddharth Mathur, Travis Scholl, Tuomas Tajakka, and Alex Voet.

I cannot thank my fellow graduate students enough, both for their mathematical contributions to my education but also for making the last six years a joy.

I want to thank Cathy Hsu for her constant support and for always being there to help me when I needed it most.

I also want to thank my family and friends, without whom I could not have made it this far.

DEDICATION

to Scooter

Chapter 1

INTRODUCTION

Multiview geometry is a subset of computer vision which considers problems arising from the geometry of reconstructing a three-dimensional object using only pictures of the object. The reconstruction of a three-dimensional model of Rome from only pictures posted on Flickr [2] is a particularly dramatic illustration of such a problem. The mathematical structure that makes this possible is the study the of underlying projective geometry and related algebraic objects.

This thesis studies multiview geometry from the perspective given to us by modern functorial algebraic geometry. Specifically, we use techniques such as linear systems, moduli spaces, and deformation theory to shed new light on unsolved problems and their inherent geometric structure. We follow a standard program in algebraic geometry to construct a geometric space (called a moduli space) whose points parametrize all possible configurations of cameras. Using this moduli space, we develop an new approach to reconstructing a three-dimensional model from images of a certain important type of camera called *calibrated cameras*. This multidisciplinary approach to these types of questions in multiview geometry is novel, and this thesis explores some of its many consequences and applications.

The literature on multiview geometry is extensive, and so we are not be able to mention every article that deserves recognition. However, throughout this chapter and thesis, we highlight several references that have been especially influential and helpful for this work. Indeed, we now mention a few such references. First, the bible of multiview geometry is [8], which contains a thorough treatment of all the foundational topics of multiview geometry. Other useful surveys of much of the material can be found in [15] and [18]. Second, the paper that inspired our functorial approach to multiview geometry is [3]. This is the original paper to explicitly make a connection between multiview geometry and a moduli perspective. Much of the work presented in the first part of Chapter 4 reproves results from [3] and then expands on them.

1.1 Multiview geometry and moduli problems

1.1.1 Background on camera configurations

In order to study the data of images from a mathematical point of view, we model a camera as a focal point through which rays of light are projected onto a flat plane to produce an image. In standard coordinates, this map is highly non-linear, but if we model the world as projective space, $\mathbf{P}^3$, and the image as a projective plane, $\mathbf{P}^2$, this map is simply a 3×4 matrix, or in the language of algebraic geometry a rational linear map from $\mathbf{P}^3$ to $\mathbf{P}^2$.

It is impossible to recover much geometric information about an 3D object from a single photo, so we instead consider the situation where several cameras take a photo simultaneously, which we call a *camera configuration*. With n cameras, we obtain n maps from $\mathbf{P}^3$ to $\mathbf{P}^2$, and combining them gives a rational linear map from $\mathbf{P}^3$ to the product space $(\mathbf{P}^2)^n$. We define the *image* of such a camera configuration to be the image of this map inside of $(\mathbf{P}^2)^n$, and we can recover the image of any particular camera in the configuration by projecting onto the appropriate component. This leads us to the fundamental question of multiview geometry:

Question. *What information about a camera configuration's image is required to recover the actual camera configuration, i.e., the rational linear map from* $\mathbf{P}^3$ *to* $(\mathbf{P}^2)^n$*?*

When $n \geq 2$, it is well-known [8] that a camera configuration is uniquely determined (up to an automorphism of $\mathbf{P}^3$) by its image in $(\mathbf{P}^2)^n$ when $n \geq 2$, so we start with a clear upper bound on the amount of information required to reconstruct a camera configuration. However, it is actually quite difficult to determine the image in $(\mathbf{P}^2)^n$ from a collection of images because it is not clear how the points in one image correspond to those in another. Fortunately, computers are able to locate key points in each photo, such as the corner of a building, and identify them between photos, thus producing a finite collection of points in $(\mathbf{P}^2)^n$ that must be in the camera configuration's image. We call each point in such a collection a *point correspondence*; if we have sufficiently many point correspondences, then we can reconstruct the entire camera configuration.

In the specific case of reconstruction for $n = 2$, a classical result in the literature [8] shows how we can recover a configuration of a pair of cameras from just seven point correspondences. We briefly recall this well-known construction because it (i) illustrates, in a concrete way, the usefulness of a parametrizing space and (ii) explains an algorithm that is closely related to my results in Chapter 5. Note that the case $n = 2$ is especially important because real-world reconstruction algorithms often first reconstruct pairs of cameras and then patch these reconstructions together to give a complete reconstruction.

In the case of two cameras, the image of $\mathbf{P}^3$ in $\mathbf{P}^2 \times \mathbf{P}^2$ is a codimension one subvariety. In the projective coordinates $[x_0, x_1, x_2] \times [y_0, y_1, y_2]$, the image is defined by a single bilinear equation

$$a_{00}x_0y_0 + a_{10}x_1y_0 + a_{20}x_2y_0 + a_{01}x_0y_1 + a_{11}x_1y_1 + a_{21}x_2y_1 + a_{02}x_0y_2 + a_{12}x_1y_2 + a_{22}x_2y_2 = 0$$

which up to scaling, we can represent using matrix multiplication

$$\begin{pmatrix} x_0 & x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} y_0 \\ y_1 \\ y_2 \end{pmatrix}.$$

In fact, images of a pair of cameras are exactly those matrices of rank two. If we view such matrices as a 9-dimensional vector space, then the corresponding *linear system* is the projectivization of this vectors space, which is isomorphic to $\mathbf{P}^8$. The set of rank deficient matrices is a codimension one variety of degree three cut out by the determinantal vanishing locus; we call this subvariety the *fundamental variety* and denote it by F.

Now, a point correspondence is as a solution to the bilinear equation defining the image and thus is a linear constraint on the coefficients or equivalently a hyperplane in our linear system $\mathbf{P}^8$. The intersection of this hyperplane with the fundamental variety is exactly those camera configurations that contain the point correspondence in their image. Seven generic point correspondences will correspond to seven linear constraints and over $\mathbf{C}$, this will give us exactly three solutions since F is degree three. Generically two of these solutions are complex, but one must be real, and that real point in $\mathbf{P}^8$ corresponds to the unique camera configuration satisfying those seven point correspondences. In other words, we found the unique image in $\mathbf{P}^2 \times \mathbf{P}^2$ that contains these seven point correspondences and can use that image to reconstruct the camera configuration.

1.1.2 A functorial approach to reconstruction

While previous methods for reconstructing camera configurations use linear algebra and explicit computations in a fixed coordinate system to study the space of images of cameras, this thesis tackles the problem from a completely different perspective which is rooted in functorial algebraic geometry. Rather than studying the space of images of cameras, we study the space of camera configurations themselves. Initially, this new approach contains many obstacles, the first of which is reinterpreting our question of camera reconstruction in terms of a functor. Specifically, we must define cameras over arbitrary rings rather than only C or R and then extend previous results to our new context. Then we construct a *moduli space* Cam_n, and relate it to the image of camera configurations by embedding it into a Hilbert scheme, denoted $\mathrm{Hilb}_{(\mathbf{P}^2)^n}$:

Theorem 1.1.1. *There exists a smooth irreducible variety* Cam_n *parametrizing* n*-view camera configurations.*

The map sending a configuration to its multiview variety defines a locally closed embedding

$$\mathrm{Cam}_n \longrightarrow \mathrm{Hilb}_{(\mathbf{P}^2)^n}$$

$$\Phi \longrightarrow \mathrm{im}(\Phi)$$

that is an open immersion when the camera centers are not collinear or $n > 4$.

For example, Cam_2 is isomorphic to the rank two locus of the fundamental variety, and the embedding of $\mathbf{F}$ into $\mathbf{P}^8$ is the same as the embedding of Cam_2 into the Hilbert scheme.

In general, the points of Cam_n parametrize the set of possible camera configurations and we develop tools to identify camera configurations by studying the geometry of this moduli space. In particular, because the Hilbert scheme is an extremely well-studied moduli space, we immediately gain a finer understanding of the geometry of Cam_n from our embedding of Cam_n into $\mathrm{Hilb}_{(\mathbf{P}^2)^n}$. This result extends previous work [3] and lays a foundation for a functorial perspective.

1.2 New results for calibrated cameras

1.2.1 Calibrated cameras

The real value of this functorial approach becomes apparent when we consider a more specialized question. So far, we have assumed no additional information about the structure of our cameras, but in many real-world application, we actually know some internal parameters of each camera. For example, we might know how a camera distorts its image from having non-square pixels or how it skews its image in post processing. Such features, which are completely unrelated to the camera's position or angle in space, are called *calibration data*. If we have access to a camera, such as on a self-driving car, we can determine this calibration data beforehand and use this additional information to reproduce more accurate reconstructions from fewer point correspondences. In fact, by using calibration data, there only exists two camera configurations [8] (up to translation, scaling and rotation of world coordinates) corresponding to a given image, which is a major improvement over the usual ambiguity of a projective automorphism. This 2-to-1 reconstruction phenomenon will be explained in section 2.3.

A significant portion of the multiview geometry literature is devoted to studying calibrated camera configurations because it has important application but also because it is more complicated. To see how these complications arise let us again consider the case of two cameras and the linear system of bilinear forms on $\mathbf{P}^2 \times \mathbf{P}^2$. It was easy to determine if a point of this linear system came from a camera configuration because they are exactly the rank two points, but determining if that system is calibrated is much harder. It was not until 1988 that Demazure [5] showed that a system of 10 cubic equations define the set of all calibrated cameras, called the essential variety and denoted E, inside of $\mathbf{P}^8$. The essential variety is a degree 10 codimension 3 subvariety of $\mathbf{P}^8$. These equations allow us to run reconstruction algorithms (now with just 5 point correspondences) but it is more complicated due to the complexity of the equations.

1.2.2 New approach to calibration

Applying our strategy of using moduli spaces to understand calibrated camera configurations faces many new challenges. Until now there did not exists any construction that paramatrized calibrated cameras or their images for more than three cameras. The largest hurdle was that there are many equivalent ways to define and keep track of a cameras calibration and the traditional approach of thinking of calibrated cameras as special kinds of 3×4 matrix do not easily lend themselves to a moduli space construction. A more geometric approach is to keep track of calibration by using a conic curve in $\mathbf{P}^3$ and keeping track of its image in $\mathbf{P}^2$ under the camera map. Specifically, we define a pair (Φ, C) to be a *normalized camera configuration* if Φ is a camera configuration and C is a degree two curve in $\mathbf{P}^3$ such that Φ maps C to the conic curve in $\mathbf{P}^2$ defined by the equation $x^2 + y^2 + z^2 = 0$. Note that the automorphism of $\mathbf{P}^3$ that fix C are exactly those that do not alter a cameras calibration, so by keeping track of this pair we are keeping track of calibration data.

This new geometric conceptualization of calibration is the key to constructing a moduli space CalCam_n that parametrizes normalized camera configurations, which we do in chapter 4. As in the case of Cam_n, we want to embed this space into a Hilbert scheme, and in order to do so with CalCam_n, we generalize the Hilbert scheme to a new space which we call a *diagram Hilbert scheme*. While the Hilbert scheme keeps track of the embedding of the image into $\mathbf{P}^2 \times \mathbf{P}^2$, the diagram Hilbert scheme keeps track of the entire data of the image of $\mathbf{P}^3$ as well as the the image of the smooth calibrating conic C.

Theorem 1.2.1. *There exists a smooth irreducible variety* CalCam_n *parametrizing* n*-view calibrated camera configurations.*

There is a natural locally closed embedding

$$\mathrm{CalCam}_n \longrightarrow \mathrm{Hilb}_{C\times\cdots\times C\subset(\mathbf{P}^2)^n}$$

$$\Phi \longrightarrow \mathrm{im}(C) \subset \mathrm{im}(\Phi)$$

that is an open immersion when the camera centers are not collinear or $n > 4$.

This construction, along with its embedding into a relatively well-understood space, gives us a powerful tool to study systems of calibrated cameras.

1.2.3 Applications to calibrated reconstruction

We now revisit the case of two cameras, but this time include calibration data. The points of CalCam_2 parametrize pairs $(\mathbf{\Phi}, C)$, where $\mathbf{\Phi}$ is a camera configuration and $C \subset \mathbf{P}^3$ is a degree two curve used to keep track of the calibration data. By forgetting the calibration data, we obtain a map $\mathrm{CalCam}_2 \to \mathrm{Cam}_2$, and the image of this map is exactly the essential variety $\mathbf{E}$. Moreover, we show that for every point (Φ, C) there exists a unique calibrating curve C' so that the point (Φ, C') is distinct from the first in CalCam_2. This tells us that the map $\mathrm{CalCam}_2 \to E$ is two to one, and in fact is a smooth cover:

Theorem 1.2.2. *The map* $\pi\colon \mathrm{CalCam}_2 \to \mathbf{E}^{\mathrm{sm}}$ *is a two to one finite étale cover.*

By further analyzing this cover and applying the theory of linear systems, we discover that simply knowing the calibration curve is enough to recover the entire image of a pair of calibrated cameras, and this gives us an entirely new description of the cover:

Theorem 1.2.3. *The variety* CalCam_2 *is given by the vanishing of a single bilinear form on* $\mathbf{P}^3 \times \mathbf{P}^3$.

This simple description has some key advantages, namely that we can use it to create a new reconstruction algorithm. Instead of a point correspondence inducing a linear constraint on $\mathbf{P}^8$, it can induce a bilinear constraint on $\mathbf{P}^3 \times \mathbf{P}^3$. Thus, with five point correspondences we get a system of six bilinear equations (the five correspondences and the equation defining CalCam_2) that should have twenty solutions. Solving this problem may may have advantages over the system of ten cubics and five linear equations that the standard algorithm relies on.

1.3 A brief outline

In Chapter 2, we develop the foundations of our modern algebraic approach. The goal of this chapter is to reframe the most important definitions and propositions from the literature (such as [8]) in the language of algebraic geometry. For clarity and simplicity, we work over $\mathbf{R}$ or $\mathbf{C}$ in order to give the correct intuition.

In Chapter 3, we take a closer look at three nuances that can occur in the reconstruction process, namely critical configurations, rank one reconstructions, and the question of whether all possible constraints arise from point correspondences. The critical configurations are studied extensively in [7], and we show how our methods yield a new proof of one of their main results which may give clarity to the geometric point of view. Rank one reconstructions are studied in [1] and once again we give a new proof to a main result to illustrate the geometric point of view. There seems to be no complete study of the last topic, but we present a partial answer.

In Chapter 4, we turn our attention to the task of constructing moduli spaces parametrizing camera configurations. To do this, we first need to define cameras and related objects over an arbitrary space rather than just R or C. This allows us to construct the moduli spaces Cam_n and CalCam_n, which parametrize camera configuration and calibrated camera configurations, respectively. We then embed Cam_n into a Hilbert scheme and construct a generalization of Hilbert schemes, called a diagram Hilbert scheme, which embed CalCam_n into. We end the chapter with a discussion of the morphism $\nu_n : \mathrm{CalCam}_n \to \mathrm{Cam}_n$ that forgets calibration data.

In Chapter 5, we restrict our attention to CalCam_2. Specifically, the hard work of formulating moduli-theoretic foundations gives new insight into structure of the Essential variety as well as the 2-1 twisted pair covering. We conclude by exploring some connections to previous work and briefly note how this novel construction leads to a new reconstruction algorithm.

Chapter 2

ALGEBRAIC GEOMETRIC FOUNDATIONS FOR MULTIVIEW GEOMETRY

In this chapter, we begin by presenting foundations for multiview geometry from the perspective of an algebraic geometer. In particular, we define the central objects of study within this thesis and prove basic results about them. We highlight key differences between our perspective and the classical approach when possible. We also aim to give context to algebraic geometers who might be unaware of the multiview geometry literature (although the interested reader is encouraged to read [8] to see a complementary perspective).

2.1 Cameras

Throughout this chapter, we work over the real and complex numbers. Note that in Section 4.1, we define these objects over an arbitrary space.

Definition 2.1.1. A *pinhole camera* is a perspective transformation $P : \mathbf{R}^3 \dashrightarrow \mathbf{R}^2$.

We can describe P as follows: given a point $o \in \mathbf{R}^3$ and a plane $I \subset \mathbf{R}^3$, we define a map

$$\mathbf{R}^3 \setminus \{o\} \to I$$

by sending a point p to the intersection between the line $\overline{po}$ and the plane I. A beautiful algebraic description of such a P goes like this: embed $\mathbf{R}^3$ into $\mathbf{P}^3_{\mathbf{R}}$, with plane at infinity $I_\infty \subset \mathbf{P}^3$ and $\mathbf{R}^2$ into $\mathbf{P}^2_{\mathbf{R}}$ with line at infinity L_∞. (We will omit the R subscript in what follows for readability.) Then there is a unique linear projection $\widetilde{P} : \mathbf{P}^3 \dashrightarrow \mathbf{P}^2$ such that the preimage of L_∞ equals I_∞, the center of $\widetilde{P}$ (i.e., the point at which $\widetilde{P}$ is undefined) equals $o \in \mathbf{R}^3 \subset \mathbf{P}^3_{\mathbf{R}}$, and such that the induced map

$$\widetilde{P}|_{\mathbf{R}^3} : \mathbf{R}^3 \dashrightarrow \mathbf{R}^2$$

equals P. That is, pinhole cameras are just restrictions of linear projections of projective spaces. Because of simplicity of using linear maps and because of the techniques we use, we redefine a camera to be the corresponding map of projective spaces.

Definition 2.1.2. A *camera* is a surjective rational map $\varphi : \mathbf{P}^3 \dashrightarrow \mathbf{P}^2$ given by three linearly independent sections of $\mathscr{O}_{\mathbf{P}^3}(1)$. The *center* of the camera is the unique point $p \in \mathbf{P}^3$ at which φ is undefined.

Alternatively, one can think of a camera as a projective 3×4 matrix with full rank. This is the prevailing approach used in the literature, and so we will try to draw connections to this point of view when appropriate.

Example 2.1.3. The camera defined by $x_0 - x_3, x_2 - 2x_3, x_1 \in \mathscr{O}_{\mathbf{P}^3}(1)$ induces the map $[x_0 : x_1 : x_2 : 1] \mapsto [x_0 - 1 : x_2 - 2 : x_1]$ and the 3×4 matrix it corresponds to is

$$\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & -2 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

The camera center is the common vanishing of these three sections or equivalently the kernel of this matrix: $[1 : 0 : 2 : 1]$.

Definition 2.1.4. In real world situations the camera center will not be on the plane at infinity, and we call such cameras *finite cameras*.

Note that for finite cameras we can always decompose the corresponding matrix as $[A|t]$ where A is a full rank 3×3 matrix. Factoring, we can write it as $A[I| - c]$ where I is the 3×3 identity matrix and $c = (x, y, z)$ is the camera center.

Example 2.1.5. For the camera in the previous example we have

$$\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & -2 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix}.$$

2.1.1 Camera configurations

The study of the relationships between multiple cameras is at the heart of Multiview geometry, and so we now define a multiview configuration:

Definition 2.1.6. A *multiview configuration* is a collection of cameras

$$\varphi_1, \dots, \varphi_n : \mathbf{P}^3 \dashrightarrow \mathbf{P}^2.$$

Notation 2.1.7. We will generally use $\mathbf{\Phi} : \mathbf{P}^3 \dashrightarrow (\mathbf{P}^2)^n$ to denote a multiview configuration, writing $\mathbf{\Phi}_i = \mathrm{pr}_i \circ \mathbf{\Phi}$ for its components when necessary. The *length* of $\mathbf{\Phi}$ is the number of cameras; we will denote it $\mathrm{len}(\mathbf{\Phi})$. Write $\mathrm{Center}(\mathbf{\Phi}) \subset \mathbf{P}^3$ for the tuple of camera centers. Write $\pi : \mathrm{Res}(\mathbf{\Phi}) \to \mathbf{P}^3$ for the blowup of $\mathbf{P}^3$ at the reduced closed subscheme supported at the camera centers; if two cameras have the same center we only count it once. Given an index i, let E_i denote the exceptional divisor over the ith camera center, with canonical inclusion $\iota_i : E_i \hookrightarrow \mathrm{Res}(\mathbf{\Phi})$. By the previous convention, this means that there can be $i \neq j$ for which $E_i = E_j$.

Definition 2.1.8. A multiview configuration $\mathbf{\Phi}$ is *general* if the camera centers are all distinct. It is *non-collinear* if the camera centers do not all lie on a single line, and *collinear* otherwise.

Definition 2.1.9. A *isomorphism* between multiview configurations $\mathbf{\Phi}^1$ and $\mathbf{\Phi}^2$ of common length n is an automorphism $\varepsilon : \mathbf{P}^3 \to \mathbf{P}^3$ fitting into a commutative diagram

$$\begin{tikzcd}
\mathbf{P}^3 \arrow[dd, "\varepsilon"'] \arrow[dr, dashed, "\boldsymbol{\Phi}^1"] & \\
& (\mathbf{P}^2)^n \\
\mathbf{P}^3 \arrow[ur, dashed, "\boldsymbol{\Phi}^2"'] &
\end{tikzcd}$$

Lemma 2.1.10. *Let Y be a scheme, and let $(\mathscr{L}, s_0, \dots, s_n)$ be an invertible sheaf with n sections. If Z is the zero scheme of $s_0, \dots, s_n$ then the rational map induced by this linear series extends uniquely to a morphism $\operatorname{Bl}_Z Y \to \mathbf{P}^n$.*

Proof. By definition the sections $s_0, \dots, s_n$ define a surjection

$$\mathscr{O}_Y^{n+1} \twoheadrightarrow \mathscr{L} \otimes \mathscr{I}_Z,$$

which extends to a surjective map of $\mathscr{O}_Y$-algebras

$$\operatorname{Sym}^*(\mathscr{L}^\vee)^{\oplus n+1} \twoheadrightarrow \bigoplus \mathscr{I}^n.$$

The induced map on relative **Proj** constructions gives the desired morphism. □

Proposition 2.1.11. *Given a multiview configuration $\boldsymbol{\Phi}$, there is a unique commutative diagram*

$$\begin{tikzcd}
\operatorname{Res}(\boldsymbol{\Phi}) \arrow[dr, "\rho"] & \\
& (\mathbf{P}^2)^{\operatorname{len}(\boldsymbol{\Phi})} \\
\mathbf{P}^3. \arrow[uu, dashed, "\pi^{-1}"] \arrow[ur, dashed, "\boldsymbol{\Phi}"'] &
\end{tikzcd}$$

The diagram has the property that for each i, the composition

$$E_i \xrightarrow{\iota_i} \operatorname{Res}(\boldsymbol{\Phi}) \xrightarrow{\rho} (\mathbf{P}^2)^{\operatorname{len}(\Phi)} \xrightarrow{\operatorname{pr}_i} \mathbf{P}^2$$

is an isomorphism.

Proof. Lemma 2.1.10 shows the existence and uniqueness of the desired diagram. To check that the composition is an isomorphism on exceptional divisors one can see that each map is locally isomorphic to the morphism $\mathrm{Bl}_0 \mathbf{A}^3 \to \mathbf{P}^2$ that resolves the canonical presentation $\mathbf{A}^3 \smallsetminus \{0\} \to \mathbf{P}^2$, and here one can simply check that the induced map from the exceptional divisor to the plane is an isomorphism. We omit the details. □

2.2 Reconstruction

We now discuss the reconstruction problem for more than two cameras.

Definition 2.2.1. A *point correspondence in n views* is a point $(p_i) \in (\mathbf{P}^2)^n$.

The problem of reconstruction can be phrased like this:

Problem 2.2.2. *Given m point correspondences in n views $(p_i)_1, \dots, (p_i)_m \in (\mathbf{P}^2)^n$, find n cameras $P_i : \mathbf{P}^3 \dashrightarrow \mathbf{P}^2$ and m world points $\xi_j \in \mathbf{P}^3$ such that $P_i(\xi_j) = (p_j)_i$ for all i and j.*

There is a standard way to further convert this problem into linear algebra. Endowing projective space with homogeneous coordinates, we can represent a camera with a 3×4-matrix $P \in \mathbf{R}^{3\times 4}$, up to scaling. In other words, two matrices P and P' represent the same camera if and only if they differ by multiplication by a non-zero real number. A point correspondence in n views corresponds to a tuple (p_i) of vectors in $\mathbf{R}^3$, and a world point is an element $\xi \in \mathbf{R}^4$. The problem can then be rephrased as follows.

Problem 2.2.3 (Rephrased in linear algebra)**.** *Given m tuples $(p_i)_j \in (\mathbf{R}^3)^n$, find matrices $P_1, \dots, P_n \in \mathbf{R}^{3\times 4}$ and vectors $\xi_1, \dots, \xi_m \in \mathbf{R}^4$ such that for all i, j we have $P_i \xi_j = \lambda_{ij} (p_j)_i$ for some non-zero scalars $\lambda_{ij} \in \mathbf{R}^\times$.*

If we wish to work purely geometrically, there is still another way to phrase the problem.

Problem 2.2.4 (Rephrased geometrically)**.** *Given m points $\alpha_1, \dots \alpha_m \in (\mathbf{R}^2)^n$, characterize all maps $\varphi : \mathbf{R}^3 \dashrightarrow (\mathbf{R}^2)^n$ such that*

1. *the components* $\mathrm{pr}_i \circ \varphi$ *are pinhole cameras, and*
2. *each* α_i *is in the image of* φ*.*

This will be especially useful in what follows, as we will begin to illustrate in Section 2.2.1.

Remark 2.2.5. There are a few natural questions to ask about this situation. For example, for a given number n of views, how many correspondences will yield only finitely many reconstructions? (This is an example of a *minimal problem* [13, 14, 16].) Once we know this, how many solutions do we expect? This is related to the algorithms one might use to solve the problem in applications, since if we know there are only a few solutions, we might hope to solve for them analytically, whereas a large number of expected solutions will force the use of numerical methods.

2.2.1 The case of two cameras

It is especially illuminating to consider the reconstruction problem for two views, since the geometry becomes simple. (On the other hand, as we will describe below, the geometry of two views is also misleading in certain ways.) Since we will ultimately prove everything here in greater generality elsewhere in this paper, we freely omit or only sketch proofs.

First consider the geometric formulation of reconstruction problem. We have m points $\alpha_1, \dots, \alpha_n$ of $\mathbf{P}^2 \times \mathbf{P}^2$, and we wish to find a pair of cameras $\mathbf{P}^3 \dashrightarrow \mathbf{P}^2 \times \mathbf{P}^2$ that contains each α_i in its image. This is a subtle problem, but we can first try to solve a compactified form of the problem. We can replace the image of $\mathbf{P}^3 \dashrightarrow \mathbf{P}^2 \times \mathbf{P}^2$ with its closure, which is now a divisor in $\mathbf{P}^2 \times \mathbf{P}^2$ (assuming that centers of the two cameras are distinct). This divisor is called the *multiview variety* associated to the camera pair.

Lemma 2.2.6. *Given a pair of linear projections* $P_1, P_2 : \mathbf{P}^3 \dashrightarrow \mathbf{P}^2$ *with distinct centers, the closure of* $\mathrm{im}(P_1 \times P_2)$ *is a divisor in the linear system* $|\mathscr{O}_{\mathbf{P}^2 \times \mathbf{P}^2}(1,1)|$*. (That is, it is* *given by a bilinear form in the homogeneous coordinates.)*

Proof. One can use intersection theory and the fact that the components are linear projections. □

Which divisors arise in this way? One immediately notices two things: (1) the divisor in question must be singular, since the line connecting the camera centers gets contracted by $P_1 \times P_2$; (2) a general member of $|\mathscr{O}(1,1)|$ is smooth (by Bertini's theorem). Thus, the divisors that occur from pairs of cameras are a special locus in $\mathbf{P}^8 \cong |\mathscr{O}(1,1)|$. What locus is this?

Viewing elements of $|\mathscr{O}(1,1)|$ as divisors associated to bilinear forms in the homogeneous coordinates, one can realize the space of such divisors as 3×3 matrices modulo scalar multiplication. More precisely, if X_0, X_1, X_2 are coordinates on the first copy of $\mathbf{P}^2$ and Y_0, Y_1, Y_2 are coordinates on the second copy, we can realize an element of $|\mathscr{O}(1,1)|$ by choosing a 3×3-matrix A and considering the equation

$$(X_0, X_1, X_2)A(Y_0, Y_1, Y_2)^T = 0.$$

Proposition 2.2.7. *The multiview varities are precisely the divisors in* $|\mathscr{O}(1,1)|$ *corresponding to matrices of rank* 2*, and these are precisely the divisors whose singular locus consists of a single closed point.*

Proof. Let us give a brief geometric explanation of why A must have rank 2. If $o \in \mathbf{P}^3$ is the center of the first camera P_1, then we can resolve the rational map P_1 to a morphism $\mathrm{Bl}_o \to \mathbf{P}^2$, and the exceptional fiber E surjects into the first image plane. On the other hand, the second camera sends o to some point $b = (b_0 : b_1 : b_2) \in \mathbf{P}^2$. In matrix terms, this says that for every (a_0, a_1, a_2) we have

$$(a_0, a_1, a_2)A(b_0, b_1, b_2)^T = 0,$$

which says precisely that $(b_0, b_1, b_2)^T$ is in the right kernel of A. □

Example 2.2.8. The matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

corresponds to the divisor $x_0y_0 + x_1y_1$ which is singular at the point $[0:0:1] \times [0:0:1]$.

It will be useful later to understand the rank one loci as well.

Proposition 2.2.9. *The divisors in* $|\mathscr{O}(1,1)|$ *that correspond to matrices of rank one are isomorphic to* $L_1 \times \mathbf{P}^2 \cup \mathbf{P}^2 \times L_2$ *where* L_1, L_2 *are lines in* $\mathbf{P}^2$.

Proof. A rank one matrix can be decomposed as the product of two vectors $(a_0, a_1, a_2)^T(b_0, b_1, b_2)$. The corresponding variety is defined by $(a_0x_0 + a_1x_1 + a_2x_2)(b_0y_0 + b_1y_1 + b_2y_2)$ which reduces into two components of the desired form. □

Example 2.2.10. The matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

corresponds to the divisor x_0y_0 which reduces into the two components $V(x_0)$ and $V(y_0)$.

A final observation. Historically, the reconstruction problem we are considering is only asked *up to projective equivalence*, that is, up to an automorphism of $\mathbf{P}^3$. This is simply asking that we perform the reconstruction up to a change of homogeneous coordinates, which have been imposed from without (since there is not intrinsic coordinate frame on the world). It turns out that the notion of joint image interacts with projective equivalence especially nicely.

Lemma 2.2.11. *Two camera configurations in two views* $(P_1, P_2) : \mathbf{P}^3 \dashrightarrow (\mathbf{P}^2)^2$ *and* $(\widetilde{P}_1, \widetilde{P}_2) : \mathbf{P}^3 \dashrightarrow (\mathbf{P}^2)^2$ *are projectively equivalent if and only if their joint images are equal (as closed subschemes).*

Proof. We include a proof over C in 2.4 and general proof in 4.1.11 below. The interesting direction is the implication that equal images yield a projective equivalence, since we only a priori get a birational isomorphism that conjugates the configurations. Regularity of this birational map ultimately follows from the linearity of the projections and a brief analysis of the implications for the equations of the birational map. □

Finally, we come to the reconstruction problem. We have the space of all multiview varieties sitting inside the determinantal locus of $\mathbf{P}^8 = \mathbf{P}(\mathrm{M}_3)$. What constraints are imposed by a point correspondence?

Lemma 2.2.12. *A point correspondence* $(p_1, p_2) \in (\mathbf{P}^2)^2$ *determines a hyperplane in* $\mathbf{P}^8 = |\mathscr{O}(1,1)|$.

Proof. This follows from the fact that an element D of $|\mathscr{O}(1,1)|$ contains (p_1, p_2) if and only if a defining equation of D vanishes at (p_1, p_2), which means that the locus of divisors containing (p_1, p_2) is the kernel of an evaluation map, making it a linear subspace of $\mathbf{P}^8$ of codimension 1, as desired. □

Since we know that the determinantal locus is a cubic hypersurface (the matrices in question being 3×3), we immediately see that seven general point correspondences will collectively result in three complex solutions to the structure from motion problem. This has one nice feature: cubics have an analytic solution, so we can actually analytically produce all (complex) candidates for the solution from the list of correspondences. (Algorithm: write down the pencil in coordinates – a linear algebra problem – and then solve the cubic induced by the determinant map. Given the matrix, one can manually produce the transform comparing the two cameras, as in [8, Section 9.5.3].)

Summary 2.2.13. Let us briefly summarize the main results in two views.

1. The multiview variety of a pair of cameras $P_1, P_2 : \mathbf{P}^3 \dashrightarrow \mathbf{P}^2$ is a closed subscheme of $\mathbf{P}^2 \times \mathbf{P}^2$ that is cut out by a single bilinear form in the homogeneous coordinates on the factors.

2. The multiview variety of (P_1, P_2) uniquely characterizes the pair up to projective equivalence.

3. The space of all joint images is open in the determinantal locus $(\det(A) = 0) \subset \mathbf{P}^8$.

4. A single point correspondence gives a hyperplane in $\mathbf{P}^8$ and thus seven general point correspondences define a line, and this line will intersect the locus of joint images in three (complex) points, at least one of which must be real. (The situation for *non-general* point correspondences is quite interesting. See [1] and Chapter 3 for interesting examples showing that things can be badly behaved.)

Remark 2.2.14. We can interpret the $\mathbf{P}^8 = |\mathscr{O}(1,1)|$ as a component of the Hilbert scheme of $\mathbf{P}^2 \times \mathbf{P}^2$, and we see that the space of camera configurations up to projective equivalences is a locally closed subscheme of the Hilbert scheme. In more views this situation gets considerably more interesting. This is one of the beautiful observations of [3]. Since the methods of [ibid.] are deeply computational, it was in our effort to understand them geometrically that we were led to the present effort to recast the classical theory in functorial algebraic geometry. As we discovered, one can exploit the resulting deformation theory of camera configurations to give a geometric proof of a refinement and generalization of this key result of [ibid.], and extend the theory to the calibrated realm.

2.3 Calibrated cameras

There is a nagging problem: we have always been working up to projective equivalence. If one looks at the kinds of distortions of real images one can produce using projective transformations (some examples comparing different kinds of reconstruction can be found in Sections 10.2, 10.3, and 10.4 of [8]), one is tempted to require not simply a projective solution to the structure from motion problem, but rather a *Euclidean* solution.

This is a function of the way in which the world (domain) and image plane are coordinatized. In other words, given a plane $I \subset \mathbf{R}^3$ in space and a camera center $o \in \mathbf{R}^3$, we can produce a canonical projection $\mathbf{R}^3 \dashrightarrow I$ with center o by using the description after Definition 2.1.1. Moreover, we can give I canonical coordinates using the metric structure on $\mathbf{R}^3$: the center of the coordinate system is the point of I closest to o (the so-called *principal point*), and the x- and y-axes are chosen to be perpendicular lines spanned by unit vectors in I. With these choices, we can write the transformation as $(x, y, z) \mapsto (fx/z, fy/z)$, where f is the distance from o to I (the "focal length").

If one imagines taking a photograph of Seattle and comparing it to a photograph of a perfect diorama of Seattle, one can see that precomposing any camera with a similarity transformation (a composition of translations, orthogonal transformations, and scaling) will leave the intrinsic geometry of the camera intact. That is, the rays back-projected from two points will have the same angles before and after the similarity transformation, the camera will have the same focal length, and so on.

Remark 2.3.1. Our terminology in this section differs from the terminology in the literature. We will use the word "normalized" in place of what is commonly called "calibrated". We reserve the word "calibrated for a more general notion that we introduce below. Although this rewording can be confusing we believe it is more appropriate and leads to a clearer understanding of the subject.

Definition 2.3.2. A pinhole camera $P : \mathbf{R}^3 \dashrightarrow \mathbf{R}^2$ is *normalized* if it differs from a canonical projection $(x, y, z) \mapsto (fx/z, fy/z)$ by a similarity.

It turns out that there is a beautiful way of understanding when the *homogeneous coordinates* on $\mathbf{P}^3$ and $\mathbf{P}^2$ can be chosen to make a linear projection $\mathbf{P}^3 \dashrightarrow \mathbf{P}^2$ compatible with the metric structures up to similarity. It requires us to keep track of two key conics determined by the Euclidean metric: the conic $x^2 + y^2 + z^2 = 0$ in the plane at infinity $w = 0$, the so-called "absolute conic" which we denote Q_∞, and something we will call the "Euclidean conic" $x^2 + y^2 + z^2 = 0$ in $\mathbf{P}^2$, which we denote D. (We call it the Euclidean conic because it arises from the Euclidean metric on $\mathbf{R}^3$.)

Lemma 2.3.3. *A camera* $P : \mathbf{P}^3 \dashrightarrow \mathbf{P}^2$ *is normalized if and only if it takes the absolute conic isomorphically to the Euclidean conic.*

Proof. Any linear form on $\mathbf{P}^3$ that vanishes on the absolute conic is a multiple of the equation of the plane at infinity, and thus any such form is uniquely determined by its value at one additional point not on that plane. It follows that any camera $\mathbf{P}^3 \dashrightarrow \mathbf{P}^2$ (whose center does not lie on the plane at infinity) is uniquely determined by its center and its restriction to the absolute conic. Moreover, any automorphism of the absolute conic extends to an automorphism of $\mathbf{P}^3$, and we can simultaneously act on a conic and swap two points that lie in the complement of the plane spanned by a conic. Thus, any camera that sends the absolute conic to the circle is projectively equivalent to the camera centered at $(0:0:0:1)$ for which the induced map of conics is the identity map. This camera is precisely $(x, y, z) \mapsto (x/z, y/z)$ in affine coordinates. $\square$

Remark 2.3.4. The proof of Lemma 2.3.3 does not use anything about the base field. If one restricts to the real numbers, one can recover a more classical description of normalized cameras. A real projective camera $P : \mathbf{P}^3_{\mathbf{R}} \dashrightarrow \mathbf{P}^2_{\mathbf{R}}$ can be described by a matrix $A \in \mathrm{M}_{3\times 4}(\mathbf{R})$. Assuming that the camera center does not lie on the plane at infinity, we can write the matrix A as

$$A = [M| - MC],$$

where M is an invertible 3×3-matrix and C is the vector of affine coordinates of the camera center. The RQ-factorization yields an expression

$$A = K[R| - RC],$$

where

$$K = \begin{pmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{pmatrix}$$

with $\alpha_x, \alpha_y > 0$ and $R \in \mathrm{O}(3)$ is an orthogonal matrix. The matrix K is called the "calibration matrix" and measures the discrepancy between the camera image coordinate system and the orthonormal coordinate system induced by the Euclidean structure of the ambient space (containing the image plane). For example, the camera output may be in pixel coordinates on a CCD, and there is no reason for the pixels to be square, or for the coordinate system to have the principal point as its origin. (This is described beautifully in Chapter 2 of [8].)

Using this matrix formulation, the proof of Lemma 2.3.3 for real cameras comes down to the statement that the bilinear form associated to the image of the absolute conic is $(KK^T)^{-1}$, and that this also gives the unique Cholesky factorization of that matrix, thus showing that K uniquely determines the image of the absolute conic. (See [8, Section 8.5.1] for more details.)

Since normalized cameras are subject to additional constraints, one expects their multiview variety to lie in a smaller subspace of $\mathbf{P}^8$. And, indeed, this is the case. There is a five-dimensional subvariety of the rank 2 locus, called the *essential variety*, that corresponds to pairs of normalized cameras. The matrices that parametrize the corresponding bilinear forms are called *essential matrices*, and they are charmingly characterized by the property that their two non-zero singular values are equal [8, Section 9.6.1].

2.3.1 Cameras calibrated with respect to arbitrary conics

In this thesis, we avoid this classical description of cameras over $\mathbf{R}$ in terms of matrix factorization and instead use Lemma $2.3.3$ to redefine calibration. The particular equations defining the "absolute conic" and "euclidean conic" do not play a role in the proof of Lemma 2.3.3. It is possible to substitute any pair of fixed conics in their place and this leads to a more flexible theory.

There are three main reasons to use this more flexible approach:

(1) it leads to the "right definition" of the moduli space of calibrated camera configurations (Chapter 4);

(2) by always forcing the absolute conic to map to the Euclidean conic, one makes it impossible to study modular boundary points where the absolute conic is flattened until it collapses (yielding degenerate calibrations). As we will describe below, these degenerate calibrations give geometrically meaningful compactifications of the space of calibrated camera configurations.

(3) In 5 we will use this geometric definition to great advantage in giving a new description of the moduli of pairs of calibrated cameras.

Definition 2.3.5. A *calibrated plane* is a pair $(\mathbf{P}^2, D)$ with D a smooth conic.

Definition 2.3.6. A *calibration datum* for a pinhole camera φ is a pair of degree 2 curves $C \subset \mathbf{P}^3$ and $D \subset \mathbf{P}^2$ such that

1. D is a smooth conic;

2. φ is regular along C;

3. φ_C factors through D.

If C is smooth, the calibration datum will be called *smooth*; otherwise it will be called *degenerate*. If a calibrated plane $(\mathbf{P}^2, D)$ is fixed, a *relative calibration datum* for a pinhole camera $\mathbf{\Phi}$ is a curve $C \subset \mathbf{P}^3$ such that (C, D) is a calibration datum for $\mathbf{\Phi}$.

Remark 2.3.7. If C is smooth then it follows from the linearity of the camera projection that $\mathbf{\Phi}$ must map C isomorphically to D, and that the center of $\mathbf{\Phi}$ is not contained in the plane spanned by C. If C is degenerate, it must be a divisor-theoretic sum of two lines on the quadric cone in $\mathbf{P}^3$ generated by D under the projection $\mathbf{\Phi}$ (i.e., a union of two distinct rulings or a double ruling). As we will see below, a union of two distinct rulings cannot occur as limit of calibration data

Remark 2.3.8. A given camera with calibrated image plane $(\mathbf{P}^2, D)$ has infinitely many relative calibration data: one can take any plane section of the quadric cone in $\mathbf{P}^3$ lying over D. Once we look at configurations of two or more cameras, there will be at most two calibration data (smooth or degenerate). This is described at length in Chapter 5.

Definition 2.3.9. A *calibrated camera* is a pair $(\varphi, (C, D))$ where φ is a pinhole camera and (C, D) is a calibration datum for φ.

2.3.2 Calibrated camera configurations

When the cameras are adorned with calibration data, we track these data through the diagrams.

Definition 2.3.10. Given a multiview configuration $\mathbf{\Phi} : \mathbf{P}^3 \dashrightarrow (\mathbf{P}^2)^n$, a *multiview calibration datum* is a pair $(C, (C_1, \ldots, C_n))$ such that for each $i = 1, \ldots, n$ the pair (C, C_i) is a calibration datum for $\mathbf{\Phi}_i$. Given a tuple of calibrated planes $(\mathbf{P}^2, C_i)$ for $i = 1, \ldots, n$, a *relative calibration datum* for $\mathbf{\Phi}$ is a curve $C \subset \mathbf{P}^3$ such that $(C, (C_1, \ldots, C_n))$ is a calibration datum for $\mathbf{\Phi}$.

Notation 2.3.11. We will write $\mathbf{C}$ for a calibration datum $(C, (C_i))$, and then $\mathbf{C}_0 = C$ and $\mathbf{C}_i = C_i$ for $i = 1, \ldots, n$.

Notation 2.3.12. A calibrated multiview configuration $(\mathbf{\Phi}, \mathbf{C})$ will be called *non-degenerate* if the calibration datum is non-degenerate.

Definition 2.3.13. An *isomorphism* between multiview configurations with calibration data $(\mathbf{\Phi}^1, \mathrm{C}^1)$ and $(\mathbf{\Phi}^2, \mathrm{C}^2)$ of common length n is an isomorphism $\varepsilon : \mathbf{\Phi}^1 \to \mathbf{\Phi}^2$ of multiview configurations as in Definition 2.1.9 such that $\varepsilon(\mathrm{C}_0^1) = \mathrm{C}_0^2$ and such that for $i = 1, \ldots, n$ we have $\mathrm{C}_i^1 = \mathrm{C}_i^2$.

2.4 A characterization of isomorphic configurations

We now briefly consider when two multiview configurations $\mathbf{\Phi}^1$ and $\mathbf{\Phi}^2$ are isomorphic (and similarly when they are endowed with calibration data). This will play a role in studying a particular map from the moduli space to Hilbert schemes in later sections of this paper.

As we will gradually see, the following lemma is the key result connecting the abstract moduli problems we study here to Hilbert schemes.

Lemma 2.4.1. *The derived adjunction map $\mathscr{O}_{\mathrm{Sch}(\mathbf{\Phi})} \to \mathbf{R}\,\rho_* \mathscr{O}_{\mathrm{Res}(\mathbf{\Phi})}$ is a quasi-isomorphism.*

Proof. This amounts to showing that $\rho^\sharp : \mathscr{O}_{(\mathbf{P}^2)^n} \to \rho_* \mathscr{O}_{\mathrm{Res}(\mathbf{\Phi})}$ is surjective and that all higher direct images $\mathbf{R}^i\,\rho_* \mathscr{O}_{\mathrm{Res}(\mathbf{\Phi})}$ (with $i > 0$) vanish.

For the surjectivity statement, note that $\rho_* \mathscr{O}_{\mathrm{Res}(\mathbf{\Phi})}$ is a finite $\mathscr{O}_{(\mathbf{P}^2)^n}$-algebra by properness. Moreover, since every non-empty fiber of ρ is geometrically integral (it being an intersection of lines, hence either a point or a line), we see that $\rho^\sharp$ is surjective after base change to any point of $(\mathbf{P}^2)^n$. By Nakayama's lemma, $\rho^\sharp$ is surjective.

Now we show that the higher direct images vanish. By the Theorem on Formal Functions, the completion of $\mathbf{R}^i\,\rho_* \mathscr{O}$ at a point p is isomorphic to $\lim \mathrm{H}^i(X_m, \mathscr{O}_{X_m})$, where X_m is the mth infinitesimal neighborhood of the fiber of ρ over p. When the fiber is empty or a point, this vanishes. The only interesting case is the unique singular point that is the image of the strict transform of the line through all camera centers, in the collinear case. Note that $\mathscr{O}_{X_m}$ is filtered by subquotients that are symmetric powers of the ideal sheaf $\mathscr{I}_{X_0}$ restricted to X_0. Given a line L in $\mathbf{P}^3$, we have that $\mathscr{I}_L|_L \cong \mathscr{O}_L(-1)^{\oplus 2}$. For each point on L that we blow up, the ideal sheaf gets twisted by 1 (functions from $\mathbf{P}^3$

vanish to extra order on the strict transform along the intersection with the exceptional divisor). In fact, if we are blowing up n points, we have that $\mathscr{I}_{X_0}|_{X_0} \cong \mathscr{O}_{X_0}(n-1)^{\oplus 2}$. The ℓth symmetric power will be a sum of copies of $\mathscr{O}_{X_0}(\ell(n-1))$. All such sheaves have vanishing H^i for all $i > 0$.

Write $\mathscr{I}_m$ for the ideal sheaf of X_m in $\mathrm{Res}(\mathbf{\Phi})$. Consider the standard exact sequences

$$0 \to \mathscr{I}_{m-1}/\mathscr{I}_m \to \mathscr{O}_{X_m} \to \mathscr{O}_{X_{m-1}} \to 0.$$

The above calculations show inductively that $\mathrm{H}^i(X_n, \mathscr{O}_{X_n}) = 0$ for all $n \geq 0$ and all $i > 0$. This concludes the proof. ☐

Corollary 2.4.2. *If Φ is a non-collinear multiview configuration then the map $\rho : \mathrm{Res}(\Phi) \to (\mathbf{P}^2)^n$ is a closed immersion.*

Proof. By the non-collinearity assumption, the geometric fibers of ρ all have length at most 1. Thus, ρ is proper and quasi-finite, hence finite. Applying Lemma 2.4.1 then shows that ρ is a closed immersion. ☐

Lemma 2.4.3. *Suppose $\varphi_1, \varphi_2 : \mathbf{P}^3 \dashrightarrow \mathbf{P}^2$ are cameras and $\alpha : \mathbf{P}^3 \dashrightarrow \mathbf{P}^3$ is a birational automorphism such that $\varphi_2 = \varphi_1 \circ \alpha$. If α and $\varphi_1 \circ \alpha$ are both regular on an open subset $U \subset \mathbf{P}^3$ whose complement has codimension at least 2 then α extends to a unique regular automorphism $\mathbf{P}^3 \to \mathbf{P}^3$.*

Proof. Removing the center of φ_1 if necessary, we may assume that there is an open subscheme $U \subset \mathbf{P}^3$ on which φ_1, φ_2, and α are all regular and $\mathrm{codim}(\mathbf{P}^3, \mathbf{P}^3 \setminus U) \geq 2$. By assumption, $\varphi_i^* \mathscr{O}(1) = \mathscr{O}_U(1)$. Thus, $\alpha^* \mathscr{O}(1) = \mathscr{O}(1)$. Since $\Gamma(U, \mathscr{O}(1)) = \Gamma(\mathbf{P}^3, \mathscr{O}(1))$, we conclude from the universal property of projective space that the morphism $\alpha : U \to \mathbf{P}^3$ extends to a unique endomorphism $\widetilde{\alpha}$ of $\mathbf{P}^3$. Since α is birational, $\widetilde{\alpha}$ is an isomorphism, as desired. ☐

Proposition 2.4.4. *Two multiview configurations $\mathbf{\Phi}^1$ and $\mathbf{\Phi}^2$ of length n are isomorphic if and only if their associated multiview schemes in $(\mathbf{P}^2)^n$ are equal.*

Proof. Since $\mathbf{\Phi}^i$ is birational onto its image for $i = 1, 2$, we see that if $\mathrm{Sch}(\mathbf{\Phi}^1) = \mathrm{Sch}(\mathbf{\Phi}^2)$ then there is a birational automorphism $\alpha : \mathbf{P}^3 \dashrightarrow \mathbf{P}^3$ such that $\mathbf{\Phi}^2 = \mathbf{\Phi}^1 \circ \alpha$. Moreover, $\mathrm{pr}_1 \circ \mathbf{\Phi}^1$, α, and $\mathrm{pr}_1 \circ \mathbf{\Phi}^1 \circ \alpha$ are all regular on the open subscheme of $\mathbf{P}^3$ that is the complement of the line joining the centers of $\mathbf{\Phi}^1$ (as this maps isomorphically to the smooth locus of $\mathrm{Sch}(\mathbf{\Phi}^1)$). Applying Lemma 2.4.3, we see that α is regular, as desired. $\square$

Definition 2.4.5. Given a calibrated multiview configuration $(\mathbf{\Phi}, C)$ with calibrated image planes $(\mathbf{P}^2, C_i)$, $i = 1, \ldots, n$, the *associated multiview flag*, denoted $\mathrm{Flag}(\mathbf{\Phi}, C)$, is the flag $C \subset \mathrm{Sch}(\mathbf{\Phi})$ contained in $C_1 \times \cdots \times C_n \subset (\mathbf{P}^2)^n$.

Corollary 2.4.6. *Two calibrated multiview configurations* $(\mathbf{\Phi}^1, C^1)$ *and* $(\mathbf{\Phi}^2, C^2)$ *of length* n *are isomorphic if and only if their associated multiview flags are equal.*

Chapter 3

NUANCES IN RECONSTRUCTION

To illustrate the utility of the foundations we have developed so far, we take a moment to explore some nuanced phenomenon that can arise during reconstruction. The first two topics discussed in this chapter only occur in non-generic situations and so are not expected to interfere with standard reconstruction algorithms. However, with noise in data, it can be important to study these "critical" loci of camera configurations since they may create unstable conditions for points nearby. From a mathematical point of view, these situations are intrinsically interesting cases that deserve consideration for their own sake.

3.1 Critical Configurations

We first consider the reconstruction problem for a pair of cameras as discussed in 2.2.1. Indeed, in our initial inspection in Chapter 2, we concluded that a generic seven point correspondence yields a unique reconstruction. However, in certain non-generic circumstances, it is possible for there to exist multiple reconstructions that are not protectively isomorphic.

Definition 3.1.1. The data of a camera configuration Φ and world points z_i is called a *critical configuration* if there exists a non-isomorphic camera configuration Φ' and world points z_i' such that $\Phi(z_i) = \Phi'(z_i')$ for all i. The pairs (Φ, z_i) and (Φ', z_i') are called *conjugate configurations*.

In [7], the question of when critical configurations occur is answered. The following is a summary of a surprising result concerning which configurations are critical.

Proposition ([7], Lemma 5.10). *Let Φ be a camera configuration of length two. If (Φ, z_i) is a critical configuration, than the world points z_i lie on a quadric.*

The proof given by [7] uses linear algebra to construct the quadric and verify the claim. Here we give a geometric proof.

Proof. Let (Φ', z'_i) be a conjugate configuration. Let s and s' denote the sections of $\mathcal{O}(1,1)$ corresponding the images of Φ and Φ' respectivly. The image points $(x_i, y_i) = \Phi(z_i) = \Phi'(z'_i)$ lie on the vanishing of both s and s'. Since Φ pulls back $\mathcal{O}(1,1)$ to $\mathcal{O}_{\mathbf{P}^3}(2)$ and $s \neq s'$ we see that it pulls back s' to a quadric in $\mathbf{P}^3$. Since all the image points (x_i, y_i) lie on s', their preimages, the world points z_i, must also lie on this quadric. $\square$

3.2 Rank One Reconstructions

Recall that the set of multivew varieties is a locally closed variety of rank two matrices in $|\mathcal{O}_{\mathbf{P}^2 \times \mathbf{P}^2}(1,1)$. The locus of rank one matrices is a codimension four subvariety, contained in the closure of the rank two variety.

It is possible for the line given by seven point correspondences to miss the rank two locus entirely and only intersect the rank one locus. In this case, unique reconstruction fails because no valid camera configurations match the point correspondence. In [1]. a criterion on the point correspondences is given to determine when this occurs; we give a rephrased version of their main result here:

Theorem 3.2.1 ([1], Theorem 2.2). *For a seven point correspondence (x_i, y_i), all of the sections of $\mathcal{O}_{\mathbf{P}^2 \times \mathbf{P}^2}(1,1)$ vanishing at these points lie in the rank one locus if and only if one of the following holds:*

(1) There is a nonempty proper subset $\tau \subset \{1, \ldots, 7\}$ such that as points in $\mathbf{R}^2$, $\{x_i : i \in \tau\}$ are collinear and $y_i = y_j$ for all $i, j \notin \tau$.

(2) There is a nonempty proper subset $\tau \subset \{1, \ldots, 7\}$ such that as points in $\mathbf{R}^2$, $\{y_i : i \in \tau\}$ are collinear and $x_i = x_j$ for all $i, j \notin \tau$.

We give a new geometric proof to explain this condition.

Proof. A seven point correspondence yields a line in the linear system $|\mathscr{O}(1,1)|$. This line is a pencil of section of $|\mathscr{O}(1,1)|$. The point correspondences must lie in the intersection of this pencil. If two or more of those reconstruction are rank one sections we can write every section in the pencil as a linear combination of those sections. A rank one divisor is of the form $\mathbf{P}^2 \times L_1 \cup L_2 \times \mathbf{P}^2$ by 2.2.9 so every section in the pencil is of the form

$$\alpha(\mathbf{P}^2 \times L_1 \cup L_2 \times \mathbf{P}^2) + \beta(\mathbf{P}^2 \times L_1' \cup L_2' \times \mathbf{P}^2).$$

Because these are distinct sections of $\mathscr{O}(1,1)$ we know that either $L_1 \neq L_1'$ or $L_2 \neq L_2'$. Without loss of generality we assume it is the former so that the intersection must be contained in $\mathbf{P}^2 \times q_1 \cup L_2 \times \mathbf{P}^2$ where q_1 is the intersection $L_1 \cap L_1'$. We divide the point correspondences into two groups, those in the component $\mathbf{P}^2 \times q_1$ and those in the component $L_2 \times \mathbf{P}^2$. In the first group, all the y-coordinates are equal. In the second group, all the x-coordinates are collinear. This is exactly condition (1). A symmetric argument shows that if $L_2 \neq L_2'$ then condition (2) will hold. Conversely, if conditions (1) or (2) hold then we can find such a pencil of quadrics. $\square$

3.3 Constraints induced by point correspondences

A point correspondence of k points induces a codimension k plane in the linear system $|\mathscr{O}_{\mathbf{P}^2\times\mathbf{P}^2}(1,1)|$. A natural question arises:

Question 3.3.1. *Which codimension k-planes of $|\mathscr{O}_{\mathbf{P}^2\times\mathbf{P}^2}(1,1)|$ arise from a point correspondence of k points?*

The natural space to parametrize codimension k planes is the Grassmannian, $Gr(k,n)$, the space of codimension k planes of $\mathbf{P}^n$. A more precise version of our question is

Question 3.3.2. *Describe the image of the map*

$$(\mathbf{P}^2 \times \mathbf{P}^2)^k \dashrightarrow Gr(k, 8)$$

sending a k point correspondence to its corresponding codimension k plane.

In this section, we give a partial answer to these questions.

3.3.1 Codimension one planes

To begin we consider which hyperplanes arise from a single point correspondence. Given a point $p \times q$ the corresponding hyperplane is given by flattening the matrix pMq^T. In other words it is the map

$$\mathbf{P}^2 \times \mathbf{P}^2 \to Gr(1, 8) \simeq \mathbf{P}^8 \simeq |\mathscr{O}_{\mathbf{P}^2 \times \mathbf{P}^2}(1, 1)|$$

given by the Segre embedding

$$[p_0 : p_1 : p_2] \times [q_0 : q_1 : q_2] \to [p_0q_0 : p_1q_0 : p_2q_0 : p_0q_1 : p_1q_1 : p_2q_1 : p_0q_2 : p_1q_2 : p_2q_2].$$

The image is the Segre variety, a degree six, codimension four subvariety of $\mathbf{P}^8$.

3.3.2 Higher codimension

For higher codimension planes, we consider the intersection of hyperplanes arising from single point correspondences. Each of these hyperplanes corresponds to a point on $Gr(1, 8)$ via identifying $Gr(1, 8)$ as the projective dual of $\mathbf{P}^8$. The intersection of two such planes corresponds to the line connecting to two points in the projective dual. In general, the intersection of k hyperplanes corresponds to a $k - 1$ plane in $Gr(1, 8)$ spanned by the k points corresponding to those hyperplanes. Thus, we seek to describe the set of all $k - 1$ planes spanned by k points on the Segre variety.

3.3.3 Grassmannian of secant varieties

This problem is studied in in [4]. Following [4], we work over C and let $X \subseteq \mathbb{P}^r$ be an integral, non-degenerate variety of dimension n. For $k \leq r$, a general tuple of $k+1$ points in X span a k plane in $\mathbb{P}^r$. Consider the rational map

$$\begin{aligned}\Phi : & X^{k+1} \dashrightarrow G(k,r) \\ & (x_0,\dots,x_k) \mapsto \mathrm{Span}\{x_0,\dots,x_k\}\end{aligned}$$

where $G(k,r)$ is the sets of k-planes in $\mathbb{P}^r$. Define $G_k(X) = \overline{\Phi(X^{k+1})}$, i.e., the smallest subvariety of the Grassmannian containing the k planes spanned by $k+1$ points of X.

Proposition 3.3.3 (Proposition 1.1 [4])**.** $\dim G_k(X) = \min\{n(k+1),(r-k)(k+1)\}$

Note that $(r-k)(k+1)$ is the dimension of $Gr(k,r)$.

This result allows us to compute dimension of $G_k(\mathbb{P}^2 \times \mathbb{P}^2)$ and compare it to the dimension of the Grassmannian in which it lives, $G(k,8)$. Letting $X = \mathbb{P}^2 \times \mathbb{P}^2 \subset \mathbb{P}^8$, we compute the dimension of the space of subspaces spanned by $k+1$ points of the Segre:

k	$\dim G_k(\mathbb{P}^2 \times \mathbb{P}^2)$	$\dim G(k,8)$
0	4	8
1	8	14
2	12	18
3	16	20
4	20	20
5	18	18
6	14	14
7	8	8
8	0	0

From this we see that for correspondences of 5,6 or 7 points we do generically get all possible codimension k-planes as restrictions. It is still an open question to give a complete description of the actual image and not its closure.

Chapter 4

MODULI OF CAMERAS

In this chapter, we explore how one might construct a moduli space of cameras. This first requires a reworking of the foundational definitions of Multiview geometry so that they work over an arbitrary space. We use the language of stacks in order to give the "correct" definitions for these moduli space, but after some deformation theory arguments we find that all of our moduli spaces can be viewed as projective schemes. After proving the existence of Cam_n and CalCam_n, we study their embeddings into Hilbert (or diagram Hilbert) schemes as well as the decalibration morphism $\nu : \mathrm{CalCam}_n \to \mathrm{Cam}_n$.

4.1 Relativization

We begin by generalizing the definitions and results of Chapter 2 to families of cameras over an arbitrary base space. Because our end goal is to construct a moduli space of cameras, we need to define a single pinhole camera in such a way that is compatible with the sheaf condition. In particular, the existence of Brauer-Severi schemes, which are locally isomorphic to $\mathbf{P}^3$ but not globally, lead us to allow more general domains for camera morphisms.

Definition 4.1.1. Given a scheme S, a *relative pinhole camera* over S is a rational map $p : \mathbf{P} \dashrightarrow \mathbf{P}^2_S$ over S uniquely determined by the following information:

1. the scheme $\mathbf{P}$ is a Zariski form of $\mathbf{P}^3_S$;

2. there is a map $\sigma : \mathscr{O}_{\mathbf{P}}^{\oplus 3} \to \mathscr{O}_{\mathbf{P}}(1)$ whose cokernel is an invertible sheaf supported exactly over a section Z of $\mathbf{P} \to S$, called the *camera center*;

3. a representative of p is given by the morphism $\mathbf{P} \smallsetminus Z \to \mathbf{P}^2_S$ determined by the quotient $\sigma_{\mathbf{P}\smallsetminus Z}$ and the universal property of projective space.

Since we assume that there is a section of $\mathbf{P} \to S$, it follows from the basic theory of Brauer-Severi schemes that it is in fact a Zariski form of $\mathbf{P}^3_S$. For reasons of descent theory, we do not make this an *a priori* assumption.

Definition 4.1.2. Given a scheme S, a *relative multiview configuration of length* n over S is given by a proper S-scheme $\mathbf{P} \to S$ of finite presentation and a rational map $\mathbf{\Phi} : \mathbf{P} \dashrightarrow (\mathbf{P}^2_S)^n$ over S such that for each i the composition $\mathrm{pr}_i \circ \mathbf{\Phi}$ is a relative pinhole camera as in Definition 4.1.1.

Two relative multiview configurations

$$\mathbf{\Phi}^i : \mathbf{P}_i \dashrightarrow \mathbf{P}^2_S, \quad i = 1, 2$$

are *isomorphic* if there is an isomorphism $\varepsilon : \mathbf{P}_1 \xrightarrow{\sim} \mathbf{P}_2$ such that $\mathbf{\Phi}^2 = \mathbf{\Phi}^1 \circ \varepsilon$.

Notation 4.1.3. Given a multiview configuration $\mathbf{\Phi} : \mathbf{P} \to (\mathrm{P}^2)^n$ of length n, we will write

1. $S(\Phi)$ for the domain $\mathbf{P}$ of $\mathbf{\Phi}$;
2. $Z_1(\mathbf{\Phi}), \ldots, Z_n(\mathbf{\Phi}) \subset \mathbf{\Phi}$ for the camera centers;
3. $Z(\mathbf{\Phi})$ for the scheme-theoretic union $Z_1(\mathbf{\Phi}) \cup \cdots \cup Z_n(\mathbf{\Phi})$;
4. $\mathrm{Res}(\mathbf{\Phi})$ for the blowup of $S(\mathbf{\Phi})$ in Z.

Definition 4.1.4. A relative multiview configuration $\mathbf{\Phi}$ over S is *general* if the camera centers $Z_1, \ldots, Z_{\mathrm{len}(\mathbf{\Phi})}$ are pairwise disjoint closed subschemes of $\mathbf{P}$.

Definition 4.1.5. A relative multiview configuration $\mathbf{\Phi} : \mathbf{P} \dashrightarrow (\mathbf{P}^2_S)^n$ over S is *collinear* if there is a closed subscheme $\mathbf{L} \subset S(\mathbf{\Phi})$ that is a relative line over S and that contains $Z(\mathbf{\Phi})$. It is *nowhere-collinear* if it is not collinear upon any basechange $S' \to S$.

Definition 4.1.6. Given a relative multiview configuration $\mathbf{\Phi}$ of length n over S, a *(smooth) calibration datum* for $\mathbf{\Phi}$ is a pair $(C, (C_1, \dots, C_n))$ where

1. $C \subset \mathbf{P}$ is a (smooth) degree two curve over S;
2. $C_i \subset \mathbf{P}^2_S$ is a relative smooth conic over S for $i = 1, \dots, n$;
3. some representative of $\mathbf{\Phi}$ is regular along C;
4. and the induced morphisms $(\mathrm{pr}_i \circ \mathbf{\Phi})_C$ factors through C_i for $i = 1, \dots, n$.

If C is smooth, the calibration datum will be called *smooth*; otherwise it will be called *degenerate*.

Proposition 4.1.7. *Given a general relative multiview configuration $\mathbf{\Phi}$ over S, there is a unique commutative diagram*

$$\begin{tikzcd}
\mathrm{Res}(\mathbf{\Phi}) \arrow[dr, "\rho"] & \\
& (\mathbf{P}^2)^{\mathrm{len}(\mathbf{\Phi})} \\
\mathbf{P}. \arrow[uu, dashed, "\pi^{-1}"] \arrow[ur, dashed, "\mathbf{\Phi}"'] &
\end{tikzcd}$$

The diagram has the property that for each i, the composition

$$E_i \xrightarrow{\iota_i} \mathrm{Res}(\mathbf{\Phi}) \xrightarrow{\rho} (\mathbf{P}^2)^{\mathrm{len}(\Phi)} \xrightarrow{\mathrm{pr}_i} \mathbf{P}^2$$

is an isomorphism. Moreover, this diagram is compatible with arbitrary base change on S.

Proof. The arrow ρ exists again by Lemma 2.1.10, and the functoriality follows from the functoriality of Lemma 2.1.10 and the flatness of everything over S. Finally, the isomorphism condition can be checked on geometric fibers, which reduces it to Proposition 2.1.11. ☐

Write $\mathrm{MVC}_n(S)$ for the groupoid of general relative multiview configurations of length n over S. Write $\mathrm{RVC}_n(S)$ for the groupoid of tuples $(\mathbf{P}, (Z_1, \ldots, Z_n), f)$ where $\pi : \mathbf{P} \to S$ is an fppf form of $\mathbf{P}^3 \to S$, the $Z_i \in \mathbf{P}(S)$ are pairwise non-intersecting sections of π, and $f : \widetilde{\mathbf{P}} \to (\mathbf{P}^2)^n$ is a morphism from the blowup of $\mathbf{P}$ along $\cup Z_i$ to $(\mathbf{P}^2_S)^n$ such that $\mathrm{pr}_i \circ f$ induces an isomorphism from the ith exceptional divisor $E_i \subset \widetilde{\mathbf{P}}$ to $\mathbf{P}^2_S$.

Corollary 4.1.8. *Proposition 4.1.7 defines a canonical equivalence of categories* $\mathrm{MVC}_n(S) \to \mathrm{RVC}_n(S)$.

Proof. The proof is tautological. (This corollary mainly serves to establish notation.) □

Definition 4.1.9. Given a general multiview configuration $\mathbf{\Phi}$ of length n, the image of the morphism ρ described in Lemma 4.1.7 is the *multiview scheme* of $\mathbf{\Phi}$.

Notation 4.1.10. The multiview scheme of $\mathbf{\Phi}$ will be denoted $\mathrm{Sch}(\mathbf{\Phi})$.

Proposition 4.1.11. *Two general multiview configurations* $\mathbf{\Phi}^1, \mathbf{\Phi}^2$ *of length* n *over* S *are isomorphic if and only if* $\mathrm{Sch}(\mathbf{\Phi}^1) = \mathrm{Sch}(\mathbf{\Phi}^2)$ *as closed subschemes of* $(\mathbf{P}^2_S)^n$.

The proof of Proposition 4.1.11 is a modification of that of Proposition 2.4.4. We require a modification of Lemma 2.4.3.

Lemma 4.1.12. *Suppose* A *is a ring and* $U \subset \mathbf{P}^3_A$ *is an open subset such that for every geometric point* $A \to \kappa$ *the fiber* $U_\kappa \subset \mathbf{P}^3_\kappa$ *has complement of codimension at least* 2. *Suppose* $\alpha : U \to \mathbf{P}^3_A$ *is a morphism such that* $\alpha^*\mathscr{O}(1) = \mathscr{O}_U(1)$. *Then* α *extends to a unique automorphism of* $\mathbf{P}^3_A$.

Proof. By the universal property of projective space, it suffices to show that restriction defines an isomorphism

$$\Gamma(\mathbf{P}^3_A, \mathscr{O}(1)) \tilde{\to} \Gamma(U, \mathscr{O}(1)).$$

To show this, it suffices to show that the adjunction map $\nu(1) : \mathscr{O}_{\mathbf{P}^3}(1) \to \iota_*\mathscr{O}_U(1)$ is an isomorphism of sheaves. By the projection formula, it suffices to show that the adjunction map for the structure sheaf

$$\nu : \mathscr{O}_{\mathbf{P}^3_A} \to \iota_* \mathscr{O}_U$$

is an isomorphism. But this is precisely Proposition 3.5 of [9]. □

Proposition 4.1.13. *If $\mathbf{\Phi}$ is a general multiview configuration over S then for all base changes $T \to S$ we have that the natural morphism*

$$\mathrm{Sch}(\mathbf{\Phi}) \times_S T \to \mathrm{Sch}(\mathbf{\Phi} \times_S T)$$

is an isomorphism. That is, formation of the asssociated multiview scheme is compatible with base change. Furthermore, $\mathrm{Sch}(\mathbf{\Phi})$ *is flat over the base.*

Proof. By Lemma 2.4.1 the structure morphism $\mathscr{O}_{(\mathbf{P}^2)^n} \to \rho_* \mathscr{O}_{\mathrm{Res}(\mathbf{\Phi})}$ is surjective. Consider the triangle in the derived category

$$I \to \mathscr{O}_{(\mathbf{P}^2)^n} \to \mathbf{R}\,\rho_* \mathscr{O}_{\mathrm{Res}(\mathbf{\Phi})} \xrightarrow{+} .$$

Let $i : (\mathbf{P}^2)^n_q \to (\mathbf{P}^2)^n$ be an embedding of a fiber. Pulling back to the fiber and using cohomology and base change we have

$$\begin{aligned} \mathbf{L}\,i^* \,\mathbf{R}\,\rho_* \mathscr{O}_{\mathrm{Res}(\mathbf{\Phi})} &\simeq \mathbf{R}\,\rho_* \,\mathbf{L}\,i^*_{\mathrm{Res}(\mathbf{\Phi})} \mathscr{O}_{\mathrm{Res}(\mathbf{\Phi})} \\ &\simeq \mathbf{R}\,\rho_* (\mathscr{O}_{\mathrm{Res}(\mathbf{\Phi})})_q \\ &\simeq (\mathscr{O}_{\mathrm{Res}(\mathbf{\Phi})})_q \end{aligned}$$

Applying [10, Lemma 3.31] to $\mathbf{R}\,\rho_* \mathscr{O}_{\mathrm{Res}(\mathbf{\Phi})}$, we see that it is quasi-isomorphic to a sheaf flat over the base. But $\mathscr{H}^0(\mathbf{R}\,\rho_* \mathscr{O}_{\mathrm{Res}(\mathbf{\Phi})})$ is $\rho_* \mathscr{O}_{\mathrm{Res}(\mathbf{\Phi})}$. Thus, we conclude that the short exact sequence

$$0 \to \mathscr{I} \to \mathscr{O}_{(\mathbf{P}^2)^n} \to \rho_* \mathscr{O}_{\mathrm{Res}(\mathbf{\Phi})} \to 0$$

consists of S-flat sheaves and is compatible with arbitrary base change. This establishes the result. □

4.2 Moduli of uncalibrated camera configurations

In this section, we describe the basic moduli problem attached to uncalibrated camera configurations. In Section 4.2.1 we will study the deformation theory of a configuration $\mathbf{\Phi}$, especially as it relates to the deformation theory of the associated scheme $\mathrm{Sch}(\mathbf{\Phi})$. Ultimately this will allow us to embed the moduli space into the Hilbert scheme.

Definition 4.2.1. Given a positive integer n, the *stack of camera configurations of length* n, denoted Cam_n, has as objects over a scheme S the groupoid of general relative multiview configurations of length n.

Proposition 4.2.2. *Let* $\mathrm{M}^n \subset \mathrm{M}^n_{3\times 4}$ *be the locus of* n*-tuples of full rank* 3×4 *matrices. There is an equivalence of stacks between* $[\mathrm{M}^n \,/\, \mathrm{GL}_4]$ *and* Cam_n.

Proof. Let $Q \to T$ be a GL_4 torsor. By [19, Theorem 4.46], there is a canonical equivalence between $\mathrm{Cam}_n(T)$ and $\mathrm{Cam}_n^{\mathrm{GL}_4}(Q)$, the GL_4-equivariant objects of $\mathrm{Cam}_n(T)$. We claim that $[\mathrm{M}^n \,/\, \mathrm{GL}_4](T)$ is equivelant to $\mathrm{Cam}_n^{\mathrm{GL}_4}(Q)$ as well.

An object of $[\mathrm{M}^n \,/\, \mathrm{GL}_4](T)$ is a GL_4-torsor, $Q \to T$, with an equivariant map to M^n. The map $Q \to \mathrm{M}^n$ induces a rational linear map $A : \mathbf{P}^3_Q \dashrightarrow (\mathbf{P}^2_Q)^n \in \mathrm{Cam}_n(Q)$. Let $\alpha : \mathrm{GL}_4 \times Q \to Q$ be the group action and let pr_2 be projection onto Q. Equivariance of the map $Q \to \mathrm{M}^n$ is precisely the statement that the action of GL_4 on $\mathbf{P}^3$ induces an isomorphism $\varphi : \alpha^* A \to \mathrm{pr}_2^* A$ of objects of $\mathrm{Cam}_n(\mathrm{GL}_4 \times Q)$.

$$\begin{tikzcd}
\mathbf{P}^3_{\mathrm{GL}_4 \times Q} \arrow[dd] \arrow[dr, "\alpha^* A"] & \\
& (\mathbf{P}^2_{\mathrm{GL}_4 \times Q})^n \\
\mathbf{P}^3_{\mathrm{GL}_4 \times Q} \arrow[ur, "\mathrm{pr}_2^* A"'] &
\end{tikzcd}$$

Similarly, equivariance gives us a cocycle condition for φ satisfying [19, Proposition 3.49], so A is a GL_4-equivariant object.

An isomorphism of objects Q and Q' of $[\mathrm{M}^n / \mathrm{GL}_4](T)$ is an morphism of torsors commuting with their maps to M^n. Let the corresponding objects of $\mathrm{Cam}_n(Q)$ be A and A'. Then

$$\begin{tikzcd}
\mathrm{pr}_2^* A \arrow[r] \arrow[d] & A \arrow[d] \\
\mathrm{pr}_2^* A' \arrow[r] & A'
\end{tikzcd}$$

commutes, so by [19, Proposition 3.48] this is a GL_4-equivariant morphism.

This defines a morphism of stacks $[\mathrm{M}^n / \mathrm{GL}_4] \to \mathrm{Cam}_n$. We can check this is an equivalence over strictly Hensalian local rings. In this case, every $\mathbf{P}^3$ form is trivial and the description above is an equivalence. □

Corollary 4.2.3. *The stack* Cam_n *is a smooth algebraic space of finite type over* $\mathrm{Spec}\,\mathbf{Z}$*.*

Proof. The space of full rank 3×4 matrices is smooth and has trivial stabilizers. □

Notation 4.2.4. We will write $\mathrm{Cam}_n^{\mathrm{nc}}$ for the locus of non-collinear configurations. This is an open substack.

4.2.1 Deformations of multiview configurations

In this section, we study the relationship between the infinitesimal deformation theory of a camera configuration and the deformation theory of its associated multiview scheme. We get strong results for non-collinear cameras for arbitrary n and for collinear cameras for $n > 4$.

As we will see below, this gives strong results on the relationship between Cam_n and $\mathrm{Hilb}_{(\mathbf{P}^2)^n}$, clarifying and improving the groundbreaking results of [3]. In particular, our infinitesimal analysis will apply at all points, showing density in the Hilbert scheme; special points are handled well for $n > 4$ by a straightforward argument using the cotangent complex, giving the enhancement in those cases. These methods are very different from the ideal-theoretic methods of [3]. It would be especially interesting to understand how the cotangent complex argument of Section 4.2.1 relate to the Gröbner basis calculations in [3].

Definition 4.2.5. Fix a ring A containing an ideal I such that $I^2 = 0$ and let $A_0 = A/I$. Suppose $\mathbf{\Phi}^0$ is a relative multiview configuration of length n over A_0. An *infinitesimal deformation of* $\mathbf{\Phi}^0$ *to* A is a pair $(\mathbf{\Phi}, \varepsilon)$, where $\mathbf{\Phi}$ is a multiview configuration of length n over A and $\varepsilon : \mathbf{\Phi} \otimes_A A_0 \xrightarrow{\sim} \mathbf{\Phi}^0$ is an isomorphism of relative multiview configurations.

An *isomorphism between infinitesimal deformations* $(\mathbf{\Phi}, \varepsilon)$ and $(\mathbf{\Phi}', \varepsilon')$ of $\mathbf{\Phi}^0$ is an isomorphism $\alpha : \mathbf{\Phi} \xrightarrow{\sim} \mathbf{\Phi}'$ of relative multiview configurations such that $\varepsilon' \circ \alpha \otimes_A A_0 = \varepsilon$.

Our goal in this section is to prove the following.

Theorem 4.2.6. *If* $\mathbf{\Phi}$ *is a general multiview configuration of length* n *with associated multiview variety* $V \subset (\mathbf{P}^2)^n$ *then, assuming either that* $\mathbf{\Phi}$ *is non-collinear or that* $n > 4$*, we have that the infinitesimal deformations of* $\mathbf{\Phi}$ *are in bijection with the infinitesimal deformations of* V *as a closed subscheme of* $(\mathbf{P}^2)^n$.

The proof will work roughly as follows.

1. First, we will study the abstract deformations of V as a scheme. As we will see, V has a property that we will call *essential rigidity*.

2. Using this essential rigidity, we will show that any deformation of V as a closed subscheme of $(\mathbf{P}^2)^n$ arises from a deformation of $\mathbf{\Phi}$. In the collinear case this is non-trivial, because $\mathrm{Res}(\Phi) \to (\mathbf{P}^2)^n$ contracts a line. As we will explain, this can be worked around by a mild study of the cotangent complex of the contraction map in the collinear case, as long as $n > 4$.

3. Using Proposition 4.1.11, we have that two deformations of $\mathbf{\Phi}$ give rise to the same deformation of V if and only if they are isomorphic, completing the proof.

It is worth noting (as hinted at in this outline) that the proof we give here is almost purely geometric. We do not rely on dimension estimates, ideal-theoretic calculations, masses of cohomology, etc. The arguments are simple variants of classical Italian geometric arguments, first used to study the geometry of projective surfaces.

Essential rigidity of blowups of $\mathbf{P}^3$

In this section we fix a commutative ring A_0, a square-zero extension

$$I \subset A \to A_0,$$

and a collection of pairwise everywhere-disjoint sections

$$\sigma_i : \operatorname{Spec} A_0 \to \mathbf{P}^3_{A_0}.$$

We write P_0 for the blowup $\mathrm{Bl}_{Z_0} \mathbf{P}^3_{A_0}$, where Z_0 is the reduced closed subscheme of $\mathbf{P}^3_{A_0}$ supported on the union of the images of the σ_i.

Proposition 4.2.7. *Given a deformation P of P_0 over A, there is a unique morphism*

$$\beta : P \to \mathbf{P}^3_A$$

deforming the canonical blow-down map

$$\beta_0 : P_0 \to \mathbf{P}^3_{A_0},$$

up to infinitesimal automorphism of $\mathbf{P}^3_A$. Moreover, β realizes P as the blowup of $\mathbf{P}^3_A$ at a closed subscheme Z that deforms Z_0 (and Z is a union of n sections of $\mathbf{P}^3_A$).

Proof. Via the universal property of projective space, the morphism β_0 is given by the natural map

$$\mathscr{O}_{P_0}^{\oplus 4} \to \beta_0^* \mathscr{O}(1)$$

arising from pulling back the natural map of sheaves

$$\mathscr{O}^{\oplus 4} \to \mathscr{O}(1)$$

on $\mathbf{P}^3_{A_0}$. By the Theorem on Formal Functions, the adjunction map

$$\alpha : \mathscr{O}_{\mathbf{P}^3_{A_0}} \to \mathbf{R}(\beta_0)_* \mathscr{O}_{P_0}$$

is an isomorphism. The deformation theory of $\beta_0^* \mathscr{O}(1)$ to P is governed by the cohomology groups $\mathrm{H}^2(P, \mathscr{O})$ (where obstructions live) and $\mathrm{H}^1(P, \mathscr{O})$ (acting simply transitively on deformations). Since α is an isomorphism, the projection formula shows that these groups are naturally isomorphic to $\mathrm{H}^i(\mathbf{P}^3_{A_0}, \mathscr{O}) \otimes_{A_0} I$, $i = 1, 2$, which vanish by cohomology and base change and the calculation of the cohomology of projective space.

This shows two things: first, that $\beta_0^* \mathscr{O}(1)$ admits a unique deformation L to P, and second that all sections of $\beta_0^* \mathscr{O}(1)$ admit lifts to sections of L over P. There is thus a commutative diagram

$$\begin{tikzcd}
\mathscr{O}_P^{\oplus 4} \arrow[r] \arrow[d] & L \arrow[d] \\
\mathscr{O}_{P_0} \arrow[r] & \beta_0^* \mathscr{O}(1)
\end{tikzcd}$$

of coherent sheaves on P (where the bottom row is the pushforward of the displayed sheaves from P_0), yielding a commutative diagram

$$\begin{tikzcd}
P \arrow[r, "\beta"] & \mathbf{P}^3_A \\
P_0 \arrow[r, "\beta_0"] \arrow[u, hook] & \mathbf{P}^3_{A_0}. \arrow[u, hook]
\end{tikzcd}$$

We claim that this deformed morphism is also a blow-down. One way to see this is the following. For each exceptional divisor $E \subset P_0$, the normal sheaf is $\mathscr{O}_E(-1)$. Cohomology and base change tells us that

$$\mathrm{H}^0(E, \mathscr{O}_E(-1)) = 0 = \mathrm{H}^1(E, \mathscr{O}_E(-1)),$$

which shows that each E_i has a unique deformation to an A-flat divisor in P. The invertible sheaf L is trivial on each E_i, so they are all collapsed under β. More concretely, by Nakayama's Lemma the Stein factorization of

$$\sqcup E_i \to P$$

produces a union of sections $Z \subset \mathbf{P}^3_A$ deforming $Z_0 \subset \mathbf{P}^3_{A_0}$. The pullback of the ideal sheaf of Z to P is precisely the ideal sheaf of $\sqcup E_i$, showing that there is a unique factorization

$$\begin{tikzcd}
P \arrow[r] \arrow[d, "\gamma"'] & \mathbf{P}^3_A \\
\operatorname{Bl}_Z \mathbf{P}^3_A. \arrow[ur] &
\end{tikzcd}$$

The morphism γ becomes an isomorphism over A_0, whence it must be an isomorphism over A by Nakayama's Lemma and the A-flatness of P and $\operatorname{Bl}_Z \mathbf{P}^3_A$. $\square$

Lifting deformation for non-collinear configurations

In this section, we explain how any deformation of a non-collinear multiview scheme lifts to a deformation of the associated multiview configuration. Fix a deformation situation

$$I \subset A \to A_0$$

and a non-collinear multiview configuration $\mathbf{\Phi}^0$ of length n over A_0 with scheme $\operatorname{Sch}(\mathbf{\Phi}^0)$.

Proposition 4.2.8. *If $X \subset (\mathbf{P}^2)^n_A$ is an A-flat deformation of $\operatorname{Sch}(\mathbf{\Phi}^0)$ then there is a deformation $\mathbf{\Phi}$ of $\mathbf{\Phi}^0$ such that $\operatorname{Sch}(\mathbf{\Phi}) = X$ as closed subschemes of $(\mathbf{P}^2)^n$. Moreover, $\mathbf{\Phi}$ is unique up to unique isomorphism of deformations of $\mathbf{\Phi}^0$ over A.*

Proof. Since $\mathbf{\Phi}^0$ is non-collinear, the natural morphism

$$\operatorname{Res}(\mathbf{\Phi}^0) \to \operatorname{Sch}(\mathbf{\Phi}^0) \subset (\mathbf{P}^2)^n$$

is an isomorphism. By Proposition 4.2.7, any deformation of $\operatorname{Sch}(\mathbf{\Phi}^0)$ is a blowup P of $\mathbf{P}^3_A$ at n disjoint sections over $\operatorname{Spec} A$. The deformation thus results in a rational map

$$\mathbf{\Phi} : \mathbf{P}^3_A \dashrightarrow (\mathbf{P}^2_A)^n$$

extending $\mathbf{\Phi}^0$. We wish to show that $\mathbf{\Phi}$ is a relative multiview configuration in the sense of Definition 4.1.4. To do this, it suffices to check that composition with each projection is a relative pinhole camera. Write $p : \mathbf{P}^3_A \dashrightarrow \mathbf{P}^2_A$ for one such projection; we will abuse notation and also write p for the corresponding map $P \to \mathbf{P}^2_A$ from the blowup. We will write E for the exceptional divisor associated to p and Z for the section blown up to make E. That is, we assume that p is the ith projection of Φ and that E is the preimage of the ith section in $\mathbf{P}^3_A$, which we call Z, uniformly omitting i from the notation. By the pinhole camera assumptions on Φ^0, $p|E_{A_0}$ maps E isomorphically to $\mathbf{P}^2_{A_0}$. It follows from Nakayama's lemma that $p|_E$ maps E isomorphically to $\mathbf{P}^2_A$.

Write $U \subset \mathbf{P}^3_A$ for the complement of the sections that are blown up to resolve Φ. By the previous paragraph, we see that $U_{A_0} \subset \mathbf{P}^3_{A_0}$ is precisely the complement of the camera centers of Φ^0. By the universal property of projective space, the morphism p is given by a surjective morphism

$$\lambda : \mathscr{O}_P^{\oplus 3} \to \mathscr{L}$$

for some $\mathscr{L}$ in $\operatorname{Pic}(P)$. Write $\pi : P \to \mathbf{P}^3_A$ for the blow-down map. We know from the definition of pinhole cameras, the rigidity of invertible sheaves on P, and the canonical way to extend morphisms generically across blowups that $\mathscr{L} \cong \pi^*(\mathscr{O}(1))(-E)$. Moreover, the resulting arrow

$$f : \pi_* \mathscr{O}^{\oplus 3} \to \mathscr{O}_{\mathbf{P}^3_A}(1)$$

has the property that its image is precisely $\mathscr{O}_{\mathbf{P}^3_A}(1) \otimes \mathscr{I}_Z$, where $\mathscr{I}_Z$ is the ideal sheaf of Z. (This follows from the universal property of blowing up.) This shows that the cokernel of f is an invertible sheaf supported on Z, showing that p is a relative pinhole camera, as desired.

It remains to show that any two such realizations Φ_1 and Φ_2 are conjugate by an infinitesimal automorphism of $\mathbf{P}^3$. But this follows immediately from Proposition 4.1.11. □

Lifting deformations for collinear configurations

For the sake of computational ease, in this section we consider a deformation situation $I \subset A \to A_0$ in which A is an Artinian local ring with maximal ideal $\mathfrak{m}$ and $\mathfrak{m}I = 0$. Write $k = A/\mathfrak{m}$.

We start with a multiview configuration $\mathbf{\Phi} : \mathbf{P}^3_{A_0} \dashrightarrow (\mathbf{P}^2)^n$ whose special fiber $\mathbf{\Phi}_k$ is collinear. Thus, the morphism

$$\operatorname{Res}(\mathbf{\Phi}_k) \to \operatorname{Sch}(\mathbf{\Phi}_k) \subset (\mathbf{P}^2)^n$$

contracts a line $\ell \subset \operatorname{Res}(\mathbf{\Phi}_k)$. To make things easier to read, write $R = \operatorname{Res}(\mathbf{\Phi}_k)$ and $B = \operatorname{Sch}(\mathbf{\Phi}_k)$. Write $L_{R/B}$ for the cotangent complex of the morphism $R \to B$. In addition, write $E_1, \ldots, E_n \subset R$ for the exceptional divisors. The usual calculations show that $K_R = \pi^* K_{\mathbf{P}^3} + E_1 + \cdots + E_n$.

Lemma 4.2.9. *If* $n > 4$ *then* $\operatorname{Ext}^2_R(L_{R/B}, \mathscr{O}_R) = 0$.

Proof. Consider the standard spectral sequence

$$E_2^{pq} = \operatorname{Ext}^p(\mathscr{H}^{-q}(L_{R/B}, \mathscr{O}_R)) \Rightarrow \operatorname{Ext}^{p+q}(L_{R/B}, \mathscr{O}_R). \tag{4.2.1.0.1}$$

We know that $\mathscr{H}^0(L_{R/B}) = \Omega^1_{R/B}$, and that $\mathscr{H}^{-j}(L_{R/B})$ is supported on ℓ for all $j \geq 0$. By Serre duality, we can compute the terms in the spectral sequence as

$$\operatorname{Ext}^p(\mathscr{H}^{-q}(L_{R/B}), \mathscr{O}_R) = \mathrm{H}^{3-p}(R, \mathscr{H}^{-q}(L_{R/B})(K_R))^\vee.$$

Since the cohomology sheaves of $L_{R/B}$ are all supported on ℓ, all columns of the E^2_{pq} page (4.2.1.0.1) vanish except (possibly) for $p = 2, 3$. It follows that

$$\operatorname{Ext}^2_R(L_{R/B}, \mathscr{O}_R) \cong \mathrm{H}^1(R, \Omega^1_{R/B}(K_R))^\vee.$$

A local calculation shows that $\Omega^1_{R/B}$ is annihilated by the ideal of ℓ, so that $\Omega^1_{R/B} = \Omega^1_{\ell/\operatorname{Spec} k}$, and thus

$$\mathrm{H}^1(R, \Omega^1_{R/B}(K_R))^\vee \cong \mathrm{H}^1(\ell, \mathscr{O}_\ell(K_\ell + K_R))^\vee \cong \mathrm{H}^0(\ell, \mathscr{O}_\ell(-K_R)) = \mathrm{H}^0(\ell, \mathscr{O}(4-n)) = 0,$$

as desired. □

Proposition 4.2.10. *Suppose* $n > 4$. *If* $X \subset (\mathbf{P}^2)^n_A$ *is an* A*-flat deformation of* $\mathrm{Sch}(\mathbf{\Phi}^0)$ *then there is a deformation* $\mathbf{\Phi}$ *of* $\mathbf{\Phi}^0$ *such that* $\mathrm{Sch}(\mathbf{\Phi}) = X$ *as closed subschemes of* $(\mathbf{P}^2)^n$. *Moreover,* $\mathbf{\Phi}$ *is unique up to unique isomorphism of deformations of* $\mathbf{\Phi}^0$ *over* A.

Proof. By Lemma 4.2.9 and [11, Proposition III.2.2.4], the obstruction to deforming $\mathrm{Res}(\Phi^0) \to \mathrm{Sch}(\Phi^0)$ over A vanishes, resulting in a deformation $R \to X$. Applying the results of Section 4.2.1, we see that this arises from a deformation Φ, as desired. The uniqueness of Φ up to isomorphism is an immediate consequence of Proposition 4.1.11. □

Proof of Theorem 4.2.6

In this section we complete the proof of Theorem 4.2.6. In the present terminology, this is equivalent to the following Proposition.

Proposition 4.2.11. *If* $\mathbf{\Phi}$ *is a non-collinear general multiview configuration or* $\mathrm{len}(\mathbf{\Phi}) > 4$ *then the morphism*

$$\mathrm{Sch} : \mathrm{Def}_{\mathbf{\Phi}^0} \to \mathrm{Def}_{\mathrm{Sch}(\mathbf{\Phi}^0)}$$

is an isomorphism of deformation functors.

Proof. The injectivity of Sch follows from Proposition 4.1.11. Surjectivity follows from Proposition 4.2.8 in the non-collinear case and Proposition 4.2.10 in the case $\mathrm{len}(\mathbf{\Phi}) > 4$. □

4.3 Diagram Hilbert schemes

In this section, we briefly explain a basic idea that is hard to find in the literature: diagram Hom-schemes and diagram Hilbert schemes. They are a mild elaboration of the idea of a flag Hilbert scheme. By not only remembering the data of the image but also the calibrating conics the moduli of calibrated cameras maps to a diagram Hilbert schemes in the same way that the moduli of uncalibrated cameras maps to a Hilbert scheme.

Definition and examples

Fix a base scheme S, a category I, and a functor $\underline{X} : I \to \mathfrak{AlgSp}_S$.

Definition 4.3.1. The *diagram Hilbert functor*

$$\mathrm{Hilb}_{\underline{X}} : \mathfrak{Sch}_S^{\circ} \to \mathfrak{Sets}$$

is the functor whose value on an S-scheme T is the set of isomorphism classes of natural transformations $\underline{Y} \to \underline{X} \times_S T$ of functors $I \to \mathfrak{Sch}_T$ where for each $i \in I$ the associated arrow $\underline{Y}(i) \to \underline{X}(i) \times_S T$ is a T-flat family of proper closed subschemes of $\underline{X}(i)$ of finite presentation over T.

Example 4.3.2. The usual Hilbert scheme is an example: just take I to be the singleton category. So is the flag Hilbert scheme of length n: in this case the category I is the category $\underline{n}$ associated to the poset $\{1, \ldots, n\}$, and the functor $\underline{X}$ is the constant functor $X \to X$. A natural transformation $\underline{Y} \to \underline{X}$ defines a nested sequence of closed subschemes of X. This is the flag Hilbert scheme (of length 2 flags).

There is also a stricter kind of flag scheme: suppose $X_1 \subset X_2$ is a closed immersion and one wants to parameterize pairs $Y_i \subset X_i$ such that $Y_1 \subset Y_2$. That is precisely the diagram Hilbert functor associated to the poset-category $\underline{2} = \{0 < 1\}$ with the functor $\underline{2} \to \mathfrak{Sch}_S$ sending i to X_i. This last example is the one that will arise naturally for us in the context of calibrated cameras. (We record more general results here in case someone in the future needs this general idea of diagram Hilbert scheme.)

Notation 4.3.3. If the diagram in question is a single morphism $X \to Y$, we will write $\mathrm{Hilb}_{X \to Y}$ for the associated Hilbert functor.

Representability

The main result about diagram Hilbert functors is that they are representable.

Proposition 4.3.4. *Let I be a finite category and $\underline{X} : I \to \mathfrak{AlgSp}_S$ a functor whose components are separated algebraic spaces. Then the diagram Hilbert functor $\mathrm{Hilb}_{\underline{X}}$ is representable by an algebraic space locally of finite presentation over S. If the $\underline{X}(i)$ are locally quasi-projective schemes then $\mathrm{Hilb}_{\underline{X}}$ is represented by a locally quasi-projective S-scheme.*

Proof. There is a natural functor

$$F : \mathrm{Hilb}_{\underline{X}} \to \prod_{i \in I} \mathrm{Hilb}_{\underline{X}(i)},$$

and we know that the latter is representable by algebraic spaces (resp. schemes) satisfying the desired conditions. It thus suffices to show the same for F, i.e., that F is representable by spaces of the required type.

For each $i \in I$, let

$$Z_i \subset \underline{X}(i) \times \prod \mathrm{Hilb}_{\underline{X}(i)}$$

denote the universal closed subscheme (pulled back over the product). Let A denote the set of arrows in I; for an arrow $a \in A$, let $s(a)$ and $t(a)$ denote the source and target of a. Consider the scheme

$$H := \prod_{a \in A} \mathrm{Hom}_{\mathrm{Hilb}_{\prod \underline{X}(i)}}(Z(s(a)), Z(t(a))),$$

which naturally fibers over $\prod \mathrm{Hilb}_{\underline{X}(i)}$. The standard theory of Hom-schemes shows that $H \to \prod \mathrm{Hilb}_{\underline{X}(i)}$ is representable by spaces of the desired type.

The final observation to make is that composition of two arrows gives equations $b \circ a = c$ in A, and these translate into *closed* conditions on H because all of the subschemes $Z(i)$ are separated. Since the conditions desired are stable under taking closed subspaces, we have proven the result. $\square$

4.4 Moduli of calibrated camera configurations

Let $\mathscr{C}$ denote the space of smooth conics in $\mathbf{P}^2_{\mathrm{Spec}\,\mathbf{Z}}$, and let $C_{\mathrm{univ}} \subset \mathbf{P}^2_{\mathscr{C}}$ denote the universal smooth conic. (The space $\mathscr{C}$ is an open subscheme of the bundle of sections of

$\mathscr{O}_{\mathbf{P}^2_{\operatorname{Spec} Z}}(2)$.) The tuple of conics $(C_{\text{univ}}, \ldots, C_{\text{univ}})$ inside $(\mathbf{P}^2)^n$ will be called the *universal calibration*.

Definition 4.4.1. Given a positive integer n, the *stack of calibrated camera configurations of length* n, denoted CalCam_n, is the stack over $\mathscr{C}^n$ whose value over a point $t : S \to \mathscr{C}^n$ consists of the groupoid of general relative calibrated multiview configurations of length n with calibration datum of the form $(C, t^*(C_{\text{univ}}, \ldots, C_{\text{univ}}))$.

In down-to-earth terms, we are just describing the stack of n-tuples of calibrated cameras with pairwise non-intersecting centers, together with *arbitrary but specified* calibration data. In the existing literature, the word “calibrated” usually means that one has fixed the calibrating conics to be the canonical absolute conic in space (attached to the Euclidean distance form on $\mathbf{P}^3$) and the circle in the plane. Since any two smooth conics are conjugate under a homography, this seems harmless. As we hope to describe in this section, thinking more geometrically and *tracking the conics as data instead of normalizing them* gives us a great deal of insight into the underlying moduli problem. The point of the universal conic in $\mathbf{P}^2$ is that we only want to allow the conic in $\mathbf{P}^3$ to vary; that is, we fix calibration data on the image planes when we define the moduli problem. By working with the universal conic, we allow those fixed planar data to be arbitrary.

Notation 4.4.2. Since we are fixing the calibration data on the image planes to be the universal conic, we will omit them from the notation for a calibration datum. Thus, we will write (Φ, C) for a calibrated configuration. When we need to refer to the image plane calibrating curves, we will use C_i for the curve in the ith plane, it is key to remember that while C_i can vary as the base varies (depending upon how it maps to $\mathscr{C}^n$, this is determined solely by the base and not by the object of CalCam_n over that point of the base.

The main result of this section is the following.

Proposition 4.4.3. *The stack* CalCam_n *is a smooth algebraic space of finite type over* $\mathscr{C}^n$.

Let $\tau_n : \mathrm{CalCam}_n \to \mathrm{CalCam}_{n-1} \times_{\mathscr{C}^{n-1}} \mathscr{C}^n$ be the morphism given by forgetting the last camera (and retaining the last calibrating plane conic).

Lemma 4.4.4. *The morphism* τ_n *is representable by separated schemes of finite presentation.*

Proof. Let $((\varphi_1, \ldots, \varphi_{n-1}, C), C_n)$ be a T-valued point of $\mathrm{CalCam}_{n-1} \times_{\mathscr{C}^{n-1}} \mathscr{C}^n$. The fiber of τ_n is given by the set of cameras φ_n with the same domain $\mathbf{P} \to T$ as the first $n-1$ cameras, with the following additonal properties.

1. The center of φ_n avoids the centers of φ_i for $i = 1, \ldots, n-1$.
2. The center of φ_n avoids C.
3. The restriction $\varphi_n|C$ factors through the closed subscheme $C_n \subset \mathbf{P}$.

The space of camera centers satisfying the first two conditions is an open subscheme $\mathbf{P}^\circ \subset \mathbf{P}$, and taking the center gives a natural map

$$\mathrm{CalCam}_n \to \mathbf{P}^\circ \times \mathrm{CalCam}_{n-1} \times_{\mathscr{C}^{n-1}} \mathscr{C}^n.$$

It suffices to show that this map is representable, and thus we may assume that the center is a given section $\sigma : T \to \mathbf{P}$. Blowing up along $\sigma(T)$ to yield $\widetilde{\mathbf{P}}$, with exceptional divisor E, we can then realize the cameras inside the open locus of the Hom-scheme $\mathrm{Hom}(\widetilde{\mathbf{P}}, \mathbf{P}^2)$ parametrizing maps $f : \widetilde{\mathbf{P}} \to \mathbf{P}^2$ for which $f^*\mathscr{O}_{\mathbf{P}^2}(1)$ is isomorphic to $\mathscr{O}(1)(-E)$ on each geometric fiber over T. This locus is of finite type. Finally, the condition that C lands in C_n is closed (and of finite presentation), completing the proof. □

Proposition 4.4.5. *The morphism* τ_n *is smooth.*

Proof. By Lemma 4.4.4 and [17, Tag 02H6], it suffices to show that τ_n is formally smooth. Let $A \to A_0$ be a square-zero extension of rings, and suppose that $(\varphi_1, \ldots, \varphi_n, C) \in \mathrm{CalCam}_n(A_0)$ is fixed. To show formal smoothness we can work Zariski-locally and thus assume that the domains of $\varphi_1, \ldots, \varphi_n$ are $\mathbf{P}^3_{A_0}$. Now suppose that we fix a deformation

$$((\varphi_1', \ldots, \varphi_{n-1}', C_A), C_n) \in \mathrm{CalCam}_{n-1}(A) \times_{\mathscr{C}^{n-1}(A)} \mathscr{C}^n(A).$$

(Because we are working over the universal conic in each image plane, we have to specify the deformation of the conic that we will use in attempting to deform the nth calibrated camera.) To show formal smoothness is suffices to extend φ_n to a morphism φ_n' that maps C_A to C_n.

The choice of deformation of C to C_A induces a lift of $C \to \mathbf{P}^2_{A_0}$ to $C_A \to \mathbf{P}^2_A$. This is because embeddings of degree two curves are given by choosing sections of $\mathscr{O}_{\mathbf{P}^1}(2)$, so any collection of sections embedding C can be extended to an embedding of C_A.

We are thus reduced to the following: we are given a tuple of three sections $\sigma_0, \sigma_1, \sigma_2 \in \Gamma(\mathbf{P}^3_{A_0}, \mathscr{O}(1))$, a degree two curve $C_A \subset \mathbf{P}^3_A$, and lifts of the $\sigma_j|_C$ to $\Gamma(C_A, \mathscr{O}(1))$. We wish to lift these extensions to sections $\widetilde{\sigma}_j \in \Gamma(\mathbf{P}^3_A, \mathscr{O}(1))$. We can do this one section at a time. Since C_A is a degree two curve, it is contained in a canonically defined plane in $\mathbf{P}^3$; we will write $C_A \subset \mathbf{P}^2_A \subset \mathbf{P}^3_A$ and similarly for A_0. (If the plane is not trivial, we can further shrink A to make it so; this is immaterial for the calculations and is only a notational device.)

Consider the diagrams

$$\begin{tikzcd}
0 \arrow[r] & \Gamma(\mathbf{P}^3_{A_0}, \mathscr{O}) \otimes_{A_0} I \arrow[r] \arrow[d] & \Gamma(\mathbf{P}^3_A, \mathscr{O}) \arrow[r] \arrow[d] & \Gamma(\mathbf{P}^3_{A_0}, \mathscr{O}) \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \Gamma(\mathbf{P}^3_{A_0}, \mathscr{O}(1)) \otimes_{A_0} I \arrow[r] \arrow[d] & \Gamma(\mathbf{P}^3_A, \mathscr{O}(1)) \arrow[r] \arrow[d] & \Gamma(\mathbf{P}^3_{A_0}, \mathscr{O}(1)) \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \Gamma(\mathbf{P}^2_{A_0}, \mathscr{O}(1)) \otimes_{A_0} I \arrow[r] & \Gamma(\mathbf{P}^2_A, \mathscr{O}(1)) \arrow[r] & \Gamma(\mathbf{P}^2_{A_0}, \mathscr{O}(1)) \arrow[r] & 0
\end{tikzcd}$$

and

$$\begin{tikzcd}
0 \arrow[r] & \Gamma(\mathbf{P}^2_{A_0}, \mathscr{O}(-1)) \otimes_{A_0} I \arrow[r] \arrow[d] & \Gamma(\mathbf{P}^2_{A}, \mathscr{O}(-1)) \arrow[r] \arrow[d] & \Gamma(\mathbf{P}^2_{A_0}, \mathscr{O}(-1)) \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \Gamma(\mathbf{P}^2_{A_0}, \mathscr{O}(1)) \otimes_{A_0} I \arrow[r] \arrow[d] & \Gamma(\mathbf{P}^2_{A}, \mathscr{O}(1)) \arrow[r] \arrow[d] & \Gamma(\mathbf{P}^2_{A_0}, \mathscr{O}(1)) \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \Gamma(C, \mathscr{O}(1)) \otimes_{A_0} I \arrow[r] & \Gamma(C_A, \mathscr{O}(1)) \arrow[r] & \Gamma(C, \mathscr{O}(1)) \arrow[r] & 0.
\end{tikzcd}$$

By the usual calculations of the cohomology of projective space, these two diagrams have exact columns. A simple diagram chase then shows that we can lift sections to $\mathbf{P}^3_A$ given values on $\mathbf{P}^3_{A_0}$ and C_A, completing the proof. □

Proof of Proposition 4.4.3. It remains to show smoothness. We use Proposition 4.4.5 and induction on n. For $n = 1$, we see that CalCam_1 is smooth over $\mathscr{C}$, which is itself open in a projective space, hence smooth. □

4.4.1 Deformation theory of calibrated camera configurations

In this section we prove the following analogue of Theorem 4.2.6.

Theorem 4.4.6. *If (Φ, C) is a non-degenerate calibrated general multiview configuration of length n with associated multiview flag*

$$(C \subset V) \hookrightarrow (C_1 \times \cdots \cdot C_n \subset (\mathbf{P}^2)^n)$$

then, assuming either that Φ is non-collinear or that $n > 4$, we have that the infinitesimal deformations of (Φ, C) are in bijection with the infinitesimal deformations of $C \subset V$ as a closed subscheme diagram of $C_1 \times \cdots \times C_n \subset (\mathbf{P}^2)^n$.

Proof. The proof leverages the proof of Theorem 4.2.6. In particular, we can forget the calibrations and apply Theorem 4.2.6 to see that under the given hypotheses any deformation of $\mathrm{Flag}(\Phi, C)$ induces a deformation of $\mathrm{Sch}(\Phi)$ that is the image of a deformation $\widetilde{\Phi}$ of Φ. The assumption that the deformation of $\mathrm{Sch}(\Phi)$ arises from a deformation of $\mathrm{Flag}(\mathbf{\Phi}, C)$ means that there is also an associated deformation of C. Since $\mathbf{\Phi}$ is an isomorphism onto its image in a neighborhood of C, this deformation of C canonically lifts to give a calibration of $\widetilde{\mathbf{\Phi}}$. □

4.5 Comparison morphisms

4.5.1 Embedding into the Hilbert scheme

The following describes the main result relating the moduli problems Cam_n and CalCam_n to Hilbert schemes. Because the statements and proofs are so similar, we combine everything into a single omnibus Proposition. This gives the generalization of the results of [3, Section 6], leveraging the novel methods of this paper to give more information about the uncalibrated case and the appropriate result in the calibrated case.

Proposition 4.5.1. *The associations* $\mathbf{\Phi} \mapsto \mathrm{Sch}(\mathbf{\Phi})$ *and* $(\mathbf{\Phi}, C) \mapsto \mathrm{Flag}(\mathbf{\Phi}, C)$ *define monomorphisms*

$$\mathrm{Sch} : \mathrm{Cam}_n \to \mathrm{Hilb}_{(\mathbf{P}^2)^n/\operatorname{Spec}\mathbf{Z}}$$

and

$$\mathrm{Flag} : \mathrm{CalCam}_n \to \mathrm{Hilb}_{C^n_{\mathrm{univ}} \subset (\mathbf{P}^2)^n/\mathscr{C}^n}$$

such that

1. *the restriction of* Sch *(resp.* Flag*) to the non-collinear locus* $\mathrm{Cam}_n^{\mathrm{nc}}$ *(resp.* $\mathrm{CalCam}_n^{\mathrm{nc}}$*) is an open immersion into the smooth locus* $\mathrm{Hilb}^{\mathrm{sm}}_{(\mathbf{P}^2)^n/\operatorname{Spec}\mathbf{Z}}$ *(resp.* $\mathrm{Hilb}^{\mathrm{sm}}_{C^n_{\mathrm{univ}} \subset (\mathbf{P}^2)^n/\mathscr{C}^n}$*);*

2. *when* $n > 4$*, the morphism* Sch *(respectively,* Flag*) itself is an open immersion into* $\mathrm{Hilb}^{\mathrm{sm}}_{(\mathbf{P}^2)^n/\operatorname{Spec}\mathbf{Z}}$ *(respectively,* $\mathrm{Hilb}^{\mathrm{sm}}_{C^n_{\mathrm{univ}} \subset (\mathbf{P}^2)^n/\mathscr{C}^n}$*);*

3. *the arrows* Sch *and* Flag *together with the forgetful maps give a commutative*

diagram

$$\begin{tikzcd}
\mathrm{CalCam}_n \arrow[r, "\mathrm{Flag}"] \arrow[d, "\nu_n"'] & \mathrm{Hilb}_{C^n_{\mathrm{univ}} \subset (\mathbf{P}^2_{\mathscr{C}^n})^n/\mathscr{C}^n} \arrow[d] \\
\mathrm{Cam}_n \times_{\mathrm{Spec}\,\mathbf{Z}} \mathscr{C}^n \arrow[r, "\mathrm{Sch}"'] & \mathrm{Hilb}_{(\mathbf{P}^2_{\mathscr{C}^n})^n/\mathscr{C}^n} .
\end{tikzcd}$$

In particular, every geometric fiber of Sch *over* $\mathrm{Spec}\,\mathbf{Z}$ *is a dominant monomorphism of* Cam_n *into a single irreducible component of the Hilbert scheme, and when* $n > 4$ *the fiber of* Sch *itself is open in the smooth locus of the Hilbert scheme, and similarly for geometric fibers of* Flag *and components of the diagram Hilbert scheme.*

Proof. Proposition 4.1.13 and Proposition 4.1.11 show that Flag is well-defined and a monomorphism. Since CalCam_n is smooth over $\mathscr{C}^n$, we have that Flag is an open immersion in a neighborhood of any point where it induces an isomorphism of deformation functors. Theorem 4.4.6 then applies to give the two desired statements. ☐

4.5.2 The decalibration morphism $\nu_n : \mathrm{CalCam}_n \to \mathrm{Cam}_n \times \mathscr{C}^n$

In this section, we study a natural morphism

$$\mathrm{CalCam}_n \to \mathrm{Cam}_n \times \mathscr{C}^n$$

given by forgetting the camera calibration datum.

Definition 4.5.2. The *decalibration morphism* is the morphism

$$\nu_n : \mathrm{CalCam}_n \to \mathrm{Cam}_n \times \mathscr{C}^n$$

given by sending $(\mathbf{\Phi}, C)$ to $\mathbf{\Phi}$.

Our main result is that ν_n is unramified and non-injective. Thus, while CalCam_n is smooth over $\mathscr{C}^n$, its image in $\mathrm{Cam}_n \times \mathscr{C}^n$ need not be smooth. And this happens in practice: for $n = 2$, if we take the standard circle in each image plane as calibration datum, the morphism ν_n becomes (in the fiber over the "circles" calibration datum) a map into the variety of fundamental matrices whose image is the subvariety of essential matrices. The latter is singular. It would be interesting to understand precisely how its singularities arise from the point of view we take here. (Perhaps this singular locus is precisely the locus where there is only a single calibrating conic in $\mathbf{P}^3$.)

4.5.3 Intersections of conic cones

Before we delve into the geometry of ν_n, we need a few preliminaries about intersections of conic cones in $\mathbf{P}^3$. We thank Bianca Viray for pointing out an omission in the previous version of this Proposition and helping us think through the correct list of possibilities.

Proposition 4.5.3. *Let X_1 and X_2 be two conic cones in $\mathbf{P}^3$ with distinct cone points. The intersection $X_1 \cap X_2$ is one of the following.*

1. *An irreducible curve of degree 4.*
2. *A union of a twisted cubic and a line.*
3. *A union of two smooth conics.*
4. *A union of one smooth conic and a doubled line.*

It can never be a doubled conic, a quadrupled line, or contain two distinct lines.

Proof. Sections of $\mathscr{O}_{\mathbf{P}^3}(2)$ correspond to symmetric 4×4-matrices (at least if 2 is invertible on the base scheme). The conic cones correspond to the rank 3 matrices. Thus, they form a dense open in a hypersurface in $|\mathscr{O}_{\mathbf{P}^3}(2)|$ of degree 4.

The intersection $X_1 \cap X_2$ is an effective Cartier divisor on X_1 with class $\mathscr{O}_{X_1}(2)$. Since a general pencil of sections of $\mathscr{O}_{\mathbf{P}^3}(2)$ will intersect the rank 3 locus in four points, a general pair of cones span a general pencil, and thus they will have smooth intersection by Bertini's theorem. Thus, the intersection $X_1 \cap X_2$ can be (in fact, usually is) a smooth curve of degree 4.

Given a twisted cubic $C \subset \mathbf{P}^3$, choose two distinct points $x_1, x_2 \in C$. Consider linear projections $\pi_i : \mathbf{P}^3 \dashrightarrow \mathbf{P}^2$ centered at x_i. The image of C under a general such projection will be a smooth conic curve $D_i \subset \mathbf{P}^2$. Intersecting the cones X_1 and X_2 over D_1 and D_2 yields $C \cup L$, where L is a line. (Indeed, the total curve must have degree 4, so the residual curve has degree 1 and must be a line.) Note that $X_1 \cap X_2$ cannot contain a singular cubic space curve. Indeed, any reduced singular cubic must be a plane curve (since otherwise a general projection from a smooth point on the curve will map it to an irreducible singular conic in the plane, which is impossible), and a tripled line on $X_1 \cap X_2$ cannot be Cartier at the cone points, hence $X_1 \cap X_2$ must contain another line, which forces the cone points to coincide.

If the intersection contains a conic (which can be arranged by fixing a conic C and noting that the kernel of the restriction map

$$\Gamma(\mathbf{P}^3, \mathscr{O}_{\mathbf{P}^3}(2)) \to \Gamma(C, \mathscr{O}_C(2)) = \Gamma(\mathbf{P}^1, \mathscr{O}(4))$$

has dimension 5, thus producing many pencils through C), then the residual curve is either another conic, a doubled line, a pair of intersecting lines, or another copy of the same conic (i.e., the conic is doubled). We analyze these separately.

First, one can have two smooth conics. It suffices to find a single example to see that this is general behavior within the locus of pencils with non-smooth intersection. A simple example is furnished by the conic cones given in homogeneous coordinates by $X^2 + Y^2 + Z^2 = 0$ and $Y^2 + Z^2 + W^2 = 0$.

What if the residual curve is two distinct lines meeting at a point? Since this must be in both cones, and any such pair of lines must meet at the cone point, we would conclude that X_1 and X_2 have the same cone point, contrary to our assumption. Thus, this cannot happen.

The residual curve *can* be a doubled line. An example is given by the pair of cones $X^2 - YZ$ and $(X - \alpha W)^2 - (Y - \alpha W)(Z - \alpha W)$ for any non-zero α. (The first cone is being translated along one of its rulings. The resulting cones are tangent along this ruling, leading to a double line of intersection.)

The intersection cannot be a quadrupled line. A quadrupled line is the intersection of two doubled planes. So we can take a pencil generated by two doubled planes and show that it cannot contain any rank 3 forms. After a change of coordinates, we may assume that the doubled planes are given by $X^2 = 0$ and $Y^2 = 0$. The pencil they span cannot contain any form of rank greater than 2.

It remains to rule out a doubled conic. Note that a doubled conic is the intersection of X_1 with a doubled plane $2P \in \mathscr{O}_{\mathbf{P}^3}(2)$. We can rule out this case if we can show that the pencil spanned by X_1 and a doubled plane not containing its cone point does not contain any more conic cones. This readily reduces to the following matrix calculation. We can represent the cone X_1 by the matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and an arbitrary doubled plane missing the cone point by the matrix

$$\begin{pmatrix} a^2 & ab & ac & a \\ ab & b^2 & bc & b \\ ac & bc & c^2 & c \\ a & b & c & 1 \end{pmatrix}$$

for $a, b, c \in k$. Searching for a conic cone in the pencil corresponds to finding λ such that the matrix

$$\begin{pmatrix} a^2 + \lambda & ab & ac & a \\ ab & b^2 + \lambda & bc & b \\ ac & bc & c^2 + \lambda & c \\ a & b & c & 1 \end{pmatrix}$$

has rank 3. But row-reducing that matrix yields

$$\begin{pmatrix} \lambda & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ a & b & c & 1 \end{pmatrix},$$

and this matrix can never have rank 3. □

Lemma 4.5.4. *Fix a smooth conic $C \subset \mathbf{P}^3$. The space $\Xi \subset \mathbf{P}^9$ of conic cones in the linear system $|\mathscr{O}_{\mathbf{P}^3}(2)|$ that contain C has dimension 3.*

Proof. The restriction map

$$\Gamma(\mathbf{P}^3, \mathscr{O}(2)) \to \Gamma(\mathbf{P}^2, \mathscr{O}(2))$$

induces a surjective linear projection of linear systems

$$\rho : \mathbf{P}^9 \dashrightarrow \mathbf{P}^5.$$

Moreover, since any pair of smooth conics in $\mathbf{P}^3$ are conjugate under an automorphism of $\mathbf{P}^3$, we have that the fibers over any two smooth conics are isomorphic. Since smooth conics are general points of $\mathbf{P}^5$, we see that the fiber over a smooth conic has dimension 4 (i.e., $9 - 5$).

On the other hand, the locus $S \subset \mathbf{P}^9$ of conic cones is open in the hypersurface of singular members of $|\mathscr{O}_{\mathbf{P}^3}(2)|$. Since the cone over C contains C, we have that $S \cap \rho^{-1}([C])$ is non-empty, hence is a threefold in $\rho^{-1}([C])$. □

Lemma 4.5.5. *Suppose $A, B \in \Gamma(\mathbf{P}^3, \mathscr{O}_{\mathbf{P}^3}(2))$ is a regular sequence of elements (i.e., they intersect properly everywhere). Let $I = A \cap B$ the scheme-theoretic intersection. Then the kernel of*

$$\Gamma(\mathbf{P}^3, \mathscr{O}(2)) \to \Gamma(I, \mathscr{O}(2))$$

is the subspace spanned by A and B.

Proof. By the regular sequence assumption, we have a resolution

$$0 \to \mathscr{O}(-4) \to \mathscr{O}(-2)^{\oplus 2} \to \mathscr{O} \to \mathscr{O}_I \to 0.$$

Twisting by 2 and using the vanishing of $\mathrm{H}^1(\mathbf{P}^3, \mathscr{O}(a))$ (for all a) gives the result. □

4.5.4 The geometry of ν_n

Fix a point ξ of $\mathrm{Cam}_n \times \mathscr{C}^n$. That is, fix conics $C_1, \ldots, C_n$ in $\mathbf{P}^2$ and a multiview configuration $\mathbf{\Phi}$. In this section we compute the fiber of ν_n over ξ.

Proposition 4.5.6. *The scheme-theoretic fiber $\nu_n^{-1}(\xi)$ is a reduced $\kappa(\xi)$-scheme of length at most 2.*

Proof. The fiber $\nu_n^{-1}(\xi)$ is precisely the scheme of smooth conics in the intersection of the cones over the image conics C_i inside the ambient $\mathbf{P}^3$. The result is thus immediate from Proposition 4.5.3. (In particular, the lack of doubled conic means that the fibers are discrete.) □

Corollary 4.5.7. *The morphism ν_n is unramified.*

Proof. This is an immediate consequence of Proposition 4.5.6. □

Proposition 4.5.8. *The morphism ν_n is proper.*

Proof. Suppose we have a multiview configuration $\mathbf{\Phi}$ of length 2 over a complete dvr R with fraction field K, degree two curves $C_1, \ldots, C_n \subset \mathbf{P}^2_R$ and a degree two curve $C_K \subset \mathbf{P}^3_K$ such that $\mathbf{\Phi}_K$ maps C_K isomorphically to the generic fiber of each C_i. By the valuative criterion for properness it suffices to extend C_K to a degree two curve C_R.

Assume we have a multiview configuration $\mathbf{\Phi}$ of length 2 over a complete dvr R with fraction field K, and suppose we have conics $C_1, \ldots, C_n \subset \mathbf{P}^2_R$ in each image plane. Write $\overline{C}_i \subset \mathbf{P}^3$ for the cone over C_i under $\mathrm{pr}_i \circ \mathbf{\Phi}$ and $I = \overline{C}_1 \cap \cdots \cap \overline{C}_n$. Finally, assume that there is a conic $C_K \subset \mathbf{P}^3_K$ such that $\mathbf{\Phi}_K$ maps C_K isomorphically to the generic fiber of each C_i; that is, $C_K \subset I_K$. Let C_R be the specialization of C_K in the closed fiber C_0. The curve C_R is degree 2, giving us a calibrated configuration over R. □

Note that even if C_k is a non-degenerate conic, C_0 need not be. This is why we need to add degenerate conics.

Proposition 4.5.9. *The morphism ν_2 has smooth image and general fiber of length 2. For any $n > 2$ the morphism ν_n is generically injective.*

Proof. The projective closure of the image of a fiber of CalCam_2 over $\mathscr{C}^2$ under ν_2 is known as the "essential variety", and its singularities are well-known (see [6, Proposition 2.1]); none of its singular points lies in the image of ν_2. To study the general fiber, it suffices by the irreducibility of all spaces involved to produce a single example of a camera configuration of length two such that the fiber of ν_n has length 2. (Indeed, the locus of quartic curves in $\mathbf{P}^3$ containing a conic has a union of two conics as its generic point.) To do this, it further suffices to find a single example of two conic cones $C_1, C_2 \subset \mathbf{P}^3$ whose intersection is a pair of smooth conics. (Indeed, general projections from the two cone points give image planes together with calibrating conics that give rise to fibers of ν_2 of length 2.) But this has already been written down in the proof of Proposition 4.5.3.

We now show that ν_n is generically injective for $n > 2$. By Lemma 4.5.4, given a smooth conic C in $\mathbf{P}^3$, the locus in $|\mathscr{O}_{\mathbf{P}^3}(2)|$ consisting of conic cones containing C is 3-dimensional. Thus, we can find three non-collinear conic cones that contain any given smooth conic C. On the other hand, by Lemma 4.5.5, given two conic cones C_1, C_2, the set of conic cones that vanish on their entire intersection $C_1 \cap C_2$ is contained in the pencil spanned by the C_i. We conclude that if $C_1 \cap C_2$ is reducible, then we can choose general cones $C_3, \ldots, C_n$ containing a smooth conic in $C_1 \cap C_2$ such that C_i is not in the pencil spanned by C_1 and C_2 for each $i > 2$. The joint vanishing locus $C_1 \cap C_2 \cap C_3 \cap \cdots \cap C_n$ is a smooth conic. Since this is generic behavior, this shows that ν_n is generically injective for all $n > 2$. □

It is a potentially interesting problem to characterize the locus over which ν_n is not injective, and the singular locus of its image (the "variety of calibrated n-focal tensors", which is studied for $n = 3$ in coordinatized form in [12]).

Corollary 4.5.10. *The morphism ν_n is finite.*

Proof. We have shown that ν_n is quasi-finite and proper and thus, finite. ☐

Chapter 5

THE ESSENTIAL VARIETY

In this chapter, we study the map $\nu_2 : \mathrm{CalCam}_2 \to \mathrm{Cam}_2$. More specifically, in Section 5.1 we use our moduli results to abstractly describe a 2-1 cover of the essential variety that compactifies the twisted pair covering. In Section 5.2, we give an entirely different description starting from basic principles to produce this 2-1 cover as well as explicitly give equations describing CalCam_2.

5.1 Twisted pairs and moduli

We begin by studying the morphism ν_2 in more detail, showing how the Hilbert scheme gives a natural compactification of the classical "twisted pair" construction. To explicitly compare this new treatment with the literature, in this section we will fix the calibrating conics to be $v(x_0^2 + x_1^2 + x_2^2) \subset \mathbf{P}^2$. Also, we will often think of an essential matrix as the corresponding pair of calibrated cameras in normalized coordinates. In these coordinates we can fix notation $P_1 = [I|0]$ and $P_2 = R[I|t]$ where $t = (a, b, c)$.

5.1.1 Twisted pairs

As shown in Section 5.2 of [15], the locus $\mathbf{E}^{\mathrm{sm}}$ of essential matrices is smooth (over $\mathbf{C}$) and admits an étale surjection $\mathrm{SO}(3) \times \mathbf{P}^2 \to \mathbf{E}^{\mathrm{sm}}$. We can understand this surjection as follows. Given a calibrated camera $P : \mathbf{P}^3 \dashrightarrow \mathbf{P}^2$, we can make a new camera Q by composing with a rotation (an element of $\mathrm{SO}(3)$) and a translation (an element of $\mathbf{A}^3 \smallsetminus \{0\}$). Since there is always a scaling ambiguity in reconstruction, we may assume we are translating by a unit vector, so the translation is really an element of $\mathbf{P}^2$. (This scaling ambiguity also allows us to replace any arbitrary orthogonal coordinate transformation with an element of $\mathrm{SO}(3)$, by scaling by -1.) This gives a map $\pi : \mathrm{SO}(3) \times \mathbf{P}^2 \to \mathbf{E}^{\mathrm{sm}}$. In terms of matrices we send (R, t) to the camera pair $P = [I|0], Q = [R|t]$ which has essential matrix $[t]_{\times} R$. Since coordinates can be normalized to write any pair of calibrated cameras in this form the morphism π is surjective on geometric points. Moreover, one can check in local coordinates that the map is étale [5, Proposition 3.2].

For any *real* essential matrix $M \in \mathbf{E}^{\mathrm{sm}}(\mathbf{R})$, the fiber of π over M contains two points: one can take a pair of cameras P_1, P_2 and replace it with the pair $P_1, \widetilde{P}_2$ where $\widetilde{P}_2$ results from rotating P_2 by 180 degrees around the axis connecting the centers of P_1 and P_2. In normalized coordinates, the matrix

$$R_t = \begin{pmatrix} 2a^2 - 1 & 2ab & 2ac & 0 \\ 2ab & 2b^2 - 1 & 2bc & 0 \\ 2ac & 2bc & 2c^2 - 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

is rotation by 180 degrees and $\widetilde{P}_2 = R[I|t]R_t$. (note that over the reals we can always rescale t so that $a^2 + b^2 + c^2 = 1$.) The pair $(P_1, P_2), (P_1, \widetilde{P}_2)$ is called a *twisted pair*; what we have described is a well-known construction in computer vision [8, Result 9.19]. The key thing to note is that the rotation construction described above *preserves calibrations* for real cameras. For complex cameras, things get more complicated, and for displacements (a, b, c) such that $a^2 + b^2 + c^2 = 0$, the corresponding transformation produces a new camera pair $(P_1, \widetilde{P}_2)$ for which $\widetilde{P}_2$ is no longer calibrated.

5.1.2 *Compactification of the twisted pair construction*

The morphism $\nu_2 : \mathrm{CalCam}_2 \to \mathrm{Cam}_2 \times \mathscr{C}^2$ gives a double covering of a closed subscheme that generalizes the twisted pair covering of the essential variety. A point of CalCam_2 is the datum (P_1, P_2, C) where P_1 and P_2 are cameras and C is a conic contained in the intersection of the cones defined by the preimage of C_{univ} via P_1 and P_2. Proposition

4.5.3 tells us that this intersection must contain either another non-degenerate conic or a doubled line. In either case denote this other degree two curve by $\widetilde{C}$. The general fibers of ν_2 are the triples (P_1, P_2, C) and $(P_1, P_2, \widetilde{C})$.

This double covering agrees with the twisted pairs covering on real points. In normalized coordinates $\widetilde{C}$ is defined by the simultaneous vanishing of

$$x^2 + y^2 + z^2 = 0 \text{ and } (a^2 + b^2 + c^2)w - 2(ax + by + cz) = 0.$$

When $a^2 + b^2 + c^2 = 1$, as it must over $\mathbf{R}$ (up to scaling), one can check that changing coordinates on $\mathbf{P}^3$ via the automorphism

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -2a & -2b & -2c & 1 \end{pmatrix}$$

sends the triple $(P_1, \widetilde{P}_2, C)$ to the triple $(P_1, P_2, \widetilde{C})$.

However, over the complex numbers there exist essential matrices such that $a^2 + b^2 + c^2 = 0$. This is exactly the condition that $\widetilde{C}$ is a doubled line. In this situation the twisted pair construction fails because the camera $\widetilde{P}$ no longer has a trivial calibration. Mathematically speaking, we are really discussing the fact that the twisted pairs morphism π, while always étale, is *not* finite. Allowing degenerate calibrations (doubled lines) extends the twisted pair morphism π to ν_2.

Proposition 5.1.1. *There exists a fixed-point free involution,* $\chi : \mathrm{CalCam}_2 \to \mathrm{CalCam}_2$ *over* Cam_2 *given by fixing the cameras and swapping calibrating conics. More precisely,* $\nu_2 \circ \chi = \nu_2$.

Proof. Given a pair of cameras $P_1, P_2 : \mathbf{P}^3 \to \mathbf{P}^2$ and smooth conics $D_1, D_2 \subset \mathbf{P}^2$, we can pull back to get two cones $C_{P_1}, C_{P_2} \subset \mathbf{P}^3$. Let $Y = C_{P_1} \cap C_{P_2}$. Blowing up the camera centers, the strict transform of these cones, $\tilde{C_{P_1}}, \tilde{C_{P_2}} \subset \mathrm{Bl}_{Z_1,Z_2} \mathbf{P}^3$, are smooth surfaces in

$\mathbf{P}^3$. The intersection is a relative effective Cartier divisor and $\tilde{C}_{P_1} \cap \tilde{C}_{P_2} \simeq Y$ since the cone centers are distinct.

A point in CalCam_2 is a pair (Φ, C) where C is a relative effective Cartier divisor contained in Y. By [17, Tag 0B8V] there exists another relative effective Cartier divisor C' such that $C' + C = Y$. Checking at a geometric point, Proposition 4.5.3 shows that C' is a degree 2 curve, and that no geometric point of CalCam_2 is fixed by χ. This argument is functorial and so induces the desired involution. Since χ only changes the calibrating conic we have $\nu_2 \circ \chi = \nu_2$. □

Theorem 5.1.2. *The morphism ν_2 factors as a finite étale morphism followed by a closed immersion.*

Proof. By Corollary 4.5.10, ν_2 is a finite morphism, hence closed. This yields a factorization $\mathrm{CalCam}_2 \to Z \to \mathrm{Cam}_2$ with the second arrow a closed immersion and the first scheme-theoretically surjective. Let A be a strictly Henselian local ring and $\operatorname{Spec} A \to Z$ a morphism. The finiteness of ν_2 yields a diagram

$$\begin{tikzcd}
\operatorname{Spec} B \arrow[r] \arrow[d, "\psi"] & \mathrm{CalCam}_2 \arrow[d, "\nu_2"] \\
\operatorname{Spec} A \arrow[r] & \mathrm{Cam}_2
\end{tikzcd}$$

By [17, Tag 04GH], B is the product of local Henselian rings. By Proposition 4.5.9, the general fibers of ψ are length 2, corresponding to the two possible calibrating conics, so $\operatorname{Spec} B \simeq \operatorname{Spec} B_1 \sqcup \operatorname{Spec} B_2$. By Corollary 4.5.7, ψ is unramified, and thus (by [17, Tag 04GL]) restricts to a closed embedding on each $\operatorname{Spec} B_i$.

$$\begin{tikzcd}
\operatorname{Spec} B_i \arrow[r] \arrow[dr, hook] & \operatorname{Spec} B_1 \sqcup \operatorname{Spec} B_2 \arrow[r] \arrow[d, "\psi"] & \mathrm{CalCam}_2 \arrow[d] \\
& \operatorname{Spec} A \arrow[r] & \mathrm{Cam}_2
\end{tikzcd}$$

The involution described in Proposition 5.1.1 induces an isomorphism $f : \operatorname{Spec} B_1 \to \operatorname{Spec} B_2$. In other words both components map isomorphically to the image, so ν_2 is étale over Z, as claimed. □

5.2 An explicit description of the 2-1 cover

We now prove the following result starting from first principles.

Theorem 5.2.1. *The classical twisted pair cover*

$$\mathrm{SO}(3,\mathbf{R}) \times \mathbf{P}^2(\mathbf{R}) \to \mathbf{E}(\mathbf{R})$$

admits a unique extension to a finite étale cover $\pi\colon \widetilde{\mathbf{E}} \to \mathbf{E}$ *by a projective scheme. Moreover, the preimage of* $\widetilde{\mathbf{E}}$ *over* $\mathbf{E}^{\mathrm{sm}}$ *is naturally the moduli space of normalized camera pairs. Furthermore,*

1. *The variety* $\widetilde{\mathbf{E}}$ *is given by the vanishing of a single bilinear form on* $\mathbf{P}^3 \times \mathbf{P}^3$.
2. $\mathbf{E}$ *is naturally a hyperplane section of* $\mathrm{Sym}^2 \mathbf{P}^3$.
3. *There is a canonical isomorphism* $\mathrm{Pic}(\mathbf{E}) = \mathbf{Z}$ *and the full linear system* $|\mathscr{O}_{\mathbf{E}}(1)|$ *embeds* $\mathbf{E}$ *into* $\mathbf{P}^8$ *with degree* 10*. The ideal of* $\mathbf{E}$ *in* $\mathbf{P}^8$ *is defined by exactly* 10 *cubic forms.*

Proof. We prove this theorem over the course of this section. We construct $\widetilde{\mathbf{E}}$ and prove parts 1 and 2 in Lemma 5.2.10. We prove the smooth locus of $\widetilde{\mathbf{E}}$ is isomorphic to CalCam_2 in Proposition 5.2.11 and the properties of the cover follow from Theorem 5.1.2. Finally, we prove part 3 with Proposition 5.2.15. □

The embedding of E and the 10 Demazure cubics that define it are well known, but the methods to find these cubics is entirely novel. The description of E as a hyperplane section of $\mathrm{Sym}^2 \mathbf{P}^3$ has been previously worked out by Kileel, Fløystad, and Ottaviani, once again using entirely different methods. While Kileel–Fløystad–Ottaviani *use* the results of Demazure to establish a statement equivalent to parts 2 and 3 of the Main Theorem (by showing that E is a hyperplane section of the space of symmetric 4×4-matrices of rank at most 2, and expressing the locus determinantally), the ideas we describe here accomplish everything at once. (In fact, it was only after we worked out our results that we realized Kileel, Fløystad, and Ottaviani's description of the essential variety matches ours, given a suitable isomorphism between $\operatorname{Sym}^2 \mathbf{P}^3$ and the space of symmetric 4×4-matrices of rank at most 2; see Section 5.2.5.)

The variety $\widetilde{\mathrm{E}}$ seems to be a novel construciton. It gives a descriptoin of the compactification of the twisted pair double cover as vanishing locus associated to a bilinear form on $\mathbf{P}^3 \times \mathbf{P}^3$. Point corespondences also give bilinear relations on $\mathbf{P}^3 \times \mathbf{P}^3$, leading us to wonder about the numerical properties and reconstruction algorithms using this new description of the twisted pair cover.

In this section we will prove the main theorem, give explicit equations describing all the relevant varieties, and briefly discuss how this cover might help with reconstruction.

Remark 5.2.2. Everything in the Main Theorem is actually defined over $\mathbf{Z}[1/2]$, so the corresponding statements hold over any algebraically closed field of characteristic different from 2.

5.2.1 A reducibility criterion for simultaneous calibration

Recall that we denote the euclidean conic $V(x^2 + y^2 + z^2) \subset \mathbf{P}^2$ by D.

Definition 5.2.3. A camera configuration $\boldsymbol{\Phi}$ can be *simultaneously normalized* if there exists a smooth conic $C \subset \mathbf{P}^3$ such that $(\boldsymbol{\Phi}, (C, D))$ is a normalized camera configuration.

This following proposition and corollary turn out to be the key to the proof of the Main Theorem.

Proposition 5.2.4. *Let $P_1, \ldots, P_n : \mathbf{P}^3 \dashrightarrow \mathbf{P}^2$ be a collection of projective cameras, and let $C_{P_i} \subset \mathbf{P}^3$ be the cone over D under camera P_i. There exists an $\alpha \in \mathrm{PGL}_4$ such that $P_i\alpha$ is normalized for all i if and only if there is a smooth conic $Y \subset C_{P_1} \cap \cdots \cap C_{P_n}$.*

Proof. Fix $i \in \{1, \dots, n\}$. By 2.3.3, $\psi_i \circ \alpha$ is normalized if and only if $\psi_i \circ \alpha$ sends the conic Q^2_∞ isomorphically to $Q^2 \subset \mathbf{P}^2$. Hence, $\psi_i \circ \alpha$ is normalized if and only if $\alpha(Q^2_\infty)$ is a smooth hyperplane section of C_i. Therefore, if α is a simultaneous calibrator for $\psi_1, \dots, \psi_n$, then the intersection $C_1 \cap \cdots \cap C_n$ contains the smooth conic $\alpha(Q^2_\infty)$. To prove the converse, assume that $C_1 \cap \cdots \cap C_n$ contains a smooth conic Y. Then there is some automorphism β of $\mathbf{P}^3$ that sends the conic Y to Q^2_∞. Thus, $\alpha := \beta^{-1}$ has the desired properties. □

We can be more precise by describing the irreducible components. In the case that the intersection is two smooth conics, both components are smooth. In that case that the intersection is a doubled line and a conic, the conic must contain a point on the line between the two camera centers, and that point will map to the singular point of the image. More precisely, this lead us to the following characterization of calibrated camera configurations.

Theorem 5.2.5. *Let $J \in |\mathscr{O}_{\mathbf{P}^2 \times \mathbf{P}^2}(1,1)|$ be a fundamental divisor. Then J is an essential divisor if and only if the restriction $J \cap (D \times D) \in |\mathscr{O}_{\mathbf{P}^{1,1}}(2,2)|$ contains a smooth divisor Z in the linear system $|\mathscr{O}_{\mathbf{P}^1 \times \mathbf{P}^1}(1,1)|$ such that either*

a) Z contains the singular point of J, or

b) the residual curve $J \cap (D \times D) - Z$ is a smooth member of $|\mathscr{O}_{\mathbf{P}^1 \times \mathbf{P}^1}(1,1)|$.

We will use the following lemma in the proof.

Lemma 5.2.6. *Let $\psi_1, \psi_2 \colon \mathbf{P}^3 \dashrightarrow \mathbf{P}^2$ be two projective cameras with distinct centers p_1 and p_2 and let $\pi \colon \mathrm{Bl}_{p_1,p_2} \mathbf{P}^3 \to \mathbf{P}^2$ denote the blow-up morphism. For $i = 1, 2$, let $\widetilde{\psi}_i$ be the morphism obtained from the composition of π with ψ_i and let $C_i \subset \mathbf{P}^3$ be the cone over Q^2 under camera ψ_i. Then*

$$(\widetilde{\psi}_1 \times \widetilde{\psi}_2)(\pi^{-1}(C_1 \cap C_2)) = J \cap (D \times D), \tag{5.2.1.0.1}$$

where J denotes the multiview variety of ψ_1 and ψ_2.

Proof. Note that both sides of (5.2.1.0.1) are closed and by definition, $\psi_i(C_i - \{p_i\}) = D$, so the left-hand side is contained in the right-hand side. Thus to prove equality it suffices to show that each side has the same class in $\operatorname{Pic} D \times D = \mathbf{Z} \times \mathbf{Z}$, since two linearly equivalent divisors where one contains the other must be equal.

First consider $J \cap (D \times D)$. Since J is a $(1,1)$ hypersurface in $\mathbf{P}^2 \times \mathbf{P}^2$ and $D \subset \mathbf{P}^2$ is a degree 2 curve, the restriction of J to $D \times D$ is a $(2,2)$-divisor.

Now let us consider the left-hand side of (5.2.1.0.1). Note that for any two pairs of projective cameras with distinct centers (ψ_1, ψ_2) and (ψ_1', ψ_2'), the images $(\widetilde{\psi}_1 \times \widetilde{\psi}_2)(\pi^{-1}(C_1 \cap C_2))$ and $(\widetilde{\psi}_1' \times \widetilde{\psi}_2')(\pi^{-1}(C_1' \cap C_2'))$ are linearly equivalent. Thus, it is enough to prove the claim for a single choice of camera centers.

The intersection of two quadric cones in $\mathbf{P}^3$ is a degree 4 curve of arithmetic genus 1. Since $\widetilde{\psi}_1 \times \widetilde{\psi}_2$ is a birational map, so the arithmetic genus of the image can only increase. However, any proper component of a $(2,2)$ divisor must have arithmetic genus 0, so the image of $\pi^{-1}(C_1 \cap C_2)$ must be a $(2,2)$-divisor. □

Proof of Theorem 5.2.5

Proof. Let $\psi_1, \psi_2 \colon \mathbf{P}^3 \dashrightarrow \mathbf{P}^2$ be projective cameras such that $J := \overline{\operatorname{im}(\psi_1 \times \psi_2)}$. If φ_1 and φ_2 are any other pair of projective cameras such that $J := \overline{\operatorname{im}(\varphi_1 \times \varphi_2)}$, then there is some automorphism α of $\mathbf{P}^3$ such that $\psi_i = \psi_i' \circ \alpha$ for $i = 1, 2$. Thus, asking for J to be the joint image of *normalized cameras* is equivalent to asking whether ψ_1 and ψ_2 can be simultaneously normalized.

Proposition 5.2.4 states that ψ_1 and ψ_2 can be simultaneously normalized if and only if the intersection of the conic cones $C_1 := C_{\psi_1}$ and $C_2 := C_{\psi_2}$ contains a smooth conic. By the classification of conic cones given in Proposition 4.5.3, the intersection $C_1 \cap C_2$ is one of the following

1. an irreducible quartic curve of arithmetic genus 1,

2. a union of a twisted cubic and the line joining both camera centers,

3. a union of two smooth conics, or

4. a union of a smooth conic and a double line.

Hence, it suffices to show that in cases (1) and (2), $J \cap D \times D$ does not contain a $(1,1)$ divisor Z with the desired properties and that in cases (3) and (4), $J \cap D \times D$ does contain a $(1,1)$ divisor Z with the desired properties.

By (5.2.1.0.1), we need only show that $\widetilde{\psi_1} \times \widetilde{\psi_2}(\pi^{-1}(C_1 \cap C_2))$ does or does not have the desired properties.

Let us examine the map $\widetilde{\psi_1} \times \widetilde{\psi_2}$. The strict transform of the line $\overline{p_1 p_2}$ maps to the unique singular point of J, which we denote (c_1, c_2). Outside of this line, $\widetilde{\psi_1} \times \widetilde{\psi_2}$ is injective. Therefore, $J \cap D \times D$ is an irreducible $(2,2)$ curve in case (1).

In case (2), since the line connects both camera centers, the preimage of the line under π contains both exceptional divisors. The exceptional divisor above p_1 maps to $D \times \{c_2\}$ and the exceptional divisor above p_2 maps to $\{c_1\} \times D$. Therefore, in case (2), $J \cap (D \times D)$ is the union of $\{c_1\} \times D$, $D \times \{c_2\}$ and a $(1,1)$-curve (the image of the twisted cubic). Since the twisted cubic is irreducible, the image must be irreducible and hence smooth, so the birational map is an isomorphism. Then degree considerations show that the twisted cubic can only intersect the line at the cone points, so in particular, the image of the twisted cubic does not contain the singular point of J.

Now we consider cases (3) and (4). Any smooth conic contained in a quadric cone is a hyperplane section that does not contain the cone point. In particular any smooth conic in $C_1 \cap C_2$ does not contain p_1 or p_2 and intersects any resulting in exactly one point. Thus, $\widetilde{\psi_1} \times \widetilde{\psi_2}$ maps any smooth conic in $C_1 \cap C_2$ isomorphically to a smooth $(1,1)$ curve in $D \times D$. Therefore, case (3) gives (5.2.5). For case (4), we note that the line $\overline{p_1 p_2}$ is a ruling of each cone and therefore the smooth conic intersects $\overline{p_1 p_2}$ in a point different from p_1 and p_2. Hence, the image of the smooth conic is a smooth $(1,1)$ curve containing the singular point of J, so we obtain case (5.2.5). □

5.2.2 A double cover of the essential variety

By Theorem 5.2.5, the essential variety is contained in the locus of reducible divisors of $|\mathscr{O}_{\mathbf{P}^{1,1}}(2,2)|$ which contain a smooth member of $|\mathscr{O}_{\mathbf{P}^1\times\mathbf{P}^1}(1,1)|$. By analyzing this locus directly, we construct a double cover $\widetilde{\mathrm{E}}$ of the essential variety that, when restricted to real points gives the classical twisted pair cover $\mathrm{SO}_3(\mathbf{R})\times\mathbf{P^2}(\mathbf{R})\to\mathbf{E}(\mathbf{R})$.

The idea behind the double cover is simple, given our earlier results. By Theorem 5.2.5, essential divisors correspond to $(2,2)$ divisors on $D\times D$ that can be written as a union of two $(1,1)$ divisors (with additional properties). Thus, the double cover comes from keeping track of an ordered pair of the constituent $(1,1)$ divisors.

This reducibility criterion not only allows us to construct a double cover $\widetilde{\mathrm{E}}\to\mathbf{E}$, rather than just over $\mathbf{E}(\mathbf{R})$, it also allows us to give a presentation of $\widetilde{\mathrm{E}}$ as a hyperplane section in $\mathrm{Sym}^2(\mathbf{P}^3)$.

The restriction isomorphism

The euclidean conic D is an embedding of $\mathbf{P}^1\hookrightarrow\mathbf{P}^2$ via the three sections $x_0^2-x_1^2, 2x_0x_1, ix_0^2+ix_1^2\in\mathscr{O}_{\mathbf{P}^1}(2)$. Restricting $\mathbf{P}^2\times\mathbf{P}^2$ to the product $D\times D$ induces a restriction map

$$|\mathscr{O}_{\mathbf{P}^2\times\mathbf{P}^2}(1,1)|\to|\mathscr{O}_{\mathbf{P}^1\times\mathbf{P}^1}(2,2)|.$$

Explicitly this map is given by

$$\begin{pmatrix} x_0y_0\\ x_1y_0\\ x_2y_0\\ x_0y_1\\ x_1y_1\\ x_2y_1\\ x_0y_2\\ x_1y_2\\ x_2y_2 \end{pmatrix}\mapsto\begin{pmatrix} (x_0^2-x_1^2)(y_0^2-y_1^2)\\ (2x_0x_1)(y_0^2-y_1^2)\\ (ix_0^2+ix_1^2)(y_0^2-y_1^2)\\ (x_0^2-x_1^2)(2y_0y_1)\\ (2x_0x_1)(2y_0y_1)\\ (ix_0^2+ix_1^2)(2y_0y_1)\\ (x_0^2-x_1^2)(iy_0^2+iy_1^2)\\ (2x_0x_1)(iy_0^2+iy_1^2)\\ (ix_0^2+ix_1^2)(iy_0^2+iy_1^2) \end{pmatrix}$$

In fact, this map is an isomorphism.

Example 5.2.7. Consider the fundamental matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The corresponding section of $\mathscr{O}_{\mathbf{P}^2\times\mathbf{P}^2}(1,1)$ is $x_0y_0 + x_1y_1$. Restricting to $D \times D$ we get

$$\begin{aligned}(x_0^2 - x_1^2)(y_0^2 - y_1^2) + (2x_0y_0)(2x_1y_1) &= x_0^2y_0^2 + x_1^2y_1^2 - x_0^2y_1^2 - x_1^2y_0^2 + 4x_0y_0x_1y_1 \\ &= (x_0y_0 + x_1y_1)^2 - (x_0y_1 - x_1y_0)^2 \\ &= (x_0y_0 + x_1y_1 + x_0y_1 - x_1y_0)(x_0y_0 + x_1y_1 - x_0y_1 + x_1y_0)\end{aligned}$$

which is a section of $\mathscr{O}_{\mathbf{P}^1\times\mathbf{P}^1}(2,2)$ but also shows that it is the sum of two $\mathscr{O}_{\mathbf{P}^1\times\mathbf{P}^1}(1,1)$ divisors. This demonstrates that the section $x_0y_0 + x_1y_1$ corresponds to a simultaneously calibrated pair of cameras.

Remark 5.2.8. The condition that a divisor be rank deficient is naturally a condition on the linear system $|\mathscr{O}_{\mathbf{P}^2\times\mathbf{P}^2}(1,1)$, while the condition that it have a calibrating curve is an algebraic condition on $|\mathscr{O}_{D\times D}(2,2)|$. Exploiting the isomorphism between the two will give us a complete description of calibrated camera configurations.

In summary, simultaneously calibrated pairs of cameras correspond to divisors that are in the image of the sum map

$$\Sigma\colon |\mathscr{O}_{D\times D}(1,1)| \times |\mathscr{O}_{D\times D}(1,1)| \to |\mathscr{O}_{D\times D}(2,2)| \xrightarrow{\sim} |\mathscr{O}_{\mathbf{P}^2\times\mathbf{P}^2}(1,1)|,$$

where the last isomorphism is the inverse of the restriction map. Since essential divisors are also fundamental divisors, we need only consider the subscheme $\Sigma^{-1}(\mathbf{F})$.

Corollary 5.2.9. *The essential variety is closure of the locus of divisors J in $|\mathscr{O}_{\mathbf{P}^2\times\mathbf{P}^2}(1,1)|$ such that*

1. *J lies in the determinant hypersurface $\mathbf{F} \subset |\mathscr{O}_{\mathbf{P}^2\times\mathbf{P}^2}(1,1)|$;*

2. *the image of J under the isomorphism*

$$|\mathscr{O}_{\mathbf{P}^2\times\mathbf{P}^2}(1,1)| \overset{\sim}{\to} |\mathscr{O}_{\mathbf{P}^1\times\mathbf{P}^1}(2,2)|$$

   lies in the locus of divisors of the form $E+F$ with $E \in |\mathscr{O}_{\mathbf{P}^1\times\mathbf{P}^1}(1,1)|$ a smooth member and $F \in |\mathscr{O}_{\mathbf{P}^1\times\mathbf{P}^1}(1,1)|$ any member such that either F is smooth or the singular locus of F is contained in E.

Proof. This is just a restatement of Theorem 5.2.5 including the condition that J be a fundamental divisor, stated in the language of linear systems. □

Lemma 5.2.10.

1. *We may fix coordinates on $|\mathscr{O}_{D\times D}(1,1)|$ such that the map Σ sends a point $([a_0,a_1,a_2,a_3],[b_0,b_1,b_2,b_3])$ to the $(1,1)$ form*

$$\mathbf{x}^{\mathbf{T}}\begin{pmatrix} a_0b_0-a_1b_1-a_2b_2+a_3b_3 & a_0b_1+a_1b_0+a_2b_3+a_3b_2 & a_0b_2+a_2b_0-a_1b_3-a_3b_1 \\ a_0b_1+a_1b_0-a_2b_3-a_3b_2 & -a_0b_0+a_1b_1-a_2b_2+a_3b_3 & a_1b_2+a_2b_1+a_0b_3+a_3b_0 \\ a_0b_2+a_2b_0+a_1b_3+a_3b_1 & a_1b_2+a_2b_1-a_0b_3-a_3b_0 & -a_0b_0-a_1b_1+a_2b_2+a_3b_3 \end{pmatrix}\mathbf{y}. \tag{5.2.2.0.1}$$

2. *Under this choice of coordinates, the subscheme $\Sigma^{-1}(\mathbf{F})$ is the union of three irreducible components*

$$V(a_0^2+a_1^2+a_2^2+a_3^2) \cup V(b_0^2+b_1^2+b_2^2+b_3^2) \cup V(a_0b_0+a_1b_1+a_2b_2+a_3b_3).$$

   The first two components are the images of the morphisms

$$s_1\colon |\mathscr{O}_{D\times D}(1,0)| \times |\mathscr{O}_{D\times D}(0,1)| \times |\mathscr{O}_{D\times D}(1,1)| \longrightarrow |\mathscr{O}_{D\times D}(1,1)| \times |\mathscr{O}_{D\times D}(1,1)|, \quad \text{and}$$
$$s_2\colon |\mathscr{O}_{D\times D}(1,1)| \times |\mathscr{O}_{D\times D}(1,0)| \times |\mathscr{O}_{D\times D}(0,1)| \longrightarrow |\mathscr{O}_{D\times D}(1,1)| \times |\mathscr{O}_{D\times D}(1,1)|,$$

   respectively, that come from a further decomposition of divisors. The third irreducible component is a smooth divisor that is the preimage of $\mathbf{E}$ under Σ; we denote this smooth divisor $\widetilde{\mathbf{E}}$.

3. *The map* $\widetilde{\mathbf{E}} \to \mathbf{E}$ *factors through* $\operatorname{Sym}^2 \mathbf{P}^3$ *giving us the diagram*

$$\begin{tikzcd}
\widetilde{\mathbf{E}} \arrow[r] \arrow[d] & \mathbf{E} \arrow[r] \arrow[d] & \mathbf{P}^8 \arrow[d] \\
\mathbf{P}^3 \times \mathbf{P}^3 \arrow[r] & \operatorname{Sym}^2 \mathbf{P}^3 \arrow[r] & \mathbf{P}^9
\end{tikzcd}$$

Proof. For $(i,j) \in \{(0,1),(1,2),(2,0)\}$, we write $s_{ij} := x_i y_j + x_j y_i$ and $d_{ij} := x_i y_j - x_j y_i$. One can then check that on $D \times D$ we have the following identities

$$f_0 := \frac{s_{01}}{d_{20}} = \frac{s_{20}}{d_{01}}, \quad f_1 := \frac{s_{12}}{d_{01}} = \frac{s_{01}}{d_{12}}, \quad f_2 := \frac{s_{20}}{d_{12}} = \frac{s_{12}}{d_{20}}, \tag{5.2.2.0.2}$$

$$\frac{s_{01}s_{12}}{s_{20}} = -x_0y_0 + x_1y_1 - x_2y_2, \quad \frac{s_{12}s_{20}}{s_{01}} = -x_0y_0 - x_1y_1 + x_2y_2, \tag{5.2.2.0.3}$$

$$\frac{s_{20}s_{01}}{s_{12}} = +x_0y_0 - x_1y_1 - x_2y_2, \quad \text{and} \tag{5.2.2.0.4}$$

$$\frac{d_{01}d_{20}}{s_{12}} = \frac{d_{12}d_{01}}{s_{20}} = \frac{d_{20}d_{12}}{s_{01}} = x_0y_0 + x_1y_1 + x_2y_2. \tag{5.2.2.0.5}$$

For $i = 0,1,2$, let $\Gamma_i \subset D \times D$ denote the graph of the automorphism of D that multiplies x_i by -1 and leaves all other variables fixed, and let $\Gamma \subset D \times D$ denote the graph of the identity map. One can check that the functions f_i give the linear equivalence between the divisors Γ_i and Γ, i.e., we have $\operatorname{div}(f_i) = \Gamma_i - \Gamma$. Since $\Gamma, \Gamma_0, \Gamma_1$, and Γ_2 are all $(1,1)$-divisors on $D \times D$, the functions f_0, f_1, f_2 and $f_3 := 1$ form a basis for $|\mathscr{O}_{D\times D}(1,1)|$.

Now let us compute $\Sigma((\mathrm{a},\mathrm{b}))$. The points a and b correspond to the functions $a_0f_0 + a_1f_1 + a_2f_2 + a_3f_3$ and $b_0f_0 + b_1f_1 + b_2f_2 + b_3f_3$ respectively. The morphism Σ is the composition of two maps, the first of which corresponds to taking sums of divisors, or equivalently, multiplication of functions. Therefore

$$(\mathbf{a},\mathbf{b}) \mapsto \sum_{\mathbf{i}} \mathbf{a_i b_i f_i^2} + \sum_{\mathbf{i<j}} (\mathbf{a_i b_j} + \mathbf{a_j b_i})\mathbf{f_i f_j} \in |\mathscr{O}_{\mathbf{D\times D}}(\mathbf{2},\mathbf{2})|. \tag{5.2.2.0.6}$$

We may use the algebraic relations (5.2.2.0.2) to show that if we multiply any rational function f_if_j by the fixed $(1,1)$-form $x_0y_0 + x_1y_1 + x_2y_2$ we will obtain the following polynomials, which clearly lift to $(1,1)$-form on $\mathbf{P}^2 \times \mathbf{P}^2$.

$$f_0f_3 \Rightarrow d_{12} \quad f_1f_3 \Rightarrow d_{20} \quad f_2f_3 \Rightarrow d_{01} \quad f_0^2 \Rightarrow x_0y_0 - x_1y_1 - x_2y_2 \quad f_1^2 \Rightarrow -x_0y_0 + x_1y_1 - x_2y_2$$

$$f_1f_2 \Rightarrow s_{12} \quad f_0f_2 \Rightarrow s_{20} \quad f_0f_1 \Rightarrow s_{01} \quad f_3^2 \Rightarrow x_0y_0 + x_1y_1 + x_2y_2 \quad f_2^2 \Rightarrow -x_0y_0 - x_1y_1 + x_2y_2$$

By using the above polynomials and rewriting (5.2.2.0.6) in terms of the standard basis for $|\mathscr{O}_{\mathbf{P}^2\times\mathbf{P}^2}(1,1)|$, we compute that $\Sigma((\mathbf{a},\mathbf{b}))$ has the desired form.

Now we turn to proving the second part of the lemma. Since $\mathbf{F}$ corresponds exactly to the matrices of rank at most 2, $\Sigma^{-1}(\mathbf{F})$ is defined by the vanishing of the determinant of (5.2.2.0.1), which factors as

$$(a_0^2+a_1^2+a_2^2+a_3^2)(b_0^2+b_1^2+b_2^2+b_3^2)(a_0b_0+a_1b_1+a_2b_2+a_3b_3).$$

Let us consider the locus in $|\mathscr{O}_{D\times D}(1,1)|$ that is the image of $|\mathscr{O}_{D\times D}(1,0)|\times|\mathscr{O}_{D\times D}(0,1)|$, or equivalently the functions that witness the linear equivalence between Γ and a reducible divisor. Geometrically, this is the Segre embedding $\mathbf{P}^1\times\mathbf{P}^1\to\mathbf{P}^3$, so the image is cut out by a quadratic equation. It remains to determine which quadratic equation.

For any $\lambda,\mu\in\mathbb{Q}$ such that $1+\lambda^2+\mu^2=0$, we have the following relations:

$$\begin{aligned}
\operatorname{div}(f_1+\lambda f_2+\mu f_3) &= ([\ \ \mu,-\lambda,\ \ 1]\times D)+(D\times[-\mu,-\lambda,\ \ 1])-\Gamma,\\
\operatorname{div}(f_0+\lambda f_2+\mu f_3) &= ([-\lambda,-\mu,\ \ 1]\times D)+(D\times[-\lambda,\ \ \mu,\ \ 1])-\Gamma,\quad \text{and}\\
\operatorname{div}(f_0+\lambda f_1+\mu f_3) &= ([-\lambda,\ \ 1,\ \ \mu]\times D)+(D\times[-\lambda,\ \ 1,-\mu])-\Gamma.
\end{aligned}$$

Hence, the quadratic locus contains $[0,1,\lambda,\mu]$, $[1,0,\lambda,\mu]$ and $[1,\lambda,0,\mu]$. Thus, the image of s_1 must be defined by $a_0^2+a_1^2+a_2^2+a_3^2$ and the image of s_2 must be defined by $b_0^2+b_1^2+b_2^2+b_3^2$.

It remains to show that $\widetilde{\mathbf{E}}\to\mathbf{E}$ factors through $\operatorname{Sym}^2\mathbf{P}^3$. First note that the map $\Sigma:\mathbf{P}^3\times\mathbf{P}^3\to\mathbf{P}^9$ factors through $\operatorname{Sym}^2\mathbf{P}^3$ since

$$\Sigma((a,b))=\sum_i a_ib_if_i^2+\sum_{i<j}(a_ib_j+a_jb_i)f_if_j$$

is invariant under swapping a and b. The subvariety $\widetilde{\mathbf{E}}$ is defined by the vanishing of $a_0b_0+a_1b_1+a_2b_2+a_3b_3$ which is also invariant under the action of swapping. $\square$

5.2.3 $\widetilde{\mathbf{E}}$ *is a compactification of the moduli space of normalized camera pairs*

The divisor $\widetilde{\mathbf{E}}$ defined in 5.2.10 has a modular interpretation.

The singular locus F^{sing} of precisely the rank one matrices in $|\mathscr{O}_{\mathbf{P}^3\times\mathbf{P}^3}(1,1)|$. All rank one matrices can be written in the form vw^T. Examining the map Σ given by

$$\Sigma((a,b)) = \sum_i a_i b_i f_i^2 + \sum_{i<j}(a_i b_j + a_j b_i) f_i f_j$$

we see that the preimage $\Sigma^{-1}(F^{\text{sing}})$ is exactly the intersection of the diagonal embedding $\mathbf{P}^3 \hookrightarrow \mathbf{P}^3 \times \mathbf{P}^3$ with $\Sigma^{-1}(F)$. Recall that F^{sm} is the smooth rank two locus of the fundamental variety. Write $\widetilde{\mathbf{E}}^{\text{sm}}$ for the open subscheme that is $\widetilde{\mathbf{E}}$ minus the diagonal.

Proposition 5.2.11. *The scheme* $\widetilde{\mathbf{E}}^{\circ}$ *represents the functor* CalCam_2*, whose* T*-valued points are tuples* $(P, \varphi_1, \varphi_2, C)$ *where* $P \to T$ *is a Zariski form of* $\mathbf{P}^3_T$*,* $\varphi_i\colon P \dashrightarrow \mathbf{P}^2_T$*,* $i = 1, 2$*, are surjective linear projections, and* $C \subset P$ *is a* T*-flat family of curves of degree* 2 *such that* φ_i *is defined on* C *and the restriction* $\varphi_i|_C$ *factors through the closed immersion* $D_T \subset \mathbf{P}^2_T$.

Proof. Recall that Cam_2 is isomorphic to F^{sm}. By construction of $\widetilde{\mathbf{E}}^{\text{sm}}$ we have a 2-1 cover $\Sigma : \widetilde{\mathbf{E}}^{\text{sm}} \to \operatorname{Cam}_2$.

By forgetting the curve C the data $(P, \varphi_1, \varphi_2, C)$ induces a unique map $T \to \operatorname{Cam}_2$. We will show that this map lifts uniquely to a map to $\widetilde{\mathbf{E}}^{\text{sm}}$ and that this lift is compatible with base change.

$$\begin{tikzcd}
 & \widetilde{\mathbf{E}}^{\circ} \arrow[d] \\
T \arrow[r] \arrow[ur, dashed] & \operatorname{Cam}_2
\end{tikzcd}$$

By assumption the restriction of $\operatorname{im}(\Phi)$ to $D \times D$ contains the degree two curve $\operatorname{im}(C)$. By Proposition 5.2.5 the restriction $\operatorname{im}(\Phi) \cap (D \times D) \in |\mathscr{O}_{D\times D}(2,2)|$ factors into two divisors $\operatorname{im}(C)$ and Z both in $|\mathscr{O}_{D\times D}(1,1)|$. Making the non-conanical choice for $\operatorname{im}(C)$ to be come from the first component of the map Σ we get a unique lift to $\widetilde{\mathbf{E}}^{\circ}$ as desired. This construction is compatible with base change so we are done. $\square$

Remark 5.2.12. We do *not* assume in Proposition 5.2.11 that C is a family of smooth conics, just a degree 2 curve mapping into the standard plane conics under the two cameras. Thus, $\widetilde{\mathbf{E}}^{\mathrm{sm}}$ is a compactification of the classical twisted pair covering over the rank 2 complex points of the essential variety.

5.2.4 The geometry of $\operatorname{Sym}^2 \mathbf{P}^3$

In this section we describe a few basic results on the geometry of $\operatorname{Sym}^2 \mathbf{P}^3$.

Lemma 5.2.13. *The Picard group of* $\operatorname{Sym}^2 \mathbf{P}^3$ *is freely generated by a very ample class* $\mathscr{O}(1)$ *of degree* 10 *satisfying* $h^0(\mathscr{O}(1)) = 10$. *There is a canonical embedding* $\operatorname{Sym}^2 \mathbf{P}^3 \hookrightarrow \mathbf{P}^9$*, unique up to automorphism of* $\mathbf{P}^9$*, such that any hyperplane section has dimension* 5 *and degree* 10.

Proof. omitted (for now) □

Consider the natural morphism

$$s\colon \mathbf{P}^3 \times \mathbf{P}^3 \to \mathbf{P}^9$$

given in homogeneous coordinates by

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix} \mapsto \begin{pmatrix} a_0b_0 \\ a_0b_1 + a_1b_0 \\ a_0b_2 + a_2b_0 \\ a_0b_3 + a_3b_0 \\ a_1b_1 \\ a_1b_2 + a_2b_1 \\ a_1b_3 + a_3b_1 \\ a_2b_2 \\ a_2b_3 + a_3b_2 \\ a_3b_3 \end{pmatrix} =: \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \end{pmatrix}.$$

Lemma 5.2.14. *The scheme-theoretic image of the morphism* s *above is isomorphic to the scheme* $\operatorname{Sym}^2 \mathbf{P}^3$.

Proof. omitted (for now) □

Note that the x_i represent all $(\mathbf{Z}/2\mathbf{Z})$-invariant $(1,1)$ forms on $\mathbf{P}^3 \times \mathbf{P}^3$ (the $\mathbf{Z}/2\mathbf{Z}$ action interchanges a_i and b_i for all i). The following equations show that all $(\mathbf{Z}/2\mathbf{Z})$-invariant $(2,2)$ forms on $\mathbf{P}^3 \times \mathbf{P}^3$ can be written as quadratic polynomials in the x's (here $\{i,j,k,\ell\} = \{0,1,2,3\}$).

$$\begin{aligned}
a_i^2 b_j^2 + a_j^2 b_i^2 &= (a_i b_j + a_j b_i)^2 - 2 a_i a_j b_i b_j \\
a_i^2 b_j b_k + a_j a_k b_i^2 &= (a_i b_j + a_j b_i)(a_i b_k + a_k b_i) - a_i b_i (a_k b_j + a_j b_k) \\
a_i^2 b_i b_j + a_i a_j b_i^2 &= a_i b_i (a_i b_j + a_j b_i) \\
a_i a_j b_i b_k + a_i a_k b_i b_j &= a_i b_i (a_j b_k + a_k b_j) \\
a_i a_j b_k b_\ell + a_k a_\ell b_i b_j &= (a_i b_k + a_k b_i)(a_j b_\ell + a_\ell b_j) + (a_i b_\ell + a_\ell b_i)(a_j b_k + a_k b_j) \ \& \ - (a_i b_j + a_j b_i)(a_k b_\ell + a_\ell b_k)
\end{aligned}$$

Then since the space of $(\mathbf{Z}/2\mathbf{Z})$-invariant $(2,2)$ forms has the same dimension as the space quadratic polynomials in the x_i, there are no quadratic relations among the x_i. Counting dimensions again, we see that the space of cubic polynomials in the x_i's has dimension $\binom{9+3}{3} = 220$, whereas the space of $(\mathbf{Z}/2\mathbf{Z})$-invariant $(3,3)$-forms on $\mathbf{P}^3 \times \mathbf{P}^3$ has dimension $20 + \frac{20 \cdot 19}{2} = 210$. Thus, the ideal $\mathscr{I}$ defining the scheme-theoretic image of s has 10 cubic generators, which correspond to the following relations

$$\begin{aligned}
4 a_i a_j a_k b_i b_j b_k =& a_i b_i (a_j b_k + a_k b_j)^2 + a_j b_j (a_i b_k + a_k b_i)^2 + a_k b_k (a_i b_j + a_j b_i)^2 \\
& - (a_i b_j + a_j b_i)(a_i b_k + a_k b_i)(a_j b_k + a_k b_j) \\
2(a_i b_j + a_j b_i)(a_i a_j b_k b_\ell + a_k a_\ell b_i b_j) =& (a_i b_j + a_j b_i)((a_k b_i + a_i b_k)(a_\ell b_j + a_j b_\ell) + \\
& (a_k b_j + a_j b_k)(a_\ell b_i + a_i b_\ell) - (a_i b_j + a_j b_i)(a_k b_\ell + a_\ell b_k)) \\
=& 2(a_i b_i (a_k b_j + a_j b_k)(a_\ell b_j + a_j b_\ell) + a_j b_j (a_k b_i + a_i b_k)(a_\ell b_i + a_i b_\ell) \\
& - 2 a_i b_i a_j b_j (a_k b_\ell + a_\ell b_k))
\end{aligned}$$

Proposition 5.2.15. *Let* $\mathbf{P}^8 \subset \mathbf{P}^9$ *be a hyperplane and* $\operatorname{Sym}^2 \mathbf{P}^3 \hookrightarrow \mathbf{P}^9$ *the embedding of Lemma 5.2.14. Then the embedding* $E = \operatorname{Sym}^2 \mathbf{P}^3 \cap \mathbf{P}^8 \subset \mathbf{P}^8$ *of the hyperplane section is* *the complete linear system on $\mathscr{O}_E(1)$, and the homogeneous ideal of E in $\mathbf{P}^8$ is generated by precisely 10 cubic forms.*

Proof. omitted $\square$

5.2.5 The scheme of symmetric 4×4-matrices of rank at most 2 is $\operatorname{Sym}^2 \mathbf{P}^3$

In this section, we connect the work here with the results of [6]. The key is the following proposition. Let $X_n^{\le 2}$ denote the scheme of symmetric $n \times n$-matrices of rank at most 2. This is a cone in the affine space of symmetric $n \times n$-matrices whose ideal is generated by the 3×3-minors.

Proposition 5.2.16. *The map $\gamma\colon (a,b) \mapsto ab^T + ba^T$ defines an isomorphism*

$$\operatorname{Sym}^2 \mathbf{P}^{n-1} \to \mathbf{P}(X_n^{\le 2})$$

over the base $\operatorname{Spec} \mathbf{Z}[1/2]$. Moreover, the composition

$$\operatorname{Sym}^2 \mathbf{P}^{n-1} \to \mathbf{P}(X_n^{\le 2}) \to \mathbf{P}(\mathrm{M}_n) \cong \mathbf{P}^{n^2-1}$$

pulls back $\mathscr{O}_{\mathbf{P}^{n^2-1}}(1)$ to $\mathscr{O}_{\operatorname{Sym}^2 \mathbf{P}^{n-1}}(1)$.

Proof. First, it follows from the formula that the induced map $\mathbf{P}^{n-1} \times \mathbf{P}^{n-1} \to \mathbf{P}(\mathrm{M}_n)$ is bilinear in homogeneous coordinates, so it pulls back $\mathscr{O}_{\mathbf{P}(\mathrm{M}_n)}(1)$ to $\mathscr{O}_{\mathbf{P}^{n-1} \times \mathbf{P}^{n-1}}(1,1)$. This shows that $\gamma^* \mathscr{O}_{\mathbf{P}(\mathrm{M}_n)}(1) \cong \mathscr{O}_{\operatorname{Sym}^2 \mathbf{P}^{n-1}}(1)$.

Using Macaulay2 one can check that $\mathbf{P}(X_n^{\le 2})$ is reduced. Thus, it suffices to show that γ is bijective on geometric points.

Given an algebraically closed field k of characteristic not 2, a symmetric $n \times n$-matrix A is uniquely determined by its associated quadratic form $x \mapsto x \cdot (Ax)$. It is a standard fact that there is a basis in which this quadratic form is diagonal. That is, there is a matrix V and a diagonal matrix D such that

$$(Vx) \cdot (AVx) = x \cdot (Dx).$$

In matrix form, $V^TAV = D$. Moreover, the map $(x, y) \mapsto (V^Tx, V^Ty)$ defines a bijection between solutions to the equation $xy^T + yx^T = A$ and solutions to $xy^T + yx^T = D$. Thus, we may assume that

$$A = D = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Writing $a = (a_1, a_2, \dots, a_n)$ and $b = (b_1, b_2, \dots, b_n)$, the equation $ab^T + ba^T = D$ becomes the system

$$a_ib_j + a_jb_i = \begin{cases} \lambda_i & \text{if } i = j \text{ and } i \in \{1, 2\}, \\ 0 & \text{otherwise.} \end{cases}$$

Multiplying the above equation by $2b_i$ or $2a_i$ for $i \in \{1, 2\}$, we obtain the relations

$$\lambda_ib_j + 2a_jb_i^2 = \lambda_ia_j + 2a_ib_j^2 = 0.$$

Since at least one of $\lambda_1 = 2a_1b_1$ and $\lambda_2 = 2a_2b_2$ is nonzero and $2a_jb_j = 0$, these relations combined imply that $a_j = b_j = 0$ for all $j \neq 1, 2$. Similarly if $\lambda_1 = 0$ or $\lambda_2 = 0$, we have $a_1 = b_1 = 0$ or $a_2 = b_2 = 0$, respectively. If $\lambda_1\lambda_2$ is nonzero, then $b_1^2b_2^{-2} = a_1^2a_2^{-2} = -\lambda_1\lambda_2^{-1}$ and, for $i = 1, 2$, $2a_ib_i = \lambda_i$. Hence, in any of these three cases, the relations yield a unique point of $\operatorname{Sym}^2 \mathbf{P}^{n-1}$.

□

5.2.6 Using the cover for reconstruciton

We briefly illustrate how the 2-1 cover of E may provide another means for computing an essential matrix from 5 point correspondences. The basic idea is that $\widetilde{\mathrm{E}}$ is defined by a single bilinear form on $\mathbb{P}^3 \times \mathbb{P}^3$ and that the hyperplanes H_i induced by a point correspondence also pull back to bilinears forms in $\mathbb{P}^3 \times \mathbb{P}^3$. In contrast, **E** is cut out by 10 cubic equations on $\mathbf{P}^8$ and each hyperplane is a liner form.

Fixing coordinates on $\mathbf{P}^3 \times \mathbf{P}^3$ as in Lemma 5.2.10, the variety $\widetilde{\mathrm{E}}$ is cut out by $a_0b_0 + a_1b_1 + a_2b_2 + a_3b_3$.

5.2.7 Pulling back via $\Sigma : \mathbf{P}^3 \times \mathbf{P}^3 \to \mathbf{P}^8$

Normally a point correspondence $(p_0, p_1, p_2) \times (q_0, q_1, q_2) \in \mathbf{P}^2 \times \mathbf{P}^2$ implies a linear constraint

$$\sum_{i,j} p_i q_j x_i y_j = 0$$

on $|\mathscr{O}_{\mathbf{P}^2 \times \mathbf{P}^2}(1,1)|$. Pulling back via Σ gives

$$\begin{array}{lll}
p_0q_0(a_0b_0 - a_1b_1 - a_2b_2 + a_3b_3)+ & p_0q_1(a_0b_1 + a_1b_0 + a_2b_3 + a_3b_2)+ & p_0q_2(a_0b_2 + a_2b_0 - a_1b_3 - a_3b_1)+ \\
p_1q_0(a_0b_1 + a_1b_0 - a_2b_3 - a_3b_2)+ & p_1q_1(-a_0b_0 + a_1b_1 - a_2b_2 + a_3b_3)+ & p_1q_2(a_1b_2 + a_2b_1 + a_0b_3 + a_3b_0)+ \\
p_2q_0(a_0b_2 + a_2b_0 + a_1b_3 + a_3b_1)+ & p_2q_1(a_1b_2 + a_2b_1 - a_0b_3 - a_3b_0)+ & p_2q_2(-a_0b_0 - a_1b_1 + a_2b_2 + a_3b_3)
\end{array}$$

Rearranging terms, any point correspondence $(p_0, p_1, p_2) \times (q_0, q_1, q_2)$ pulls back to the hypersurface

$$\begin{array}{l}
a_0b_0(p_0q_0 - p_1q_1 - p_2q_2) + a_0b_1(p_0q_1 + p_1q_0) + a_0b_2(p_2q_0 + p_0q_2) + a_0b_3(p_1q_2 - p_2q_1)+ \\
a_1b_0(p_0q_1 + p_1q_0) + a_1b_1(-p_0q_0 + p_1q_1 - p_2q_2) + a_1b_2(p_1q_2 + p_2q_1) + a_1b_3(-p_0q_2 + p_2q_0)+ \\
a_2b_0(p_0q_2 + p_2q_0) + a_2b_1(p_1q_2 + p_2q_1) + a_2b_2(-p_0q_0 - p_1q_1 + p_2q_2) + a_2b_3(p_0q_1 - p_1q_0)+ \\
a_3b_0(p_1q_2 - p_2q_1) + a_3b_1(-p_0q_2 + p_2q_0) + a_3b_2(p_0q_1 - p_1q_0) + a_3b_3(p_0q_0 + p_1q_1 + p_2q_2)
\end{array}$$

If we have five point correspondences we get six bilinear equations in $\mathbf{P}^3 \times \mathbf{P}^3$ and hope to get 20 solutions over $\mathbf{C}$. We then apply Σ to recover the essential matrices we are after.

BIBLIOGRAPHY

[1] Sameer Agarwal, Hon-Leung Lee, Bernd Sturmfels, and Rekha R. Thomas, *On the existence of epipolar matrices*, Int. J. Comput. Vis. **121** (2017), no. 3, 403–415. MR 3606541

[2] Sameer Agarwal, Noah Snavely, Ian Simon, Steven M Seitz, and Richard Szeliski, *Building rome in a day*, Computer Vision, 2009 IEEE 12th International Conference on, IEEE, 2009, pp. 72–79.

[3] Chris Aholt, Bernd Sturmfels, and Rekha Thomas, *A Hilbert scheme in computer vision*, Canad. J. Math. **65** (2013), no. 5, 961–988. MR 3095002

[4] Luca Chiantini and Marc Coppens, *Grassmannians of secant varieties*, Forum Mathematicum, vol. 13, Berlin; New York: De Gruyter, c1989-, 2001, pp. 615–628.

[5] M. Demazure, *Sur deux problemes de reconstruction*, Tech. Report RR-0882, INRIA, July 1988.

[6] Gunnar Fløystad, Joe Kileel, and Giorgio Ottaviani, *The Chow form of the essential variety in computer vision*, Journal of Symbolic Computation (2017), To appear.

[7] Richard Hartley and Fredrik Kahl, *Critical configurations for projective reconstruction from multiple views*, International Journal of Computer Vision **71** (2007), no. 1, 5–47.

[8] Richard Hartley and Andrew Zisserman, *Multiple view geometry in computer vision*, second ed., Cambridge University Press, Cambridge, 2003, With a foreword by Olivier Faugeras. MR 2059248

[9] Brendan Hassett and Sándor J. Kovács, *Reflexive pull-backs and base extension*, J. Algebraic Geom. **13** (2004), no. 2, 233–247. MR 2047697

[10] D. Huybrechts, *Fourier-Mukai transforms in algebraic geometry*, Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, Oxford, 2006. MR 2244106

[11] Luc Illusie, *Complexe cotangent et déformations. I*, Lecture Notes in Mathematics, Vol. 239, Springer-Verlag, Berlin-New York, 1971. MR 0491680

[12] Joe Kileel, *Minimal problems for the calibrated trifocal variety*, SIAM Journal on Applied Algebra and Geometry, To appear.

[13] Joe Kileel, *Minimal problems for the calibrated trifocal variety*, SIAM Journal on Applied Algebra and Geometry **1** (2017), no. 1, 575–598.

[14] Zuzana Kukelova, Martin Bujnak, and Tomas Pajdla, *Automatic generator of minimal problem solvers*, Computer Vision–ECCV 2008 (2008), 302–315.

[15] Stephen Maybank, *Theory of reconstruction from image motion*, Springer Series in Information Sciences, vol. 28, Springer-Verlag, Berlin, 1993. MR 1226451

[16] David Nistér, *An efficient solution to the five-point relative pose problem*, IEEE transactions on pattern analysis and machine intelligence **26** (2004), no. 6, 756–770.

[17] The Stacks Project Authors, *Stacks project*, `http://stacks.math.columbia.edu`, 2017.

[18] Matthew Trager, Martial Hebert, and Jean Ponce, *The joint image handbook*, Proceedings of the IEEE international conference on computer vision, 2015, pp. 909–917.

[19] Angelo Vistoli, *Notes on grothendieck topologies, fibered categories and descent theory*, arXiv preprint math/0412512 (2004).