General models for rational cameras and the case of two-slit projections

Matthew Trager[1,2] Bernd Sturmfels[3] John Canny[3] Martial Hebert[4] Jean Ponce[1,2]

[1]École Normale Supérieure, CNRS, PSL Research University [2]Inria [3]UC Berkeley [4]Carnegie Mellon University

Abstract

The rational camera model recently introduced in [18] provides a general methodology for studying abstract nonlinear imaging systems and their multi-view geometry. This paper builds on this framework to study "physical realizations" of rational cameras. More precisely, we give an explicit account of the mapping between between physical visual rays and image points (missing in the original description), which allows us to give simple analytical expressions for direct and inverse projections. We also consider "primitive" camera models, that are orbits under the action of various projective transformations, and lead to a general notion of intrinsic parameters. The methodology is general, but it is illustrated concretely by an in-depth study of two-slit cameras, that we model using pairs of linear projections. This simple analytical form allows us to describe models for the corresponding primitive cameras, to introduce intrinsic parameters with a clear geometric meaning, and to define an epipolar tensor characterizing two-view correspondences. In turn, this leads to new algorithms for structure from motion and self-calibration.

1. Introduction

The past 20 years have witnessed steady progress in the construction of effective geometric and analytical models of more and more general imaging systems, going far beyond classical pinhole perspective (e.g., [1, 2, 5, 6, 12, 13, 16, 18, 19, 21, 25, 26]). In particular, it is now recognized that the *essential* part of any imaging system is the mapping between scene points and the corresponding light rays. The mapping between the rays and the points where they intersect the retina plays an *auxiliary* role in the image formation process. For pinhole cameras, for example, all retinal planes are geometrically equivalent since the corresponding image patterns are related to each other by projective transforms. Much of the recent work on general camera models thus focuses primarily on the link between scene points and the *line congruences* [9, 14] formed by the corresponding rays, in a purely projective setting [1, 2]. The *rational camera* model of Ponce, Sturmfels and Trager [18] is a recent instance of this approach, and provides a unified algebraic framework for studying a very large class of imaging systems and the corresponding multi-view geometry. It is, however, *abstract*, in the sense that the mapping between visual rays and image points is left unspecified. This paper provides a *concrete* embedding of this model, by making the mapping from visual rays to a retinal plane explicit, and thus identifying physical instances of rational cameras. The imaging devices we consider are in fact the composition of three maps: the first two are purely geometric, and map scene points onto visual rays, then rays onto the points where they intersect a retinal plane. The last map is analytical: given a coordinate system on the retina, it maps image points onto their corresponding coordinates (see Figure 1). In particular, by introducing the notion of "primitive" camera model as the orbit of a camera under the action of projective, affine, similarity, and euclidean transformations, we can generalize many classical attributes of pinhole projections, such as intrinsic coordinates and calibration matrices. Our methodology applies to arbitrary (rational) cameras, but it is illustrated concretely by an in-depth study of two-slit cameras [2, 6, 12, 26], which we describe using a pair of linear 2×4 projection matrices. This simple form allows us to describe models for the corresponding primitive cameras, to introduce intrinsic parameters with a clear geometric meaning, and to define an *epipolar tensor* characterizing two-view correspondences. In turn, this leads to new algorithms for structure from motion and self-calibration.

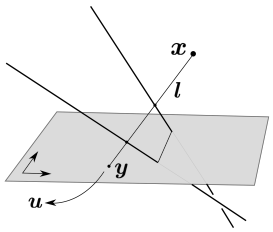

Figure 1. A general camera associates a scene point $\boldsymbol{x}$ with a visual ray $\boldsymbol{l}$, then maps the ray $\boldsymbol{l}$ to its intersection $\boldsymbol{y}$ with some retinal plane $\boldsymbol{\pi}$, and finally uses a projective coordinate system on $\boldsymbol{\pi}$ to express $\boldsymbol{y}$ as a point $\boldsymbol{u}$ in $\mathbb{P}^2$.

1.1. Background

Pinhole (or *central*) perspective has served since the XV$^{\text{th}}$ Century as an effective model of physical imaging systems, from the camera obscura and the human eye to the daguerreotype and today's digital photographic or movie cameras. Under this model, a scene point $\boldsymbol{x}$ is first mapped onto the unique line $\boldsymbol{l}$ joining $\boldsymbol{x}$ to $\boldsymbol{c}$, which is itself then mapped onto its intersection $\boldsymbol{y}$ with some retinal plane $\boldsymbol{\pi}$.[1] All retinas are projectively equivalent, and thus the *essential* part of the imaging process is the map $\lambda_{L_c} : \mathbb{P}^3 \setminus \boldsymbol{c} \to L_c$ associating a point with the corresponding visual ray in the *bundle* L_c of all lines through the pinhole $\boldsymbol{c}$ [16]. Similarly, many non-central cameras can be imagined (and actually constructed [20, 21, 24]) by replacing the line bundle L_c with a more general family of lines. For example, Pajdla [13] and Batog *et al.* [2] have considered cameras that are associated with *linear congruences*, *i.e.*, two-dimensional families of lines that obey linear constraints. This model applies two-slit [12, 23, 26], pushbroom [6], and pencil [25] cameras. For these devices, the projection can be described by a projective map A so that each point $\boldsymbol{x}$ is mapped to the line joining $\boldsymbol{x}$ and $A\boldsymbol{x}$. Recently, a generalization of this model to non-linear congruences [9, 14] has been introduced by Ponce, Sturmfels and Trager [18]. They study algebraic properties of the essential map that associates scene points to the corresponding viewing rays, and provide a general framework for studying multi-view geometry in this setting. On the other hand, they do not focus on the *auxiliary* part of the imaging process, that associates viewing rays with image coordinates, leaving this map left unspecified as an arbitrary birational map from a congruence to $\mathbb{P}^2$. We provide in Section 2 the missing link with a concrete retinal plane and (pixel) coordinate systems, which is key for defining intrinsic parameters that have physical meaning, and for developing actual algorithms for multi-view reconstruction.

1.2. Contributions

From a theoretical standpoint: 1) we present a concrete embedding of the abstract rational cameras introduced in [18], deriving general direct and inverse projection formulas for these cameras, and 2) we introduce the notion of *primitive camera models*, that are orbits of rational cameras under the action of the projective, affine, and euclidean and similarity groups, and lead to the generalization familiar concepts such as intrinsic camera parameters. From a more concrete point of view, we use this model for an in-depth study of two-slit cameras [2, 12, 26]. Specifically, 1) we introduce a pair of *calibration matrices*, which generalize to our setting the familiar upper-triangular matrices associated with pinhole cameras, and can be identified with the orbits of two-slit cameras under similarity transforms; 2) we define the *epipolar tensor*, that encodes the projective geometry of a pair of two-slit cameras, and generalizes the traditional fundamental matrix; and 3) use these results to describe algorithms for structure from motion and self-calibration.

To improve readability, most proofs and technical material are deferred to the supplementary material.

[1] This is of course an *abstraction* of a physical camera, where a "central ray" is picked inside the finite light beams of physical optical systems. As noted in [2], this does not take anything away from the idealized model adopted here. For example, a camera equipped with a lens is effectively modeled, ignoring distortions, by a pinhole camera.

1.3. Notation and elementary line geometry

Notation. We use bold letters for matrices and vectors ($\boldsymbol{A}$, $\boldsymbol{x}$, $\boldsymbol{u}$, *etc.*) and regular fonts for coordinates (a_i, x_i, u_i, *etc.*). Homogeneous coordinates are lowercase and indexed from one, and both points and planes are column vectors, *e.g.*, $\boldsymbol{x} = (x_1, x_2, x_3, x_4)^T$, $\boldsymbol{u} = (u_1, u_2, u_3)^T$. For projective objects, equality is always up to some scale factor. We identify $\mathbb{R}^3$ with points $(x_1, x_2, x_3, 1)^T$ in $\mathbb{P}^3$, and adopt the standard euclidean metric. We also use the natural hierarchy among projective transformations by the nested subgroups (dubbed from now on *natural projective subgroups*) formed by euclidean (orientation-preserving) transformations, similarities (euclidean transformations plus scaling), affine transformations, and projective ones. We will assume that the reader is familiar with their analytical form.

Line geometry. The set of all lines of $\mathbb{P}^3$ forms a four-dimensional variety, the *Grassmannian* of lines, denoted by $\mathrm{Gr}(1, 3)$. The line passing through two distinct points $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{P}^3$ can be represented using *Plücker coordinates* by writing $\boldsymbol{l} = (l_{41}, l_{42}, l_{43}, l_{23}, l_{31}, l_{12})$ where $l_{ij} = x_i y_j - x_j y_i$. This defines a point in $\mathbb{P}^5$ that is independent of the choice of $\boldsymbol{x}$ and $\boldsymbol{y}$ along $\boldsymbol{l}$. Moreover, the coordinates of any line $\boldsymbol{l}$ satisfy the constraint $\boldsymbol{l} \cdot \boldsymbol{l}^* = 0$, where $\boldsymbol{l}^* = (l_{23}, l_{31}, l_{12}, l_{41}, l_{42}, l_{43})$ is the *dual line* for $\boldsymbol{l}$. Conversely any point in $\mathbb{P}^5$ with this property represents a line, so $\mathrm{Gr}(1, 3)$ can be identified with a quadric hypersurface in $\mathbb{P}^5$. The *join* ($\vee$) and *meet* ($\wedge$) operators are used to indicate intersection and span of linear spaces (points, lines, planes). For example, $\boldsymbol{x} \vee \boldsymbol{y}$ denotes the unique line passing through $\boldsymbol{x}$ and $\boldsymbol{y}$. To give analytic formulae for these operators, it useful to introduce the *primal and dual Plücker matrices* for a line $\boldsymbol{l}$: these are defined respectively as $\boldsymbol{L} = \boldsymbol{x}\boldsymbol{y}^T - \boldsymbol{y}\boldsymbol{x}^T$ and $\boldsymbol{L}^* = \boldsymbol{u}\boldsymbol{v}^T - \boldsymbol{v}\boldsymbol{u}^T$, where $\boldsymbol{x}, \boldsymbol{y}$ are any two points on $\boldsymbol{l}$, and $\boldsymbol{u}, \boldsymbol{v}$ are any two planes containing $\boldsymbol{l}$ (so $\boldsymbol{l} = \boldsymbol{x} \vee \boldsymbol{y} = \boldsymbol{u} \wedge \boldsymbol{v}$). With these definitions, the join $\boldsymbol{l} \vee \boldsymbol{z}$ of $\boldsymbol{l}$ with a point $\boldsymbol{z}$ is the plane with coordinates $\boldsymbol{L}^*\boldsymbol{z}$, while the meet $\boldsymbol{l} \wedge \boldsymbol{w}$ with a plane $\boldsymbol{w}$ is the point with coordinates $\boldsymbol{L}\boldsymbol{w}$.

2. A physical model for rational cameras

For convenience of the reader, we briefly recall in Section 2.1 some results from [18]. We then proceed to new results in Sections 2.2 and 2.3.

2.1. Cameras and line congruences

Congruences. A *line congruence* is a two-dimensional family of lines in $\mathbb{P}^3$, or a surface in $\mathrm{Gr}(1,3)$ [9]. Since a general camera produces a two-dimensional image, the set of viewing rays captured by the imaging device will form a congruence. We will only consider *algebraic* congruences that are defined by polynomial equations in $\mathrm{Gr}(1,3)$. Every such congruence L is associated with two non-negative integers (α,β): the *order* α is the number of rays in L that pass through a generic point of $\mathbb{P}^3$, while the *class* β is the number of lines in L that lie in a generic plane. The pair (α,β) is the *bidegree* of the congruence. The set of points that *do not* belong to α distinct lines is the *focal locus* $F(L)$.

Geometric rational cameras. Order one congruences (or $(1,\beta)$-congruences) are a natural geometric model for most imaging systems [1, 2, 16]. Indeed, a $(1,\beta)$-congruence L defines a *rational map* $\lambda_L : \mathbb{P}^3 \dashrightarrow \mathrm{Gr}(1,3)$ associating a generic point $\boldsymbol{x}$ in $\mathbb{P}^3$ with the unique $\lambda_L(\boldsymbol{x})$ line in L that contains $\boldsymbol{x}$.[2] All possible maps of this form are described in [18]. The Plücker coordinates of a line $\lambda_L(\boldsymbol{x})$ are homogeneous polynomials (or *forms*) of degree $\beta+1$ in the coordinates of $\boldsymbol{x}$. For example, a family of $(1,\beta)$-congruences are defined by an algebraic curve γ of degree β and a line δ intersecting γ in $\beta-1$ points [14]. Its rays are the common transversals to γ and δ. The mapping $\lambda_L : \mathbb{P}^3 \dashrightarrow \mathrm{Gr}(1,3)$ can be expressed, using appropriate coordinates of $\mathbb{P}^3$, by

$$\lambda_L(\boldsymbol{x}) = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \vee \begin{bmatrix} x_1 f(x_1,x_2) \\ x_2 f(x_1,x_2) \\ g(x_1,x_2) \\ h(x_1,x_2) \end{bmatrix}, \tag{1}$$

where f, g and h are respectively forms of degree $\beta-1$, β and β. The remaining $(1,\beta)$-congruences for $\beta > 0$ correspond to degenerations of this case, or to transversals to twisted cubics. These can be parameterized in a similar manner [18].

Photographic rational cameras. Composing $\lambda_L : \mathbb{P}^3 \dashrightarrow \mathrm{Gr}(1,3)$ with an *arbitrary* birational map $\nu : L \dashrightarrow \mathbb{P}^2$ determines a map $\psi = \nu \circ \lambda_L : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$, which is taken as the definition of a general "photographic" rational camera in [18]. Although this model leads to effective formulations of multi-view constraints, it is *abstract*, with no explicit link between $\boldsymbol{u} = \psi(\boldsymbol{x})$ and the actual image point where the ray $\lambda_L(\boldsymbol{x})$ intersects the sensing array.

[2] A rational map between projective spaces is a map whose coordinates are homogeneous polynomial functions. In particular, this map is not well-defined on points where all these polynomials vanish. Here and in the following, we use a dashed arrow to indicate a rational map that is only well-defined on a dense, open set.

2.2. Rational cameras with retinal planes

In this paper, we adapt the framework from [18] for studying *physical instances* of photographic cameras. We will refer to the map λ_L defined by a $(1,\beta)$-congruence L as the *essential* part of the imaging process (or an "essential camera" for short). The *auxiliary* part of the process requires choosing a *retinal plane* $\boldsymbol{\pi}$ in $\mathbb{P}^3$. This determines a map $\mu_{\boldsymbol{\pi}} : L \dashrightarrow \boldsymbol{\pi}$ that associates any ray $\boldsymbol{l}$ in L not lying in $\boldsymbol{\pi}$ with its intersection $\boldsymbol{y} = \boldsymbol{l} \wedge \boldsymbol{\pi}$ with that plane. For a generic choice of $\boldsymbol{\pi}$, this map is not defined on β lines, since β lines from L will lie on $\boldsymbol{\pi}$ (recall the definition of β from Section 2.1). Together, λ_L and $\mu_{\boldsymbol{\pi}}$ (or, equivalently, L and $\boldsymbol{\pi}$) define a mapping $\mu_{\boldsymbol{\pi}} \circ \lambda_L : \mathbb{P}^3 \dashrightarrow \boldsymbol{\pi}$ that can be taken as a definition of a *geometric rational camera*. Finally, an analytic counterpart of this model is obtained by picking a coordinate system (π) on $\boldsymbol{\pi}$, that corresponds to the pixel grid where radiometric measurements are obtained. Representing (π) as a 4×3 matrix $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3]$, where columns correspond to basis points, the coordinates $\boldsymbol{u}$ in $\mathbb{P}^2$ of any point $\boldsymbol{y}$ on $\boldsymbol{\pi}$ are given by $\boldsymbol{u} = \boldsymbol{Y}^\dagger \boldsymbol{y}$, where $\boldsymbol{Y}^\dagger$ is a 3×4 matrix in the three-parameter set of pseudoinverses of $\boldsymbol{Y}$ (see for example [16]). Note that both $\mu_{\boldsymbol{\pi}}$ and $\boldsymbol{Y}^\dagger$ are *linear*, and in fact a simple calculation shows that $\boldsymbol{N} = \boldsymbol{Y}^\dagger \circ \mu_{\boldsymbol{\pi}}$ is described by the 3×6 matrix

$$\boldsymbol{N} = \boldsymbol{Y}^\dagger \circ \mu_{\boldsymbol{\pi}} = \begin{bmatrix} (\boldsymbol{y}_2 \vee \boldsymbol{y}_3)^{*T} \\ (\boldsymbol{y}_3 \vee \boldsymbol{y}_1)^{*T} \\ (\boldsymbol{y}_1 \vee \boldsymbol{y}_2)^{*T} \end{bmatrix}. \tag{2}$$

This expression does not depend on L: it represents a linear map $\mathbb{P}^5 \dashrightarrow \mathbb{P}^2$ that associates a generic line $\boldsymbol{l}$ in $\mathbb{P}^3$ with the coordinates $\boldsymbol{u}$ in $\mathbb{P}^2$ of its intersection with $\boldsymbol{\pi}$.

Example 1. Consider the plane $\boldsymbol{\pi} = \{x_3 - x_4 = 0\}$ in $\mathbb{P}^3$, equipped with the reference frame given by $\boldsymbol{y}_1 = (1,0,0,0)^T$, $\boldsymbol{y}_2 = (0,1,0,0)^T$, $\boldsymbol{y}_3 = (0,0,1,1)^T$ (with fixed relative scale). The matrix $\boldsymbol{N}$ takes in this case the form

$$\boldsymbol{N} = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}. \tag{3}$$

The null space of the corresponding linear projection is $\{p_{41} - p_{31} = p_{42} + p_{23} = p_{43} = 0\}$, which characterizes lines contained in $\boldsymbol{\pi}$. ♢

In summary, a complete analytical model of a physical photographic rational camera is a map $\psi : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ that is the composition of the essential map $\lambda_L : \mathbb{P}^3 \dashrightarrow \mathrm{Gr}(1,3)$, associated with a $(1,\beta)$-congruence L, and the linear map $\boldsymbol{N}$ of (2):

$$\boldsymbol{x} \mapsto \boldsymbol{u} = \psi(\boldsymbol{x}) = \boldsymbol{N}\lambda_L(\boldsymbol{x}). \tag{4}$$

By construction, the coordinates of $\psi(\boldsymbol{x})$ are forms of degree $\beta+1$ in $\boldsymbol{x}$. One could of course define directly a rational camera in terms of such forms (and indeed rational cameras are commonly used in photogrammetry [8]). Note,

however, that arbitrary rational maps do not, in general, define a camera, since the pre-image of a generic point $\boldsymbol{u}$ in $\mathbb{P}^2$ should be a line. From (4) we also easily recover the *inverse projection* $\chi : \mathbb{P}^2 \dashrightarrow \mathrm{Gr}(1,3)$ mapping a point in $\mathbb{P}^2$ onto the corresponding viewing ray in L:

$$\boldsymbol{u} \mapsto \boldsymbol{l} = \chi(\boldsymbol{u}) = \lambda_L(\boldsymbol{Y}\boldsymbol{u}). \tag{5}$$

The Plücker coordinates of $\boldsymbol{l}$ are also forms of degree $\beta + 1$ in $\boldsymbol{u}$. Equation (5) can be used to define epipolar or multiview constraints for point correspondences, since it allows one to express incidence constraints among viewing rays in terms of the corresponding image coordinates [21]. See also Section 4.

To conclude, let us point out that although a rational camera $\psi : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ was introduced in terms of a congruence L, a retinal plane $\boldsymbol{\pi}$, and a projective reference frame on $\boldsymbol{\pi}$, the retinal plane is often not (completely) determined given ψ. This is well known for pinhole cameras, and we will argue that a similar ambiguity arises for two-slit cameras. On the other hand, it is easy to see that the congruence L and the global mapping $\boldsymbol{N}$ can always be recovered from the analytic expression of the camera: the congruence L is in fact determined by the pre-images $\psi^{-1}(\boldsymbol{u})$ under ψ of points $\boldsymbol{u}$ in $\mathbb{P}^2$, while $\boldsymbol{N}$ is described by $\boldsymbol{N}(\psi^{-1}(\boldsymbol{u})) = \boldsymbol{u}$.

2.3. Primitive camera models

The orbit of a rational camera under the action of any of the four natural projective subgroups defined earlier is a family of cameras that are geometrically and analytically equivalent under the corresponding transformations. Each orbit will be called a *primitive camera model* for the corresponding subgroup, and we will attach to every such model, whenever appropriate and possible, a particularly simple analytical representative. A primitive camera model exhibits (analytical and geometric) *invariants*, that is, properties of all cameras in the same orbit that are preserved by the associated group of transformations. For example, the *intrinsic parameters* of a pinhole perspective camera are invariants for the corresponding euclidean primitive model; we will argue in the next section that this definition can in fact be applied for arbitrary rational cameras. Another familiar example is the *projection of the absolute conic*, which is an invariant for the similarity orbit of any imaging system. Indeed, the absolute conic, defined in any euclidean coordinate system by $\{x_4 = x_1^2 + x_2^2 + x_3^2 = 0\}$, is fixed by all similarity transformations. This property is commonly used for the self-calibration of pinhole perspective cameras [11, 15, 17, 22] but it can be used for more general cameras as well (see Section 4).

To further illustrate these concepts, we describe classical primitive models for pinhole projections. We consider the standard pinhole projection $\psi(\boldsymbol{x}) = (x_1, x_2, x_3)^T$, associated with the 3×4-matrix $[\mathrm{Id} \,|\, \boldsymbol{0}]$. Its projective orbit consists of all linear projections $\mathbb{P}^3 \dashrightarrow \mathbb{P}^2$, corresponding to 3×4 matrices of full rank, so pinhole cameras form a primitive projective model. The affine orbit of ψ is the family of *finite cameras*, *i.e.*, cameras of the form $[\boldsymbol{A} \,|\, \boldsymbol{b}]$, where $\boldsymbol{A}$ is a 3×3 full-rank matrix. The orbit of ψ under similarities or under euclidean transformations is the set of cameras of the form $[\boldsymbol{R} \,|\, \boldsymbol{t}]$ where $\boldsymbol{R}$ is a rotation matrix. The euclidean and similarity orbits in this case coincide: this is of course related to the scale ambiguity in pictures taken with perspective pinhole cameras. The euclidean/similarity invariants of a finite camera can be expressed as entries of a 3×3 upper-triangular *calibration matrix* $\boldsymbol{K}$: this follows from the uniqueness of the RQ-decomposition $\boldsymbol{A} = \boldsymbol{K}\boldsymbol{R}$. Primitive models for cameras "at infinity" include *affine*, *scaled orthographic*, and *orthographic* projections: these are defined by the orbits under natural projective groups of $\psi'(\boldsymbol{x}) = (x_1, x_2, x_4)^T$. Here the similarity and euclidean orbits are *different* (indeed, orthographic projections preserve distances). The intrinsic parameters can be expressed as entries of a 2×2 calibration matrix (see [7, Sec. 6.3]).

The remaining part of the paper focuses on a particular class of rational cameras, namely two-slit cameras, for which the general concepts introduced above can be easily applied. Two-slit cameras correspond in fact to congruences of class $\beta = 1$, so they can be viewed as the "simplest" rational cameras after pinhole projections.

3. Two-slit cameras: a case study

Two-slit cameras [3, 6, 12, 23, 26] record viewing rays that are the transversals to two fixed skew lines (the slits). The bidegree of the associated congruence is $(1,1)$, since a generic plane will contain exactly one transversal to the slits, namely the line joining the points where the slits intersect the plane. This type of congruence is the intersection of $\mathrm{Gr}(1,3)$ with a 3-dimensional linear space in $\mathbb{P}^5$ [2]. The map λ_L associated with the slits $\boldsymbol{l}_1, \boldsymbol{l}_2$ is

$$\boldsymbol{x} \mapsto \boldsymbol{l} = (\boldsymbol{x} \vee \boldsymbol{l}_1) \wedge (\boldsymbol{x} \vee \boldsymbol{l}_2). \tag{6}$$

We now fix a retinal plane $\boldsymbol{\pi}$ equipped with the coordinate system defined by $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3]$. A rational camera $\psi : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ is obtained by composing (6) with the 3×6 matrix $\boldsymbol{N}$ as in (2). A simple calculation shows that the resulting map can be written compactly as

$$\boldsymbol{x} \mapsto \boldsymbol{u} = \begin{bmatrix} \boldsymbol{x}^T \boldsymbol{P}_1^* \boldsymbol{S}_1 \boldsymbol{P}_2^* \boldsymbol{x} \\ \boldsymbol{x}^T \boldsymbol{P}_1^* \boldsymbol{S}_2 \boldsymbol{P}_2^* \boldsymbol{x} \\ \boldsymbol{x}^T \boldsymbol{P}_1^* \boldsymbol{S}_3 \boldsymbol{P}_2^* \boldsymbol{x} \end{bmatrix}, \tag{7}$$

where $\boldsymbol{P}_1^*, \boldsymbol{P}_2^*$ are the dual Plücker matrices associated with the slits $\boldsymbol{l}_1, \boldsymbol{l}_2$, while $\boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{S}_3$ are Plücker matrices for $\boldsymbol{y}_2 \vee \boldsymbol{y}_3$, $\boldsymbol{y}_3 \vee \boldsymbol{y}_1$, $\boldsymbol{y}_1 \vee \boldsymbol{y}_2$ respectively [3, 26].

Example 2. Let us fix the slits to be the lines $\boldsymbol{l}_1 = \{x_1 = x_3 = 0\}, \boldsymbol{l}_2 = \{x_2 = x_3 + x_4 = 0\}$. The corresponding

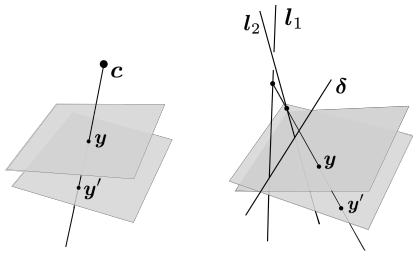

Figure 2. *Left:* A pinhole $\boldsymbol{c}$ induces a homography between any two retinal planes not containing $\boldsymbol{c}$. *Right:* Two skew lines $\boldsymbol{l}_1, \boldsymbol{l}_2$ induce a homography between planes intersecting at a transversal $\boldsymbol{\delta}$ to $\boldsymbol{l}_1, \boldsymbol{l}_2$.

essential map $\mathbb{P}^3 \dashrightarrow \mathrm{Gr}(1,3)$ is given by

$$\lambda(\boldsymbol{x}) = (x_1(x_3+x_4), x_2x_3, x_3(x_3+x_4), x_2x_3, 0, -x_1x_2)^T \tag{8}$$

Composing with the projection $\boldsymbol{N}$ in Example 1, we obtain the formula for a rational camera with slits $\boldsymbol{l}_1, \boldsymbol{l}_2$:

$$\boldsymbol{u} = \psi(\boldsymbol{x}) = (x_1(x_3+x_4), 2x_2x_3, x_3(x_3+x_4))^T \tag{9}$$

The same expression can be deduced from (7). ♢

It is noted in [3, 26] that using a different retinal plane $\boldsymbol{\pi}$ in (7) corresponds, in general, to composing the rational camera $\psi : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ with a *quadratic* change of image coordinates $\mathbb{P}^2 \dashrightarrow \mathbb{P}^2$. However, this transformation is in fact *linear* when $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ intersect along a transversal to the slits. This follows from the following property:

Lemma 1. *Let $\boldsymbol{l}_1, \boldsymbol{l}_2$ be two skew lines in $\mathbb{P}^3$. For any point $\boldsymbol{x}$ not on the these lines, we indicate with $\lambda(\boldsymbol{x})$ the unique transversal to $\boldsymbol{l}_1, \boldsymbol{l}_2$ passing through $\boldsymbol{x}$. If $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ are two planes intersecting at a line $\boldsymbol{\delta}$ that intersects $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$, then the map $f : \boldsymbol{\pi} \dashrightarrow \boldsymbol{\pi}'$ defined, for points $\boldsymbol{y}$ not on $\boldsymbol{\delta}$, as*

$$f(\boldsymbol{y}) = \lambda(\boldsymbol{y}) \wedge \boldsymbol{\pi}', \tag{10}$$

can be extended to a homography between $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$.

This Lemma also implies that two retinal planes that intersect at a transversal to the slits can define the same rational camera (using appropriate coordinate systems). Note the similarity with the traditional pinhole case, where the choice of the retinal plane is completely irrelevant since the pinhole $\boldsymbol{c}$ induces a homography between any planes $\boldsymbol{\pi}, \boldsymbol{\pi}'$ not containing $\boldsymbol{c}$. See Figure 2.

3.1. A projective model using linear projections

Contrary to the case of pinhole cameras, two-slit cameras of the form (7) are *not* all projectively equivalent. This can be argued by noting that the coordinates $\boldsymbol{u}_1, \boldsymbol{u}_2$ in $\mathbb{P}^2$ of the points $\boldsymbol{y}_1 = \boldsymbol{l}_1 \wedge \boldsymbol{\pi}$, $\boldsymbol{y}_2 = \boldsymbol{l}_2 \wedge \boldsymbol{\pi}$ (the intersections of the slits with the retinal planes) are always preserved by projective transformations of $\mathbb{P}^3$. For Batog *et al.* [1, 2], the coordinates $\boldsymbol{u}_1, \boldsymbol{u}_2$ are "intrinsic parameters" of the camera; indeed, they are *projective* intrinsics (*i.e.*, invariants) of a two-slit device. Batog *et al.* also argue that choosing the points $\boldsymbol{y}_1$, $\boldsymbol{y}_2$ as points in the projective basis on $\boldsymbol{\pi}$ leads to simplified analytic expressions for the projection. Here, we develop this idea further, and observe that any two-slit camera with this kind of coordinate system can always be described by a pair of *linear* projections. More precisely, for any retinal plane $\boldsymbol{\pi}$, let us fix a coordinate system $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3]$ where $\boldsymbol{y}_1 = \boldsymbol{l}_2 \wedge \boldsymbol{\pi}$, $\boldsymbol{y}_2 = \boldsymbol{l}_1 \wedge \boldsymbol{\pi}$ and $\boldsymbol{y}_3$ is arbitrary: in this case, a straightforward computation shows that (7) reduces to

$$\boldsymbol{x} \mapsto \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} (\boldsymbol{p}_1^T\boldsymbol{x})\ (\boldsymbol{q}_2^T\boldsymbol{x}) \\ (\boldsymbol{p}_2^T\boldsymbol{x})\ (\boldsymbol{q}_1^T\boldsymbol{x}) \\ (\boldsymbol{p}_2^T\boldsymbol{x})\ (\boldsymbol{q}_2^T\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_1^T\boldsymbol{x}/\boldsymbol{p}_2^T\boldsymbol{x} \\ \boldsymbol{q}_1^T\boldsymbol{x}/\boldsymbol{q}_2^T\boldsymbol{x} \\ 1 \end{bmatrix}, \tag{11}$$

where $\boldsymbol{p}_1 = (\boldsymbol{l}_1 \vee \boldsymbol{y}_3)$, $\boldsymbol{p}_2 = -(\boldsymbol{l}_1 \vee \boldsymbol{y}_1)$, $\boldsymbol{q}_1 = (\boldsymbol{l}_2 \vee \boldsymbol{y}_3)$, $\boldsymbol{q}_2 = -(\boldsymbol{l}_2 \vee \boldsymbol{y}_2)$ are vectors representing planes in $\mathbb{P}^3$. It is easy to see that this quadratic map can be described using *two linear maps* $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1$, namely

$$\boldsymbol{x} \mapsto \begin{bmatrix} u_1 \\ u_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_1^T\boldsymbol{x} \\ \boldsymbol{p}_2^T\boldsymbol{x} \end{bmatrix} = \boldsymbol{A}_1\boldsymbol{x},\ \boldsymbol{x} \mapsto \begin{bmatrix} u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1^T\boldsymbol{x} \\ \boldsymbol{q}_2^T\boldsymbol{x} \end{bmatrix} = \boldsymbol{A}_2\boldsymbol{x}. \tag{12}$$

In other words, (11) determines the 2×4 matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ up to two scale factors, and vice-versa. Since applying a projective transformation to $\boldsymbol{x}$ in (11) corresponds to a matrix multiplication applied to both $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, we easily deduce that every pair of 2×4-matrices of full rank and with disjoint null-spaces corresponds to a two-slit camera, and that all these cameras are projectively equivalent. The two 2×4 matrices for a two-slit camera are analogues of the 3×4 matrix representing a pinhole camera: for example, the slits are associated with the null-spaces of these two matrices.[3] For two given projection matrices, the retinal plane may be any plane containing the line $\{\boldsymbol{p}_2^T\boldsymbol{x} = \boldsymbol{q}_2^T\boldsymbol{x} = 0\}$: this is the line through $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, and is the locus of points where the projection is not defined. This completely describes a primitive projective model with $7 + 7 = 14$ degrees of freedom. More precisely, there are 8 degrees of freedom corresponding to the choice of the slits, 2 for the intersection points of the retinal plane with the slits, and 4 for the choice of coordinates on the plane (since two basis points are constrained).

In the remainder of the paper, we will always assume that a two-slit photographic camera is of the form (11). This is equivalent to knowing the "projective intrinsic parameters" [2], namely the coordinates of $\boldsymbol{l}_1 \wedge \boldsymbol{\pi}$, $\boldsymbol{l}_2 \wedge \boldsymbol{\pi}$. We will also identify a camera with its two associated projection matrices.

[3] The linear maps $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1$ correspond in fact to the "line-centered" projections for the two slits. The action of a two-slit camera is arguably more natural viewed as a map $\mathbb{P}^3 \dashrightarrow \mathbb{P}^1 \times \mathbb{P}^1$, however we chose to maintain $\mathbb{P}^2$ as the image domain, since it is a better model for the retinal plane used in physical devices.

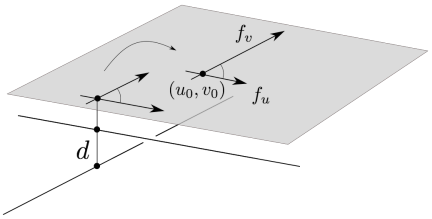

Figure 3. Physical interpretation of the entries the calibration matrices for parallel two-slit cameras: the parameters f_u, f_v, u_0, v_0 describe the change of retinal plane coordinates, with respect to some camera in the euclidean orbit of (15).

Example 3. The two-slit projection from Example 2 is of the form (11) with

$$\boldsymbol{A}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \boldsymbol{A}_2 = \begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}. \quad (13)$$

The retinal plane belongs to the pencil of planes containing $\{x_3 = x_3 + x_4 = 0\}$, *i.e.*, it is a plane of the form $x_3 - dx_4 = 0$. The choice $d = 1$ is natural since points of the form $[x_1, x_2, 1, 1]$ are mapped to $[x_1, x_2, 1]$. ♢

3.2. Orbits and calibration matrices

Using the linear model introduced above, we can easily describe affine, similarity, and euclidean orbits for two-slit cameras. For example, the affine orbit of the device in (9), (13) corresponds to

$$\boldsymbol{A}_1 = \begin{bmatrix} \boldsymbol{m}_1^T & t_1 \\ \boldsymbol{m}_3^T & t_3 \end{bmatrix}, \boldsymbol{A}_2 = \begin{bmatrix} \boldsymbol{m}_2^T & t_2 \\ \boldsymbol{m}_3^T & t_4 \end{bmatrix}, \quad (14)$$

where $\boldsymbol{m}_i$ are arbitrary 3-vectors. This is the family of two-slit cameras where the retinal plane is parallel to the slits: indeed, although this plane is not completely determined, it is constrained to contain the line $\{[\boldsymbol{m}_3^T, t_3]\boldsymbol{x} = [\boldsymbol{m}_3^T, t_4]\boldsymbol{x} = 0\}$, that intersects both slits. We will refer to (14) as a *parallel* two-slit camera. These cameras form an affine model with 12 degrees of freedom.

We now consider the family of (euclidean) parallel cameras of the form

$$\boldsymbol{A}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \boldsymbol{A}_2 = \begin{bmatrix} 2\cos\theta & 2\sin\theta & 0 & 0 \\ 0 & 0 & 1 & d \end{bmatrix}. \quad (15)$$

for $d \neq 0$ and $0 < \theta < 2\pi$ (and $\theta \neq \pi$). The slits for this camera are at an angle of θ and distance d. Note that (9) is of this form, with $\theta = \pi/2$ and $d = 1$.

Using (15) as a family of canonical euclidean devices, we can introduce expressions for the "intrinsic parameters" of two-slit cameras.

Proposition 1. *If $\boldsymbol{A}_1$, $\boldsymbol{A}_2$ describe a parallel two-slit camera (14), then we can uniquely write*

$$\boldsymbol{A}_1 = \boldsymbol{K}_1 \begin{bmatrix} \boldsymbol{r}_1^T & t_1 \\ \boldsymbol{r}_3^T & t_3 \end{bmatrix}, \boldsymbol{A}_2 = \boldsymbol{K}_2 \begin{bmatrix} \boldsymbol{r}_2^T & t_2 \\ \boldsymbol{r}_3^T & t_4 \end{bmatrix}, \quad (16)$$

where $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ are upper-triangular 2×2 matrices defined up to scale with positive elements along the diagonal, and $\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3$ are unit vectors, with $\boldsymbol{r}_3$ orthogonal to both $\boldsymbol{r}_1, \boldsymbol{r}_2$. Here, $\theta = \arccos(\boldsymbol{r}_1 \cdot \boldsymbol{r}_2)$ is the angle *between the slits, and $|t_4 - t_3|$ is the* distance *between the slits. Moreover, if the matrices $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ are written as*

$$\boldsymbol{K}_1 = \begin{bmatrix} f_u & u_0 \\ 0 & 1 \end{bmatrix}, \boldsymbol{K}_2 = \begin{bmatrix} 2f_v & v_0 \\ 0 & 1 \end{bmatrix}, \quad (17)$$

then f_u, f_v can be interpreted as "magnifications" in the u and v directions, and (u_0, v_0) as the position of the "principal point". See Figure 3.

The parameters θ and d, and the matrices $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ are clearly invariant under euclidean transformations. Moreover, within the parallel model (14), two cameras belong to the same euclidean orbit if and only if all of their parameters are the same. In fact, the 12 degrees of freedom of a parallel camera are split into 6 corresponding to the "intrinsics" θ, d, $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$, and 6 for the "extrinsic" action of euclidean motion.[4] Compared to the traditional intrinsic parameters for pinhole cameras, we note the absence of a term corresponding to the "skewness" of the image reference frame. Indeed, the angle between the two axes must be the same as the angle between the slits, as a consequence of the "intrinsic coordinate system" (the principal directions correspond in fact to the fixed basis points $\boldsymbol{y}_1, \boldsymbol{y}_2$ on the retinal plane). On the other hand, θ and d do not have an analogue for pinhole cameras. We will sometimes refer to d and θ as the "3D" intrinsic parameters, since we distinguish them from the "analytic" intrinsic parameters, that are entries of the calibration matrices $\boldsymbol{K}_1, \boldsymbol{K}_2$, and differentiate (euclidean orbits of) cameras only based on analytic part of their mapping. We also point out that for two-slit cameras in (16), the euclidean orbit and the similarity orbit do *not* coincide: this implies that when the intrinsic parameters are known, some information on the scale of a scene can be inferred from a photograph [21].

Pushbroom cameras. Pushbroom cameras are degenerate class of projective two-slit cameras, in which one of the slits lies on the plane at infinity [3]. This is quite similar to the class affine cameras for perspective projections. Pushbrooms are handled by our projective model (11), but not by our affine one (14), where both slits are necessarily finite. We thus introduce another affine model, namely

$$\boldsymbol{A}_1 = \begin{bmatrix} \boldsymbol{m}_1^T & t_1 \\ \boldsymbol{0} & 1 \end{bmatrix}, \boldsymbol{A}_2 = \begin{bmatrix} \boldsymbol{m}_2^T & t_2 \\ \boldsymbol{m}_3^T & t_3 \end{bmatrix}, \quad (18)$$

where $\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3$ are arbitrary 3-vectors. All such cameras are equivalent up to affine transformations, so this de-

[4] Intrinsic parameters describing euclidean orbits among more general (non-parallel) two-slit cameras can also be defined, but two more parameters are required. We chose to consider only two-slits with retinal plane parallel to the slits, since this is a natural assumption, and because the parameters have a simpler interpretation in this case.

scribes an affine model with 11 degrees of freedom. The corresponding rational cameras can be written as

$$\boldsymbol{x} \mapsto \left([\boldsymbol{m}_1^T, t_1]\boldsymbol{x}, \frac{[\boldsymbol{m}_2^T, t_2]\boldsymbol{x}}{[\boldsymbol{m}_3^T, t_3]\boldsymbol{x}}, 1 \right)^T . \tag{19}$$

This coincides with the linear pushbroom model proposed by Hartley and Gubta [6], who identify (19) with the 3×4 matrix with rows $[\boldsymbol{m}_1^T, t_1], [\boldsymbol{m}_2^T, t_2], [\boldsymbol{m}_3^T, t_3]$.

Let us now consider a family of affine pushbroom cameras of the form

$$\begin{bmatrix} \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \tag{20}$$

for $0 < \theta < 2\pi$ (and $\theta \neq \pi$). This represents a pushbroom camera where the direction of movement is at an angle θ with respect to the parallel scanning planes. We use this family of canonical devices to define calibration matrices and intrinsic parameters.

Proposition 2. *Let $\boldsymbol{A}_1, \boldsymbol{A}_2$ define a pushbroom camera as in (18), and let us also assume that $\boldsymbol{m}_1$ and $\boldsymbol{m}_3$ are orthogonal.*[5] *We can uniquely write*

$$\boldsymbol{A}_1 = \boldsymbol{K}_1 \begin{bmatrix} \boldsymbol{r}_1^T & t_1 \\ \boldsymbol{0} & 1 \end{bmatrix}, \boldsymbol{A}_2 = \boldsymbol{K}_2 \begin{bmatrix} \boldsymbol{r}_2^T & t_2 \\ \boldsymbol{r}_3^T & t_3 \end{bmatrix}. \tag{21}$$

where $\boldsymbol{K}_1 = \mathrm{diag}(1/v, 1)$, $\boldsymbol{K}_2 = \begin{bmatrix} f & u \\ 0 & 1 \end{bmatrix}$ (with positive v and f) and $\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3$ are unit vectors, with $\boldsymbol{r}_3$ orthogonal to both $\boldsymbol{r}_1, \boldsymbol{r}_2$. Here, $\theta = \arccos(\boldsymbol{r}_1 \cdot \boldsymbol{r}_2)$ is the angle *between the two slits (or between the direction of motion of the sensor and the parallel scanning planes). Moreover, v can be interpreted as the speed of the sensor, and f and u as the magnification and the principal point of the 1D projection.*

The entries $\boldsymbol{K}_1, \boldsymbol{K}_2$ are the "analytic" intrinsic parameters of the pushbroom camera, while θ is a "3D" intrinsic parameter.

4. Two-slit cameras: algorithms

In this section, we apply our study of two-slit cameras to develop algorithms for structure from motion (SfM). The epipolar geometry of two-slit cameras will be described in terms of a $2 \times 2 \times 2 \times 2$ *epipolar tensor*. Previously, image correspondences between two-slit cameras have been characterized using a 6×6 [3], or a 4×4 fundamental matrix [2]. The latter approach, due to Batog *et al.* [2], is similar to ours, since it is based on the "intrinsic" image reference frame that we also adopt. However, the tensor representation has the advantage of being easily described in terms of the elements of the four 2×4 projection matrices, in a form that closely resembles the corresponding expression for the traditional fundamental matrix. The definition of the epipolar tensor was already given in [18] for image coordinates in $\mathbb{P}^1 \times \mathbb{P}^1$ (and without explicit links to physical coordinates). Here, we also observe that every such tensor identifies exactly *two* projective camera configurations:

Theorem 1. *Let $(\boldsymbol{A}_1, \boldsymbol{A}_2)$, $(\boldsymbol{B}_1, \boldsymbol{B}_2)$ be two general projective two-slit cameras. The set of corresponding image points $\boldsymbol{u}$, $\boldsymbol{u}'$ in $\mathbb{P}^2$ is characterized by the following relation:*

$$\sum_{ijkl} f_{ijkl} \begin{bmatrix} u_1 \\ u_3 \end{bmatrix}_i \begin{bmatrix} u_2 \\ u_3 \end{bmatrix}_j \begin{bmatrix} u_1' \\ u_3' \end{bmatrix}_k \begin{bmatrix} u_2' \\ u_3' \end{bmatrix}_l = 0, \tag{22}$$

where $\boldsymbol{F} = (f_{ijkl})$ is a $2 \times 2 \times 2 \times 2$ "epipolar tensor". Its entries are

$$f_{ijkl} = (-1)^{i+j+k+l} \cdot \det \left[(\boldsymbol{A}_1)_{3-i}^T \ (\boldsymbol{A}_2)_{3-j}^T \ (\boldsymbol{B}_1)_{3-k}^T \ (\boldsymbol{B}_2)_{3-l}^T \right]. \tag{23}$$

Up to projective transformations of $\mathbb{P}^3$ there are two *configurations $(\boldsymbol{A}_1, \boldsymbol{A}_2), (\boldsymbol{B}_1, \boldsymbol{B}_2)$ compatible with a given epipolar tensor.*

Proof sketch. The definition of $\boldsymbol{F}$ follows by applying the incidence constraint for two lines to the "inverse line projections" (5) of image points. See the supplementary material or [18] for details. The definition of $\boldsymbol{F}$ is clearly invariant under projective transformations of $\mathbb{P}^3$. Hence, we may assume that

$$\boldsymbol{A}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ c_{11} & c_{12} & c_{13} & c_{14} \end{bmatrix}, \ \boldsymbol{A}_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ c_{21} & c_{22} & c_{23} & c_{24} \end{bmatrix}, \\ \boldsymbol{B}_1 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ c_{31} & c_{32} & c_{33} & c_{34} \end{bmatrix}. \ \boldsymbol{B}_2 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ c_{41} & c_{42} & c_{43} & c_{44} \end{bmatrix}. \tag{24}$$

The 16 entries of $\boldsymbol{F}$ are now the principal minors (*i.e.*, all minors obtained by considering subsets of rows and columns with the same indices) of the 4×4-matrix $\boldsymbol{C} = (c_{ij})$. Thus, determining the projection matrices $(\boldsymbol{A}_1, \boldsymbol{A}_2)$, $(\boldsymbol{B}_1, \boldsymbol{B}_2)$ corresponding to the tensor $\boldsymbol{F}$, is equivalent to finding the entries of a 4×4-matrix given its principal minors. This problem is studied in [10]. The set of all matrices with the same principal minors as $\boldsymbol{C}$ have the form $\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{D}$ or $\boldsymbol{D}^{-1}\boldsymbol{C}^T\boldsymbol{D}$, where $\boldsymbol{D}$ is a diagonal matrix. These two families of matrices, viewed as elements of (24), correspond to two distinct projective configurations of cameras. □

The set of all epipolar tensors forms a 13-dimensional variety in $\mathbb{P}^{15}$: this agrees with $14+14-15 = 13$, where 14 represents the degrees of freedom of two-slit cameras, and 15 is to account for projective ambiguity. Two equations are sufficient to characterize an epipolar tensor *locally*, however a result in [10] implies that a complete algebraic characterization actually requires 718 polynomials of degree 12.

Our study of canonical forms and calibration matrices in Section 3 also leads to a natural definition of *essential tensors*: for example, an essential tensor could be defined by (22) where $(\boldsymbol{A}_1, \boldsymbol{A}_2), (\boldsymbol{B}_1, \boldsymbol{B}_2)$ are all of the form (16)

[5] The more general case can also be described, but presents some technical difficulties. See the supplementary material for a discussion.

with $\boldsymbol{K}_1$, $\boldsymbol{K}_2$ being the identity. Proposition 1 then guarantees that for any pair of "parallel" two-slit cameras as in (14), we can uniquely write the epipolar tensor as

$$\boldsymbol{F}_{ijkl} = \boldsymbol{E}_{ijkl}(\boldsymbol{K}_1\boldsymbol{A})_i(\boldsymbol{K}_2\boldsymbol{A})_j(\boldsymbol{K}_1\boldsymbol{B})_k(\boldsymbol{K}_2\boldsymbol{B})_l \qquad (25)$$

where $\boldsymbol{E}_{ijkl}$ is an essential tensor. This closely resembles the analogous decomposition of fundamental matrices. Recovering an algebraic characterization of essential tensors, similar to the classical result that identifies essential matrices as fundamental matrices with two equal singular values, could be an interesting topic for future work.

Structure from motion. Using Theorem 1, we can design a *linear algorithm* for SfM, that proceeds as follows: (1) Using at least 15 image point correspondences, estimate $\boldsymbol{F}$ linearly using (22). (2) Recover two projective camera configurations that are compatible $\boldsymbol{F}$. Clearly, for noisy image correspondences, the linear estimate from step 1) will not be a valid epipolar tensor: a simple solution for this is to recover elements of $\boldsymbol{C}$ using only 13 principal minors given by the entries of $\boldsymbol{F}$. More precisely, after setting $c_{12} = c_{13} = c_{14} = 1$ (and normalizing $\boldsymbol{F}$ so that $f_{2222} = 1$), the elements on the diagonal and on the first column of $\boldsymbol{C}$ can be recovered from $\boldsymbol{F}$ using linear equalities. The remaining six entries are pairwise constrained by six elements of $\boldsymbol{F}$, leading to 8 possible matrices $\boldsymbol{C}$. In an ideal setting with no noise, exactly two of the 8 solutions will be consistent with the remaining two elements of $\boldsymbol{F}$ (more generally, we consider the two solutions that minimize an "algebraic residual"). A preliminary implementation of this approach, presented in detail the supplementary material, confirms that projective configurations of two-slit cameras can be recovered from image correspondences. It is also possible to design a *13-point algorithm* that recovers projection matrices (24) and the corresponding tensor $\boldsymbol{F}$ from a minimal amount of data, namely 13 point correspondences between images. The set of linear tensors that satisfy (22) for 13 correspondences is a two-dimensional linear space, and imposing constraints for being a valid epipolar tensor leads to a system of algebraic equations. According to [10, Remark 14] this system has 28 complex solutions for $\boldsymbol{F}$, which translate into 56 matrices $\boldsymbol{C} = (c_{ij})$. Experiments using the computer algebra system `Macaulay2` [4] confirm these theoretical results.

Self-calibration. Any reconstruction based on the epipolar tensor will be subject to projective ambiguity. On the other hand, using results from Section 3, it is possible to develop strategies for *self-calibration*. Let us assume that we have recovered a *projective* reconstruction of two-slit projections $\boldsymbol{A}_1^i, \boldsymbol{A}_2^i$ for $i = 1, \dots, n$, and also that we know that they are in fact (parallel) finite two-slit cameras. Our goal is to find a "euclidean upgrade", that is, a 4×4-matrix $\boldsymbol{Q}$ that describes the transition from a euclidean reference frame to the frame corresponding to our projective reconstruction. According to Proposition 1, we may write

$$\begin{aligned} \boldsymbol{A}_1^i \boldsymbol{Q} \boldsymbol{\Omega}^* \boldsymbol{Q}^T \boldsymbol{A}_1^{i\,T} &= \boldsymbol{K}_1^i \boldsymbol{K}_1^{i\,T} \\ \boldsymbol{A}_2^i \boldsymbol{Q} \boldsymbol{\Omega}^* \boldsymbol{Q}^T \boldsymbol{A}_2^{i\,T} &= \boldsymbol{K}_2^i \boldsymbol{K}_2^{i\,T}, \end{aligned} \qquad (26)$$

(equality up to scale) where $\boldsymbol{K}_1^i, \boldsymbol{K}_2^i$ are the unknown 2×2 matrices of intrinsic parameters for $\boldsymbol{A}_1^i, \boldsymbol{A}_2^i$, and $\boldsymbol{\Omega}^* = \mathrm{diag}(1, 1, 1, 0)$. Geometrically, (26) expresses the fact that the *dual of the image of the absolute conic* is a section of the *dual absolute quadric*. These relations are completely analogous to the self-calibration equations for pinhole cameras, so that any partial knowledge of the matrices $\boldsymbol{K}_1^i, \boldsymbol{K}_2^i$ can be used to impose constraints on $\boldsymbol{Q}\boldsymbol{\Omega}^*\boldsymbol{Q}^T$ and hence on $\boldsymbol{Q}$ (although, as for the pinhole case, we can actually only recover a "similarity" upgrade). For example, if the principal points are known to be at the origin (so $\boldsymbol{K}_1^i \boldsymbol{K}_i^{i\,T}$ and $\boldsymbol{K}_2^i \boldsymbol{K}_2^{i\,T}$ are diagonal), then (26) gives four linear equations in the elements of $\boldsymbol{Q}\boldsymbol{\Omega}^*\boldsymbol{Q}^T$ corresponding to the zero entries of $\boldsymbol{K}_1^i \boldsymbol{K}_1^{i\,T}$ and $\boldsymbol{K}_2^i \boldsymbol{K}_2^{i\,T}$. A sufficient number of views allows us to estimate $\boldsymbol{Q}\boldsymbol{\Omega}^*\boldsymbol{Q}^T$, and from a singular value decomposition we can recover $\boldsymbol{Q}$ up to a similarity transformation. We refer to the supplementary material for some experiments with synthetic data.

5. Discussion

In the first part of this presentation, we have described optical systems that can be associated with congruences of order one, and that record lines by measuring the coordinates of their intersection with some retinal plane. This setting is very general, but excludes important families of imaging devices such as (non-central) *catadioptric cameras*, or cameras with *optical distortions*. In these examples, visual rays are *reflected* or *refracted* by specular surfaces or optical lenses, leading to maps that are often not rational (for example, they may involve square-roots). These cases could be handled by noting that mirrors or lenses act on a line congruence L of (primary) visual rays by mapping it to a new congruence L' of (secondary) rays. A completely general system consists of a sequence of such steps, followed a final map where rays are intersected with a retinal plane. Partial results in [18] discuss the effect of reflecting a $(1, \beta)$-congruence off an algebraic surface, but an effective description of reflections and refractions in terms of line congruences is still missing. It will of course be of great interest to pursue this direction of research, and extend the approach proposed in this presentation to a completely general setting.

Acknowledgments. This work was supported in part by the ERC grant VideoWorld, the Institut Universitaire de France, an Inria International Chair, the Inria-CMU associated team GAYA, ONR MURI N000141010934, the US National Science Foundation (DMS-1419018) and the Einstein Foundation Berlin.

References

[1] G. Batog. *Classical problems in computer vision and computational geometry revisited with line geometry*. PhD thesis, Université Nancy II, 2011. 1, 3, 5

[2] G. Batog, X. Goaoc, and J. Ponce. Admissible linear map models of linear cameras. In *CVPR*, 2010. 1, 2, 3, 4, 5, 7

[3] D. Feldman, T. Pajdla, and D. Weinshall. On the epipolar geometry of the crossed-slits projection. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 988–995. IEEE, 2003. 4, 5, 6, 7

[4] D. R. Grayson and M. E. Stillman. Macaulay2, a software system for research in algebraic geometry. Available at http://www.math.uiuc.edu/Macaulay2/. 8

[5] M. Grossberg and S. Nayar. The raxel imaging model and ray-based calibration. *IJCV*, 61(2):119–137, 2005. 1

[6] R. Gupta and R. I. Hartley. Linear pushbroom cameras. *IEEE Transactions on pattern analysis and machine intelligence*, 19(9):963–975, 1997. 1, 2, 4, 7

[7] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 4

[8] Y. Hu, V. Tao, and A. Croitoru. Understanding the rational function model: methods and applications. *International Archives of Photogrammetry and Remote Sensing*, 20(6), 2004. 3

[9] E. Kummer. Über die algebraischen Strahlensysteme, insbesondere über die der ersten und zweiten Ordnung. *Abh. K. Preuss. Akad. Wiss. Berlin*, pages 1–120, 1866. 1, 2, 3

[10] S. Lin and B. Sturmfels. Polynomial relations among principal minors of a 4× 4-matrix. *Journal of Algebra*, 322(11):4121–4131, 2009. 7, 8

[11] S. Maybank and O. Faugeras. A theory of self-calibration of a moving camera. *IJCV*, 8(2):123–151, 1992. 4

[12] T. Pajdla. Geometry of two-slit camera. Technical Report 2002-2, Chezch Technical Ubiversity, 2002. 1, 2, 4

[13] T. Pajdla. Stereo with oblique cameras. *IJCV*, 47(1):161–170, 2002. 1, 2

[14] P. D. Poi. Congruences of lines with one-dimensional focal locus. *Portugaliae Mathematica*, 61:329–338, 2004. 1, 2, 3

[15] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch. Visual modeling with a hand-held camera. *IJCV*, 59:207–232, 2004. 4

[16] J. Ponce. What is a camera? In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1526–1533. IEEE, 2009. 1, 2, 3

[17] J. Ponce, T. Papadopoulo, M. Teillaud, and B. Triggs. The absolute quadratic complex and its application to camera self calibration. In *CVPR*, 2005. 4

[18] J. Ponce, B. Sturmfels, and M. Trager. Congruences and concurrent lines in multi-view geometry. *arXiv preprint arXiv:1608.05924*, 2016. 1, 2, 3, 7, 8

[19] P. Rademacher and G. Bishop. Multiple-center-of-projection images. In *SIGGRAPH*, pages 199–206, 1998. 1

[20] S. Seitz and J. Kim. The space of all stereo images. *ijcv*, 48(1):21–28, 2002. 2

[21] P. Sturm, S. Ramalingam, J. Tardif, S. Gasparini, and J. Barreto. Camera models and fundamental concepts used in geometric computer vision. *Foundations and Trends in Computer Graphics and Vision*, 6(1-2):1–183, 2011. 1, 2, 4, 6

[22] W. Triggs. Auto-calibration and the absolute quadric. In *CVPR*, pages 609–614, San Juan, Puerto Rico, June 1997. 4

[23] D. Weinshall, M.-S. Lee, T. Brodsky, M. Trajkovic, and D. Feldman. New view generation with a bi-centric camera. In *European Conference on Computer Vision*, pages 614–628. Springer, 2002. 2, 4

[24] J. Ye and J. Yu. Ray geometry in non-pinhole cameras: A survey. *The Visual Computer*, 30(1):93–112, 2014. 2

[25] J. Yu and L. McMillan. General linear cameras. In *Proc. ECCV*, 2004. 1, 2

[26] A. Zomet, D. Feldman, S. Peleg, and D. Weinshall. Mosaicing new views: the crossed-slits projection. *PAMI*, 25(6):741–754, 2003. 1, 2, 4, 5