Open access • Proceedings Article • DOI:10.1109/CVPR.2005.237

Multi-view geometry for general camera models — Source link

Peter Sturm

Institutions: French Institute for Research in Computer Science and Automation

Published on: 20 Jun 2005 - Computer Vision and Pattern Recognition

Topics: Pinhole camera model, Triangulation (computer vision), Projection (set theory), Catadioptric system and Perspective (graphical)

Multi-View Geometry for General Camera Models

Peter Sturm

Multi-View Geometry for General Camera Models

P. Sturm
INRIA Rhône-Alpes, 38330 Montbonnot, France

Abstract

We consider the structure from motion problem for a previously introduced, highly general imaging model, where cameras are modeled as possibly unconstrained sets of projection rays. This allows to describe most existing camera types, including pinhole cameras, sensors with radial or more general distortions, catadioptric cameras (central or non-central), etc. We introduce a hierarchy of general camera models: the most general model has unconstrained projection rays whereas the most constrained model dealt with here is the central model, where all rays pass through a single point. Intermediate models are what we call axial cameras (all rays touch a single line), and x-slit cameras (rays touch two lines). The foundations for a multi-view geometry of completely non-central cameras are given, leading to the formulation of multi-view matching tensors, analogous to the fundamental/essential matrices, trifocal and quadrifocal tensors of perspective cameras. This framework is then specialized explicitly for the two-view case, for the intermediate camera types mentioned above.

1. Introduction

Many different types of cameras including pinhole, stereo, catadioptric, omnidirectional and non-central cameras have been used in computer vision. Most existing camera models are parametric (i.e. defined by a few intrinsic parameters) and address imaging systems with a single effective viewpoint (all rays pass through one point). In addition, existing calibration or structure from motion procedures are often taylor-made for specific camera models, see examples e.g. in [3, 11, 7].

The aim of this work is to relax these constraints: we want to propose and develop calibration and structure from motion methods that should work for any type of camera model, and especially also for cameras without a single effective viewpoint. To do so, we first renounce on parametric models, and adopt the following very general model: a camera acquires images consisting of pixels; each pixel captures light that travels along a ray in 3D. The camera is fully described by [9]:

- the coordinates of these rays (given in some local coordinate frame).
- the mapping between rays and pixels; this is basically a simple indexing.

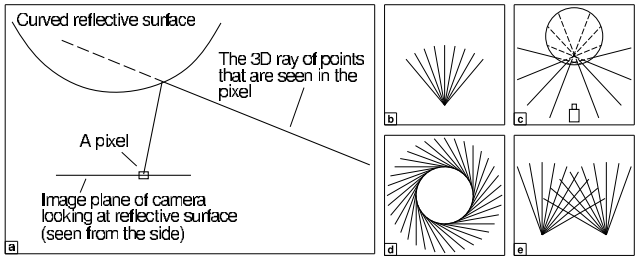

Figure 1. Examples of imaging systems; (c)–(e) are non-central devices. (a) Catadioptric system. (b) Central camera (e.g. perspective, with or without radial distortion). (c) Camera looking at reflective sphere. (d) Omnivergent imaging system [18, 21]. (e) Stereo system.

This general imaging model allows to describe virtually any camera that captures light rays travelling along straight lines. Examples are (cf. figure 1):

- a camera with any type of optical distortion, such as radial or decentering.
- a camera looking at a reflective surface, e.g. as often used in surveillance, a camera looking at a spherical or otherwise curved mirror [12]. Such systems, as opposed to central catadioptric systems [1, 6] using e.g. parabolic mirrors, do not in general have a single effective viewpoint.
- multi-camera stereo systems: put together the pixels of all image planes; they "catch" light rays that definitely do not travel along lines that all pass through a single point. Nevertheless, in the above general camera model, a stereo system (with rigidly linked cameras) is considered as a **single** camera.
- other acquisition systems, many of them non-central, see e.g. [2, 14, 17, 18, 21, 25, 26], insect eyes, etc.

In this paper, we propose the foundations for a multview geometry of the general, non-central camera model, leading to the formulation of multi-view matching tensors, analogous to the fundamental or essential matrices, trifocal and quadrifocal tensors of perspective cameras. The multiview geometry will be formulated for calibrated cameras, i.e. we do not directly work with image point correspondences, but rather with correspondences between associated camera rays in 3D.

We also introduce a natural hierarchy of camera models: the most general model has unconstrained projection rays

whereas the most constrained model dealt with here is the central model, where all rays pass through a single point. Intermediate models considered in this paper are axial and x-slit cameras. The two-view geometry, first established for non-central cameras, is specialized for these intermediate camera types in this paper. Several works exist on epipolar geometry for omnidirectional cameras, central and non-central ones [5, 8, 15, 19, 22, 24]. Most of them aimed at obtaining matching constraints between uncalibrated images, whereas in this paper, we deal with calibrated cameras and give a rather complete treatment of the problem.

The paper is organized as follows. §2 gives some background on Plücker coordinates for 3D lines, used to parameterize camera rays. A hierarchy of camera models is proposed in §3. §4 gives parameterizations of projection rays, for the different camera models. The multi-view geometry for the general camera model, as well as two-view geometry for intermediate models, is given in §5.

2. Plücker Coordinates

We represent projection rays as 3D lines, via Plücker coordinates. Several definitions exist for them; we use the following. Let $\mathbf{A}$ and $\mathbf{B}$ be the homogeneous coordinates of 3D points defining a line. The line can be represented by the skew-symmetric 4×4 Plücker matrix $\mathsf{L} = \mathbf{A}\mathbf{B}^\mathsf{T} - \mathbf{B}\mathbf{A}^\mathsf{T}$. It is independent (up to scale) of the points used to represent the line. An alternative representation for the line is its Plücker coordinate vector of length 6:

$$\mathbf{L} = \begin{pmatrix} A_4B_1 - A_1B_4 \\ A_4B_2 - A_2B_4 \\ A_4B_3 - A_3B_4 \\ A_3B_2 - A_2B_3 \\ A_1B_3 - A_3B_1 \\ A_2B_1 - A_1B_2 \end{pmatrix} \tag{1}$$

We sometimes split it in two 3-vectors $\mathbf{a}$ and $\mathbf{b}$,

$$\mathbf{a}^\mathsf{T} = \begin{pmatrix} L_1 & L_2 & L_3 \end{pmatrix} \quad \mathbf{b}^\mathsf{T} = \begin{pmatrix} L_4 & L_5 & L_6 \end{pmatrix}$$

which satisfy the so-called Plücker constraint: $\mathbf{a}^\mathsf{T}\mathbf{b} = 0$.

Consider a metric transformation defined by a rotation matrix R and a translation vector $\mathbf{t}$, acting on points via:

$$\mathbf{C} \rightarrow \begin{pmatrix} \mathsf{R} & \mathbf{t} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \mathbf{C}$$

Plücker coordinates are then transformed according to

$$\begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} \rightarrow \begin{pmatrix} \mathsf{R} & 0 \\ -[\mathbf{t}]_\times \mathsf{R} & \mathsf{R} \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}$$

Two lines intersect if the following relation holds:

$$\mathbf{L}_2^\mathsf{T} \begin{pmatrix} 0 & \mathsf{I} \\ \mathsf{I} & 0 \end{pmatrix} \mathbf{L}_1 = \mathbf{a}_2^\mathsf{T}\mathbf{b}_1 + \mathbf{b}_2^\mathsf{T}\mathbf{a}_1 = 0 \tag{2}$$

Table 1. Camera models, defined by 3D points and lines that have an intersection with all projection rays of a camera.

Points/lines cutting rays	Description
None	Non-central camera
1 point	Central camera
2 points	Camera with a single ray
1 line	Axial camera
1 point, 1 line	Central 1D camera
2 skew lines	X-slit camera
2 coplanar lines	Union of a non-central 1D camera and a central camera
3 coplanar lines without a common point	Non-central 1D camera

3. A Hierarchy of Camera Models

A **non-central camera** may have completely unconstrained projection rays, whereas for a **central camera**, there exists a point – the **optical center** – that lies on all projection rays. An intermediate case is what we call **axial cameras**, where there exists a line that cuts all projection rays – the **camera axis** (not to be confounded with optical axis). Examples of cameras falling into this class are:

- x-slit cameras [16, 27] (also called two-slit or crossed-slits cameras), and their special case of linear pushbroom cameras [10]. Note that these form a sub-class of axial cameras, as explained below.
- stereo systems consisting of 2 central cameras or 3 or more central cameras with collinear optical centers.
- non-central catadioptric cameras of the following type: the mirror is any surface of revolution and the optical center of the central camera looking at it (can be any central camera, not only pinhole), lies on its axis of revolution. It is easy to verify that in this case, all projection rays cut the mirror's axis of revolution, i.e. the camera is an axial camera, with the mirror's axis of revolution as camera axis. Note that catadioptric cameras with a spherical mirror and a central camera looking at it, are always non-central axial cameras.

These three classes of camera models may also be defined as: existence of a linear space of d dimensions that has an intersection with all projection rays: $d = 0$ defines central, $d = 1$ axial and $d = 2$ general non-central cameras.

Intermediate classes do exist. X-slit cameras are a special case of axial cameras: there actually exist 2 lines in space that both cut all projection rays. Similarly, central 1D cameras (cameras with a single row of pixels) can be defined by a point and a line in 3D. Camera models, some of which without much practical importance, are summarized in table 1. A similar way of defining camera types was suggested in [16].

Table 2. Parameterization of projection rays for different camera models (see text).

Camera model	Central		Axial		X-slit	
	finite	infinite	finite	infinite	finite+finite	finite+infinite
Parameterization of projection rays	$\begin{pmatrix}\mathbf{a}\\ \mathbf{0}\end{pmatrix}$	$\begin{pmatrix}0\\0\\a_3\\b_1\\b_2\\0\end{pmatrix}$	$\begin{pmatrix}\mathbf{a}\\b_1\\b_2\\0\end{pmatrix}$	$\begin{pmatrix}0\\a_2\\a_3\\\mathbf{b}\end{pmatrix}$	$\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\W&0&-Y&0\\0&0&0&1\\0&0&0&0\end{pmatrix}\begin{pmatrix}a_1\\a_2\\a_3\\b_2\end{pmatrix}$	$\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&W&0&0\\0&0&1&0\\0&0&0&1\\0&0&0&0\end{pmatrix}\begin{pmatrix}a_1\\a_3\\b_1\\b_2\end{pmatrix}$

It is worthwhile to consider different classes due to the following observation: the usual calibration and motion estimation algorithms proceed by first estimating a matrix or tensor by solving linear equation systems (e.g. the calibration tensors in [23] or the essential matrix [19]). Then, the parameters that are searched for (usually, motion parameters), are extracted from these. However, when estimating for example the 6×6 essential matrix of *non-central* cameras based on image correspondences obtained from *central* or *axial* cameras, then the associated linear equation system does not give a unique solution (much like when estimating a fundamental matrix from correspondences coming from coplanar 3D points). Consequently, the algorithms for extracting the actual motion parameters, can not be applied without modification.

In the following, we deal with central, axial, x-slit and fully non-central cameras.

4. Parameterizations

Multi-view geometry will be formulated in terms of the Plücker coordinates of camera rays. For other models than the fully non-central one, camera rays belong to constrained sets, as explained in the previous section. We may thus choose the cameras' local coordinate systems such as to obtain "simpler" coordinate vectors for camera rays, and in turn simpler matching constraints. Since we deal with calibrated cameras, rays are given in metric coordinate systems, and we may apply rotations and translations to fix local coordinate systems. Appropriate parameterizations for different models are explained in the following.

4.1. Central Cameras

All rays go through a single point, the optical center. We distinguish the cases of a finite and infinite optical center.

Finite optical center. We choose a local coordinate system with the optical center as origin. This leads to projection rays whose Plücker sub-vector $\mathbf{b}$ is zero, cf. table 2. This is one reason why the multi-focal tensors, e.g. the fundamental matrix, can be written with a "base size" of 3.

Infinite optical center (e.g. affine camera). We can not adopt the optical center as origin, thus choose a coordinate system where it has coordinates $(0,0,1,0)^\mathsf{T}$. Projection rays are then of the form given in the 3rd column of table 2.

4.2. Axial Cameras

All rays touch a line, the camera axis. Again, by choosing local coordinate systems appropriately, the formulation of the multi-view relations may be simplified. We distinguish the cases of a finite and an infinite camera axis.

Finite axis. Assume that the camera axis is the Z-axis. Then, all projection rays have Plücker coordinates with $L_6 = b_3 = 0$, cf. the 4th column of table 2.

Infinite axis. We choose a local coordinate system where the axis is the line at infinity with coordinates $(1,0,0)^\mathsf{T}$ (line coordinates on plane at infinity). The camera axis' Plücker coordinates are then given by $(0,0,0,1,0,0)^\mathsf{T}$. Projection rays thus have coefficients with $L_1 = a_1 = 0$ (this is obtained using (2)), cf. the 5th column of table 2.

Multi-view relations for axial cameras, with finite or infinite axis, can thus be formulated via tensors of "base size" 5, e.g. the essential matrix will be of size 5×5 (see §5.3.2).

4.3. X-Slit Cameras

As mentioned above, x-slit cameras are defined as follows: there exist two lines – *camera axes* – that cut all projection rays. The case of the two axes cutting one another, i.e. being coplanar, is not of interest here, so we consider two mutually skew axes. Two cases are thus possible: (i) both axes are finite lines or (ii) one of the two axes is a line at infinity. In any case, one axis at least is a finite line; we adopt a local coordinate system as said above for axial cameras (with finite axis). As for the second axis, we have to distinguish the two cases.

Two finite axes. Having fixed the first axis, we still have the freedom to rotate about it and translate along it. Since the two axes are skew, we may thus obtain a local coordinate system, where the second axis goes through a point on the Y-axis, and is parallel to the XZ-plane. Hence, it will be defined by two points as follows:

$$\mathbf{A}^\mathsf{T} = \begin{pmatrix}0 & Y & 0 & 1\end{pmatrix} \qquad \mathbf{B}^\mathsf{T} = \begin{pmatrix}X & 0 & Z & 0\end{pmatrix}$$

The second axis' Plücker coordinates are thus given by:

$$\mathbf{L}_2^\mathsf{T} = \begin{pmatrix}X & 0 & Z & -YZ & 0 & YZ\end{pmatrix}$$

Projection rays cut the two axes, so must be of the form:

$$\mathbf{P}^\mathsf{T} = \begin{pmatrix}a_1 & a_2 & a_3 & \left(\frac{YZ}{X}a_1 - Ya_3\right) & b_2 & 0\end{pmatrix}$$

We divide by X, but this is no problem since it can not be zero, otherwise the second axis would be parallel to the first one, and thus coplanar, which is excluded here. Let us replace $\frac{YZ}{X}$ by W. Then, each projection ray can be parameterized by 4 coefficients (which are defined up to scale), as given in the 6th column of table 2.

One finite and one infinite axis. Having fixed the first axis, we still have the freedom to rotate about it and translate along it. Translation has no effect on the infinite second axis, but we may rotate about the first axis, such that the second one has coordinates $(0, \cos\Theta, \sin\Theta)^\mathsf{T}$ (homogeneous coordinates of a line at infinity). The second axis' Plücker coordinates are thus:

$$\mathbf{L}_2^\mathsf{T} = \begin{pmatrix} 0 & 0 & 0 & 0 & \cos\Theta & \sin\Theta \end{pmatrix}$$

Projection rays cut the two axes, so must be of the form:

$$\mathbf{P}^\mathsf{T} = \begin{pmatrix} a_1 & -a_3\tan\Theta & a_3 & b_1 & b_2 & 0 \end{pmatrix}$$

For ease of notation, let us define $W = -\tan\Theta$. Then, each projection ray can be parameterized by 4 coefficients (defined up to scale), as given in the last column of table 2.

4.4. General Non-Central Cameras

No such simplification occurs, and multi-view tensors will have "base size" 6.

5. Multi-View Geometry

We establish the foundations of a multi-view geometry for general (non-central) cameras. Its cornerstones are, as with perspective cameras, matching tensors. We show how to establish them, analogously to the perspective case.

Here, we only talk about the calibrated case; the uncalibrated case is nicely treated for perspective cameras, since calibrated and uncalibrated cameras are linked by projective transformations. For non-central cameras however, there is no such link: in the most general case, every pair (pixel, camera ray) may be completely independent of other pairs.

5.1. Reminder on Perspective Multi-View Geometry

We briefly review how to derive multi-view matching relations for perspective cameras [4]. Let P_i be projection matrices of n images. Image points $\mathbf{q}_i$ are matching, if there exist a 3D point $\mathbf{Q}$ and scale factors λ_i with:

$$\lambda_i \mathbf{q}_i = \mathsf{P}_i \mathbf{Q}, \;\; \forall i = 1 \cdots n$$

This may be formulated as the following matrix equation:

$$\underbrace{\begin{pmatrix} \mathsf{P}_1 & \mathbf{q}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathsf{P}_2 & \mathbf{0} & \mathbf{q}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathsf{P}_n & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{q}_n \end{pmatrix}}_{\mathsf{M}} \begin{pmatrix} \mathbf{Q} \\ -\lambda_1 \\ -\lambda_2 \\ \vdots \\ -\lambda_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

The matrix M, of size $3n \times (4+n)$ has thus a null-vector, meaning that its rank is less than $4+n$. Hence, the determinants of all submatrices of size $(4+n) \times (4+n)$ must vanish. These determinants are multi-linear expressions in terms of the coordinates of image points $\mathbf{q}_i$. Every possible submatrix should be considered, but only those with 2 or more rows per view, give rise to constraints linking all projection matrices. Hence, constraints can be obtained up to n views with $2n \leq 4+n$, meaning that only for up to 4 views, matching constraints linking all views can be obtained.

The constraints for n views take the form:

$$\sum_{i_1=1}^{3}\sum_{i_2=1}^{3}\cdots\sum_{i_n=1}^{3} q_{1,i_1} q_{2,i_2} \cdots q_{n,i_n} T_{i_1,i_2,\cdots,i_n} = 0 \quad (3)$$

where the multi-view matching tensor T of dimension $3 \times \cdots \times 3$ depends on and partially encodes the cameras' projection matrices P_i.

Note that as soon as cameras are calibrated, this theory applies to any central camera: for a camera with radial distortion for example, the above formulation holds for distortion-corrected image points.

5.2. Multi-View Geometry of Non-Central Cameras

Here, instead of projection matrices (depending on calibration and pose), we deal with pose matrices:

$$\mathsf{P}_i = \begin{pmatrix} \mathsf{R}_i & \mathbf{t}_i \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \quad (4)$$

These are the similarity transformations that map a point from some global reference frame, into the camera's local coordinate frames (note that since no optical center and no camera axis exist, no assumptions about the local coordinate frames are made). As for image points, they are now replaced by camera rays. We will obtain expressions in terms of the rays' Plücker coordinates, i.e. we will end up with matching tensors T and matching constraints of the form (3), with the difference that tensors will have size $6 \times \cdots \times 6$ and act on Plücker line coordinates:

$$\sum_{i_1=1}^{6}\sum_{i_2=1}^{6}\cdots\sum_{i_n=1}^{6} L_{1,i_1} L_{2,i_2} \cdots L_{n,i_n} T_{i_1,i_2,\cdots,i_n} = 0 \quad (5)$$

In the following, we explain how to derive such matching constraints. Consider a set of n camera rays and let them be defined by two points $\mathbf{A}_i$ and $\mathbf{B}_i$ each; the choice of points to represent a ray is not important, since later we will fall back onto the ray's Plücker coordinates.

Now, a set of n camera rays are matching, if there exist a 3D point $\mathbf{Q}$ and scale factors λ_i and μ_i with:

$$\lambda_i \mathbf{A}_i + \mu_i \mathbf{B}_i = \mathsf{P}_i \mathbf{Q}, \;\; \forall i = 1 \cdots n$$

i.e. if the point $\mathsf{P}_i\mathbf{Q}$ lies on the line spanned by $\mathbf{A}_i$ and $\mathbf{B}_i$.

Like for perspective cameras, we group these equations in matrix form:

$$\underbrace{\begin{pmatrix} \mathsf{P}_1 & \mathbf{A}_1 & \mathbf{B}_1 & \cdots & \mathbf{0} & \mathbf{0} \\ \mathsf{P}_2 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathsf{P}_n & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_n & \mathbf{B}_n \end{pmatrix}}_{\mathsf{M}} \begin{pmatrix} \mathbf{Q} \\ -\lambda_1 \\ -\mu_1 \\ \vdots \\ -\lambda_n \\ -\mu_n \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}$$

As above, this equation shows that M must be rank-deficient. However, the situation is different here since the P_i are of size 4×4 now, and M of size $4n\times(4+2n)$. We thus consider submatrices of M of size $(4+2n)\times(4+2n)$. Furthermore, in the following we show that only submatrices with 3 rows or more per view, give rise to constraints on all pose matrices. Hence, $3n \leq 4+2n$, and again, $n \leq 4$, i.e. multi-view constraints are only obtained for up to 4 views.

Let us first see what happens for a submatrix of M where some view contributes a single row. The two columns corresponding to its base points $\mathbf{A}$ and $\mathbf{B}$, are multiples of one another: they contain only zeroes, besides a single non-zero coefficient, in the single row associated with the considered view. Hence, the determinant of the considered submatrix of M is always zero, and no constraint is available.

In the following, we exclude this case, i.e. we only consider submatrices of M where each view contributes at least two rows. Let N be such a matrix. Without loss of generality, we start to develop its determinant with the columns containing $\mathbf{A}_1$ and $\mathbf{B}_1$. The determinant is then given as a sum of terms of the following form:

$$(A_{1,j}B_{1,k} - A_{1,k}B_{1,j}) \det \bar{\mathsf{N}}_{jk}$$

where $j,k \in \{1..4\}$, $j \neq k$, and $\bar{\mathsf{N}}_{jk}$ is obtained from N by dropping the columns containing $\mathbf{A}_1$ and $\mathbf{B}_1$ as well as the rows containing $A_{1,j}$ and $A_{1,k}$. We observe several things:

- The term $(A_{1,j}B_{1,k} - A_{1,k}B_{1,j})$ is nothing else than a Plücker coordinate of the ray of camera 1 (cf. §2). By continuing with the development of the determinant of $\bar{\mathsf{N}}_{jk}$, it becomes clear that the total determinant of N can be written in the form:

$$\sum_{i_1=1}^{6}\sum_{i_2=1}^{6}\cdots\sum_{i_n=1}^{6} L_{1,i_1}L_{2,i_2}\cdots L_{n,i_n}T_{i_1,i_2,\cdots,i_n} = 0$$

  i.e. the coefficients of the $\mathbf{A}_i$ and $\mathbf{B}_i$ are "folded together" into Plücker coordinates of camera rays and T is a matching tensor relating the n cameras. Its coefficients depend exactly on the cameras' pose matrices.

- If camera 1 contributes only two rows to N, then the determinant of N becomes of the form:

$$L_{1,x}\left(\sum_{i_2=1}^{6}\cdots\sum_{i_n=1}^{6} L_{2,i_2}\cdots L_{n,i_n}T_{i_2,\cdots,i_n}\right) = 0$$

  i.e. it only contains a single coordinate $L_{1,x}$ of the ray of camera 1, and the tensor T does not depend at all on the pose of that camera. Hence, to obtain constraints relating *all* cameras, each camera has to contribute at least three rows to the considered submatrix of M.

Table 3. Cases of multi-view matching constraints for central and non-central cameras. Columns named "useful" contain entries of the form x-y-z etc. that correspond to sub-matrices of M that give rise to matching constraints linking *all* views: x-y-z refers to sub-matrices containing x rows from one camera, y from another etc.

	central		non-central	
# views	M	useful	M	useful
2	6×6	3-3	8×8	4-4
3	9×7	3-2-2	12×10	4-3-3
4	12×8	2-2-2-2	16×12	3-3-3-3

We are now ready to establish the different cases that lead to useful multi-view constraints. As mentioned above, for more than 4 cameras, no constraints linking all of them are available: submatrices of size at least $3n\times3n$ would be needed, but M only has $4+2n$ columns. So, only for $n \leq 4$, such constraints exist.

Table 3 gives all useful cases, both for central and non-central cameras. These lead to two-view, three-view and four-view matching constraints, encoded by essential matrices, trifocal and quadrifocal tensors. Deriving their forms is now mainly a mechanical task.

5.3. The Case of Two Views

We have so far explained how to formulate bifocal, trifocal and quadrifocal matching constraints between non-central cameras, expressed via matching tensors of dimension 6×6 to $6\times6\times6\times6$. To make things more concrete, we explore the two-view case in some more detail in the following. We show how the bifocal matching tensor, or essential matrix, can be expressed in terms of the motion/pose parameters. This is then specialized from non-central to axial, x-slit and central cameras. The essential matrices for these cases are summarized in table 4. That table also gives the minimum numbers of correspondences required for estimating the essential matrices using linear equations. These are not explained in detail due to lack of space, but can be derived easily by considering coefficients in essential matrices, that are zero are appear twice.

5.3.1. Non-Central Cameras

For simplicity, we assume here that the global coordinate system coincides with the first camera's local coordinate system, i.e. the first camera's pose matrix is the identity. As for the pose of the second camera, we drop indices, i.e. we express it via a pose matrix P, composed of a rotation

Table 4. Essential matrices for different camera models. The last column gives the minimum number of correspondences between projection rays required for computing essential matrices using linear equations.

Camera model	Essential matrix	Size	# corr.
Non-central	$\mathsf{E}_n = \begin{pmatrix} -[\mathbf{t}]_\times \mathsf{R} & \mathsf{R} \\ \mathsf{R} & \mathsf{0}_{3\times 3} \end{pmatrix}$	6×6	17
Axial with finite axis	$\mathsf{E}_{af} = \begin{pmatrix} -[\mathbf{t}]_\times \mathsf{R} & \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \\ R_{31} & R_{32} \end{pmatrix} \\ \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \end{pmatrix} & \mathsf{0}_{2\times 2} \end{pmatrix}$	5×5	16
Axial with infinite axis	$\mathsf{E}_{ai} = \begin{pmatrix} t_1 R_{32} - t_3 R_{12} & t_1 R_{33} - t_3 R_{13} & R_{21} & R_{22} & R_{23} \\ t_2 R_{12} - t_1 R_{22} & t_2 R_{13} - t_1 R_{23} & R_{31} & R_{32} & R_{33} \\ R_{12} & R_{13} & 0 & 0 & 0 \\ R_{22} & R_{23} & 0 & 0 & 0 \\ R_{32} & R_{33} & 0 & 0 & 0 \end{pmatrix}$	5×5	11
X-slit with two finite axes	$\mathsf{E}_{xff} = \begin{pmatrix} 1 & 0 & 0 & W_2 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -Y_2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \mathsf{E}_{af} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ W_1 & 0 & -Y_1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	4×4	13
X-slit with one finite and one infinite axis	$\mathsf{E}_{xfi} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & W_2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \mathsf{E}_{af} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & W_1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	4×4	10
Central with finite optical center	$\mathsf{E}_{cf} = -[\mathbf{t}]_\times \mathsf{R}$	3×3	8
Central with infinite optical center	$\mathsf{E}_{ci} = \begin{pmatrix} t_2 R_{13} - t_1 R_{23} & R_{31} & R_{32} \\ R_{13} & 0 & 0 \\ R_{23} & 0 & 0 \end{pmatrix}$	3×3	4

matrix R and a translation vector $\mathbf{t}$, according to (4). The matrix M is thus given as:

$$\mathsf{M}_{8\times 8} = \begin{pmatrix} \mathsf{I}_{4\times 4} & \mathbf{A}_1 & \mathbf{B}_1 & \mathbf{0} & \mathbf{0} \\ \mathsf{P} & \mathbf{0} & \mathbf{0} & \mathbf{A}_2 & \mathbf{B}_2 \end{pmatrix}$$

For a matching pair of rays, M must be rank-deficient. Here, this implies that its determinant is equal to zero. It can be developed to the following expression, where the Plücker coordinates $\mathbf{L}_1$ and $\mathbf{L}_2$ are defined as in equation (1):

$$\mathbf{L}_2^\mathsf{T} \underbrace{\begin{pmatrix} -[\mathbf{t}]_\times \mathsf{R} & \mathsf{R} \\ \mathsf{R} & \mathsf{0} \end{pmatrix}}_{\mathsf{E}_n} \mathbf{L}_1 = 0 \qquad (6)$$

We find the essential matrix E_n, as was done in [19].

5.3.2. Axial Cameras

Finite axis. As mentioned in §3, we adopt local coordinate systems where camera rays have $L_6 = 0$. Hence, the epipolar constraint (6) can be expressed by a reduced essential matrix of size 5×5, which acts on reduced Plücker vectors, consisting of the first five Plücker coordinates. This essential matrix is obtained from the non-central one E_n (6), by dropping its sixth row and column, leading to E_{af}, as given in table 4.

Note that this essential matrix is in general of full rank (rank 5), but may be rank-deficient. It can be shown that it is rank-deficient exactly if the axes of the two camera cut each other. In that case, the left and right null-vectors of E_{af} represent the camera axes of one view in the local coordinate system of the other one (one gets their Plücker vectors when adding a zero as 6th coordinate).

Infinite axis. The epipolar constraint (6) can be expressed by a reduced essential matrix E_{ai} (cf. table 4) of size 5×5, acting on reduced Plücker vectors, consisting of the last five Plücker coordinates (cf. table 2). It is always rank-deficient; its right null-vector is $(0, 0, R_{11}, R_{12}, R_{13})^\mathsf{T}$, which represents the second camera's axis, expressed in the first camera's coordinate system (to get its Plücker vector, add a zero as 1st coordinate). The left null-vector is $(0, 0, R_{11}, R_{21}, R_{31})^\mathsf{T}$, which represents the first camera's axis, expressed in the second camera's coordinate system.

5.3.3. X-Slit Cameras

Two finite axes. We get a reduced essential matrix E_{xff} (cf. table 4) of size 4×4, acting on reduced Plücker vectors of the form $(a_1, a_2, a_3, b_2)^\mathsf{T}$ (cf. §4.3).

Contrary to previous cases, the essential matrix now not only encodes motion, but also "intrinsic parameters" (the coefficients W_i and Y_i of the two cameras' second axes).

One finite and one infinite axis. We get a reduced essential matrix E_{xfi} (cf. table 4) of size 4×4, acting on reduced Plücker vectors of the form $(a_1, a_3, b_1, b_2)^\mathsf{T}$ (cf. §4.3). Again, it not only encodes motion, but also "intrinsic parameters" (the coefficients W_i of the two cameras' infinite axes).

5.3.4. Central Cameras

Finite optical center. As mentioned in §3, we here deal with camera rays of the form $(L_1, L_2, L_3, 0, 0, 0)^\mathsf{T}$. Hence, the epipolar constraint (6) can be expressed by a reduced essential matrix of size 3×3. We actually find here the "classical" 3×3 essential matrix $\mathsf{E}_{cf} = -[\mathbf{t}]_\times \mathsf{R}$ [11, 13].

Infinite optical center. The essential matrix in this case is E_{ci}, cf. table 4. This resembles the affine fundamental matrix [20], but is not the same: here, the essential matrix acts on 3D lines, not on image points. For example, the right null-vector of E_{ci} is $(0, R_{32}, -R_{31})^\mathsf{T}$, which represents the 3D line with Plücker coordinates $(0, 0, 0, R_{32}, -R_{31}, 0)^\mathsf{T}$. This is the line spanned by the two optical centers, i.e. the baseline (expressed in the first camera's coordinate system).

6. Conclusion

We have proposed a multi-view geometry for non-central cameras, the first to our knowledge. A natural hierarchy of camera models has been introduced, grouping cameras into classes depending on, loosely speaking, the spatial distribution of their projection rays. Two-view geometry was specialized in detail to different camera models. We hope that this theoretical work allows to define some common ground for recent efforts in characterizing the geometry of non-classical cameras.

Concerning possibilites for further work, geometrical relations between cameras of different types would be simple to derive along the lines used here, and all expressions can of course be transcribed in tensor notation. In this paper, we concentrated on the theory and did not address the issue of actually estimating the matching tensors and extracting motion parameters from them. It is relatively straightforward though to extract the motion parameters from the various essential matrices, due to their forms given in table 4. Experiments with the essential matrix for non-central cameras were successful, as also reported in [19], and experiments with intermediate camera types are ongoing.

Finally, we would like to note that, although motivated by the generic imaging model associating rays to pixels, the multi-view relations derived here hold naturally for any camera model that allows to attribute projection rays to image points with sub-pixel precision.

References

[1] S. Baker, S.K. Nayar. A theory of single-viewpoint catadioptric image formation. *IJCV*, 35(2), 1999.

[2] H. Bakstein, T. Pajdla. An overview of non-central cameras. *Computer Vision Winter Workshop, Ljubljana*, 2001.

[3] J.P. Barreto, H. Araujo. Paracatadioptric camera calibration using lines. *ICCV*, 2003.

[4] O. Faugeras, B. Mourrain. On the geometry and algebra of the point and line correspondences between n images. *ICCV*, 1995.

[5] D. Feldman, T. Pajdla, D. Weinshall. On the epipolar geometry of the crossed-slits projection. *ICCV*, 2003.

[6] C. Geyer, K. Daniilidis. A unifying theory of central panoramic systems and practical applications. *ECCV*, 2000.

[7] C. Geyer, K. Daniilidis. Paracatadioptric camera calibration. *PAMI*, 24(5), 2002.

[8] C. Geyer, K. Daniilidis. Mirrors in Motion: Epipolar geometry and motion estimation. *ICCV*, 2003.

[9] M.D. Grossberg, S.K. Nayar. A general imaging model and a method for finding its parameters. *ICCV*, 2001.

[10] R.I. Hartley, R. Gupta. Linear pushbroom cameras. *ECCV*, 1994.

[11] R.I. Hartley, A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[12] R.A. Hicks, R. Bajcsy. Catadioptric sensors that approximate wide-angle perspective projections. *CVPR*, 2000.

[13] H.C. Longuet-Higgins. A computer program for reconstructing a scene from two projections. *Nature*, 293, 1981.

[14] J. Neumann, C. Fermüller, Y. Aloimonos. Polydioptric camera design and 3D motion estimation. *CVPR*, 2003.

[15] T. Pajdla. Epipolar Geometry of Some Non-classical Cameras. *Computer Vision Winter Workshop, Bled*, 2001.

[16] T. Pajdla. Geometry of two-slit camera. Report CTU-CMP-2002-02, Czech Technical University, Prague, 2002.

[17] T. Pajdla. Stereo with oblique cameras. *IJCV*, 47(1-3), 2002.

[18] S. Peleg, M. Ben-Ezra, Y. Pritch. Omnistereo: Panoramic stereo imaging. *PAMI*, 23(3), 2001.

[19] R. Pless. Using many cameras as one. *CVPR*, 2003.

[20] L.S. Shapiro, A. Zisserman, M. Brady. 3D Motion Recovery via Affine Epipolar Geometry. *IJCV*, 16(2), 1995.

[21] H.-Y. Shum, A. Kalai, S.M. Seitz. Omnivergent stereo. *ICCV*, 1999.

[22] P. Sturm. Mixing catadioptric and perspective cameras. *Workshop on Omnidirectional Vision, Copenhagen*, 2002.

[23] P. Sturm, S. Ramalingam. A generic concept for camera calibration. *ECCV*, 2004,

[24] T. Svoboda. *Central Panoramic Cameras: Design, Geometry, Egomotion*. PhD Thesis, Czech Technical University, Prague, 1999.

[25] R. Swaminathan, M.D. Grossberg, S.K. Nayar. A perspective on distortions. *CVPR*, 2003.

[26] J. Yu, L. McMillan. General linear cameras. *ECCV*, 2004.

[27] A. Zomet, D. Feldman, S. Peleg, D. Weinshall. Mosaicing new views: The crossed-slit projection. *PAMI*, 25(6), 2003.