Mixing Catadioptric and Perspective Cameras

Peter Sturm

INRIA Rhône-Alpes

655 Avenue de l'Europe, 38330 Montbonnot, France

Peter.Sturm@inrialpes.fr ⊙ http://www.inrialpes.fr/movi/people/Sturm

Abstract

We analyze relations that exist between multiple views of a static scene, where the views can be taken by any mixture of para-catadioptric, perspective or affine cameras. Concretely, we introduce the notion of fundamental matrix, trifocal and quadrifocal tensors for the different possible combinations of these camera types. We also introduce the notion of plane homography for mixed image pairs. Generally speaking, these novel multi-view relations may form the basis for the typical geometric computations like motion estimation, 3D reconstruction or (self-) calibration. A few novel algorithms illustrating some of these aspects, are described, especially concerning what we call calibration transfer*, using fundamental matrices, and self-calibration from plane homographies.*

1. Introduction

This work has been motivated by the increasing interest of vision researchers and practitioners in the theory and use of omnidirectional cameras [12, 13, 3]. Our main goal is to contribute to a unified theory encompassing omnidirectional and traditional (perspective) cameras. We are especially interested in the study of geometrical and algebraic multi-view relations and their use in various geometrical computations like 3D reconstruction, self-calibration or motion estimation.

During the last decade and until today, multi-view relations between perspective views have been extensively studied [9, 5]. Among the most important concepts, one might cite the multi-linear matching constraints (fundamental matrix and trifocal tensors) that enable robust matching of images and are useful in motion estimation; self-calibration and the notion of uncalibrated 3D vision; multi-view reconstruction using factorization etc. We would like to derive analogous concepts for omnidirectional cameras. Some of these concepts are already known, e.g. the fundamental and essential matrices for para-catadioptric cameras [7, 17], epipolar geometry for general central catadioptric cameras [17], calibration [2, 6] and self-calibation [7, 10] of para-catadioptric cameras.

In this paper, we generalize some previous results and introduce several new concepts. Very important, in our opinion, is to study multi-view relations that hold between omnidirectional and perspective cameras, and their applications. An important potential application of omnidirectional cameras, especially in video-surveillance, is to locate a visual event, and to "guide" a perspective camera that might fixate and zoom in on the event, to take close-ups. A perspective camera with a large zoom is usually better modeled as an affine camera (typically, an orthographic one). So, we study the multi-view relations that hold between any combination of omnidirectional, perspective and affine cameras. Concretely, we will introduce the different types of fundamental matrices, and show the existence of trifocal and quadrifocal tensors, as well as plane homographies between pairs of views. We then briefly discuss their use for (self-) calibration, by giving novel algorithms for calibration transfer and self-calibration from planes.

Concerning the types of omnidirectional camera, our eventual goal is to treat the various types of *central catadioptric cameras* [1]. In this paper, we nearly exclusively consider *para-catadioptric* cameras, e.g. systems consisting of a parabolic mirror and an affine camera. Currently, we are not able to generalize several of our results to the other types of central catadioptric cameras (especially, those based on hyperbolic mirrors), the problem being that the multi-view relations are not multi-linear in general.

Organization. In §2, linear backprojection equations are explained, that allow to derive multi-linear matching constraints in §3. Self-calibration and calibration transfer using fundamental matrices and plane homographies for mixed types of cameras, is shown in §4. Experimental results illustrating these concepts are given in these sections. §5 concludes and describes perspectives.

Notations. We denote matrices in sans serif (e.g. R), vectors in bold (e.g. **t**), zero vectors as **0**. The symbol $\sim$ means equality of vectors or matrices up to scale, accounting for homogeneous coordinates. The 3×3 identity matrix is denoted as I. The skew-symmetric matrix associated with the cross-product is represented by $[\mathbf{v}]_\times$: $[\mathbf{v}]_\times \mathbf{w} = \mathbf{v} \times \mathbf{w}$.
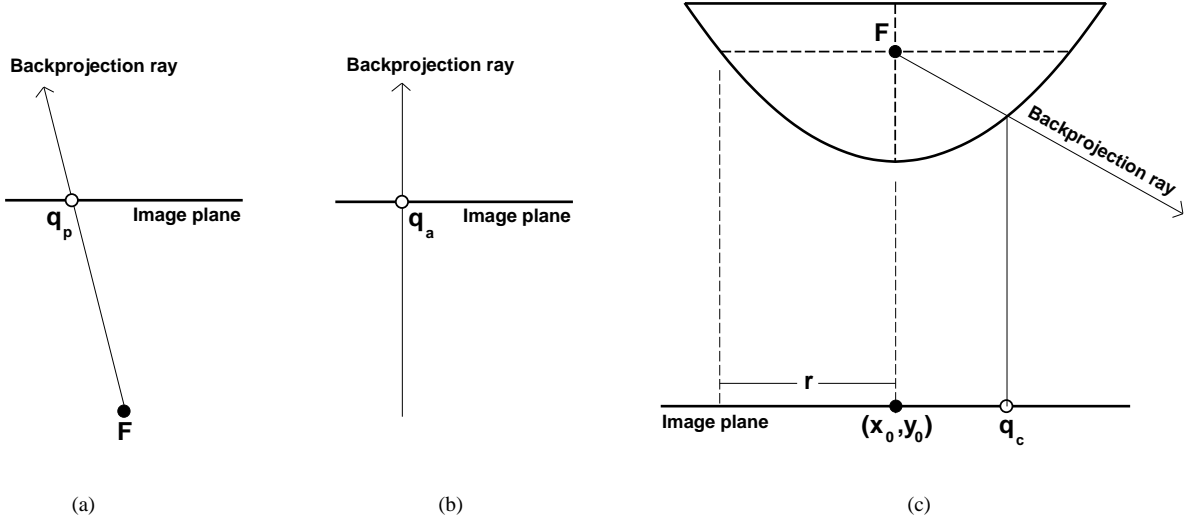
Figure 1. Camera models used in this paper. (a) Perspective projection: the optical center $\mathbf{F}$ is at position $\mathbf{t}_p$ (see text). (b) Affine projection: the (back-) projection rays are all parallel, and their direction is $\mathbf{r}_{a,3}$. (c) Para-catadioptric projection: the effective single viewpoint $\mathbf{F}$ is at position $\mathbf{t}_c$. The effective intrinsic parameters r, x_0 and y_0 are measured in pixels.

2. Camera Models

In this section, we explain the models we use for the camera types considered (see also figure 1). Since we are interested in deriving *multi-linear* constraints among views, we are keen to find linear projection equations. For perspective and affine cameras, 3×4 projection matrices *linearly* map homogeneous 3D point coordinates to homogeneous image point coordinates. As for catadioptric cameras, such linear projection equations do not seem to exist. What we will use instead are *backprojection matrices*, that map image point coordinates to the *direction* of the (back-) projection ray between the original 3D point and the (effective) optical center. It is possible to derive such mappings, that are linear, although not in standard image point coordinates, but in "lifted" ones, which shall be explained below. The backprojection equations derived in this section, will be used in section 3 to derive multi-view matching constraints.

2.1. Perspective Cameras

Let the projection matrix of a perspective camera be $\mathsf{P}_p \sim \mathsf{K}_p \mathsf{R}_p \begin{pmatrix} \mathtt{I} & -\mathbf{t}_p \end{pmatrix}$, where K_p is a calibration matrix (upper triangular 3×3), R_p a 3×3 rotation matrix and $\mathbf{t}_p$ the 3-vector of the optical center's coordinates.

All (finite) 3D points projecting onto a given image point $\mathbf{q}_p$ can be parameterized by a scale factor λ_p via:

$$\mathbf{Q} = \mathbf{t}_p + \lambda_p \mathbf{D}_p \ , \tag{1}$$

where the direction $\mathbf{D}_p$ of the projection ray is given by:

$$\mathbf{D}_p = (\mathsf{K}_p \mathsf{R}_p)^{-1} \mathbf{q}_p = \mathsf{R}_p^{\mathsf{T}} \mathsf{K}_p^{-1} \mathbf{q}_p \ . \tag{2}$$

2.2. Affine Cameras

Let the projection matrix of an affine camera be:

$$\mathsf{P}_a = \begin{pmatrix} \mathsf{K}_a \bar{\mathsf{R}}_a & \mathbf{t}_a \\ \mathbf{0}^{\mathsf{T}} & d_a \end{pmatrix} \ ,$$

with a 2×2 calibration matrix K_a, a 2-vector $\mathbf{t}_a$ and a 2×3 "amputated" rotation matrix $\bar{\mathsf{R}}_a$:

$$\bar{\mathsf{R}}_a = \begin{pmatrix} \mathbf{r}_{a,1}^{\mathsf{T}} \\ \mathbf{r}_{a,2}^{\mathsf{T}} \end{pmatrix} \ .$$

The missing third row gives the direction of the projection rays (they are all parallel). It is obtained (up to sign) as the cross-product of the other two rows: $\mathbf{r}_{a,3} = \mathbf{r}_{a,1} \wedge \mathbf{r}_{a,2}$.

All (finite) 3D points projecting onto an image point $\mathbf{q}_a$ (3-vector of homogeneous coordinates) can be parameterized by a scale factor λ_a' as follows:

$$\mathbf{Q} = \frac{1}{q_{a,3}} \mathsf{R}_a^{\mathsf{T}} \mathsf{K}_a^{-1} \begin{pmatrix} d_a & 0 & -t_{a,1} \\ 0 & d_a & -t_{a,2} \end{pmatrix} \mathbf{q}_a + \lambda_a' \mathbf{r}_{a,3} \ .$$

We will later use the following equation, obtained by multiplying the previous one by $q_{a,3}$:

$$q_{a,3} \mathbf{Q} = \underbrace{\mathsf{R}_a^{\mathsf{T}} \mathsf{K}_a^{-1} \begin{pmatrix} d_a & 0 & -t_{a,1} \\ 0 & d_a & -t_{a,2} \end{pmatrix}}_{\mathbf{B}_a} \mathbf{q}_a + \lambda_a \mathbf{r}_{a,3} \ , \tag{3}$$

with $\lambda_a = q_{a,3} \lambda_a'$ as free scale factor.

2

2.3. Para-Catadioptric Cameras

In this paper, we consider catadioptric systems consisting of a parabolic mirror and an affine camera. Concretely, the mirror is radially symmetric, and its surface may be represented by the quadric with the following matrix, for some scalar m defining the mirror's "curvature":

$$\Omega \sim \begin{pmatrix} 4m^2 & 0 & 0 & 0 \\ 0 & 4m^2 & 0 & 0 \\ 0 & 0 & 0 & -2m \\ 0 & 0 & -2m & -1 \end{pmatrix} \ .$$

Its two real focal points are the origin and the point at infinity of the Z-axis. Let the origin be the effective single viewpoint of the para-catadioptric system – we will sometimes also call it the *first focus*, whereas the point at infinity will be the *second focus*. Let P_c be the projection matrix of an affine camera, whose optical center is the second focus:

$$P_c = \begin{pmatrix} K_c & \mathbf{0} & \mathbf{t}_c \\ 0 & 0 & 0 & d_c \end{pmatrix} \ ,$$

with a 2×2 calibration matrix K_c and a 2-vector $\mathbf{t}_c$. The calibration matrix allows to represent all types of affine camera: para-perspective, weak perspective or orthographic. For easier reading, we present in the following only formulas for orthographic projection, but note that all derivations have also been done for the general affine camera. For the orthographic camera, we have:

$$P_c = \begin{pmatrix} k & 0 & 0 & t_{c,1} \\ 0 & k & 0 & t_{c,2} \\ 0 & 0 & 0 & d_c \end{pmatrix} \ .$$

Let $\mathbf{q}_c$ be the 3-vector of homogeneous coordinates of a point in the orthographic image. The direction $\mathbf{D}'_c$ of the effective (back-) projection ray (the line linking the effective viewpoint and the original 3D point $\mathbf{Q}$), can be computed as follows (we omit the subscripts c for clarity):

$$\mathbf{D}' = \begin{pmatrix} 4mkq_3 (q_1 d - q_3 t_1) \\ 4mkq_3 (q_2 d - q_3 t_2) \\ 4m^2 \left((q_1 d - q_3 t_1)^2 + (q_2 d - q_3 t_2)^2 \right) - k^2 q_3^2 \end{pmatrix} \ .$$

This is not linear in the image coordinates, however, by "lifting" them from the 3-vector $\mathbf{q}_c$ to the 4-vector[1]

$$\hat{\mathbf{q}}_c = \begin{pmatrix} q_{c,1}^2 + q_{c,2}^2 \\ q_{c,1} q_{c,3} \\ q_{c,2} q_{c,3} \\ q_{c,3}^2 \end{pmatrix} \ , \tag{4}$$

we obtain the following *linear* backprojection equation:

$$\mathbf{D}'_c = B_c \hat{\mathbf{q}}_c \ ,$$

---

[1]This is similar to the lifted coordinates in [7], although here they are obtained in a purely algebraic manner.

where (we again omit the subscripts c):

$$B = \begin{pmatrix} 0 & 4mkd & 0 & -4mkt_1 \\ 0 & 0 & 4mkd & -4mkt_2 \\ 4m^2 d^2 & -8m^2 dt_1 & -8m^2 dt_2 & 4m^2 \left(t_1^2 + t_2^2 \right) - k^2 \end{pmatrix} \ .$$

The parameters m, k, d, t_1 and t_2 are not independent, and we replace them by three *effective intrinsic parameters*: $r = \frac{k}{2md}$, $x_0 = \frac{t_1}{d}$ and $y_0 = \frac{t_2}{d}$ (cf. figure 1 (c)). With these, the backprojection matrix takes the form:

$$B_c \sim \begin{pmatrix} 0 & 2r & 0 & -2rx_0 \\ 0 & 0 & 2r & -2ry_0 \\ 1 & -2x_0 & -2y_0 & x_0^2 + y_0^2 - r^2 \end{pmatrix} \ . \tag{5}$$

All (finite) 3D points projecting onto a given image point $\mathbf{q}_c$ can now be parameterized by a scale factor λ_c via:

$$\mathbf{Q} = \mathbf{t}_c + \lambda_c \underbrace{R_c^\mathsf{T} B_c \hat{\mathbf{q}}_c}_{\mathbf{D}_c} \ , \tag{6}$$

where R_c and $\mathbf{t}_c$ represent the extrinsic parameters of the para-catadioptric system.

3. Multi-Linear Multifocal Matching Constraints

We use the backprojection equations laid out in the previous section for perspective, affine and para-catadioptric cameras, to obtain multifocal matching constraints. We proceed similarly to what has been done in the pure perspective case to derive multi-linear matching constraints [4, 19]. In the first paragraph, we describe the general scheme, and in the following ones, we concentrate on special cases.

3.1. General Scheme

Consider projections of a 3D point $\mathbf{Q}$ (non-homogeneous coordinates) in a set of views. Consider the general case of u perspective, v affine and w para-catadioptric views, with image points $\mathbf{q}_p^1, \ldots, \mathbf{q}_p^u$ in the perspective views, and analogously for the other camera types. In the following, superscripts are associated to different images. The backprojection equations (1), (3) and (6) may be grouped in a single equation system as shown in equation (7) on top of the following page. The vectors $\mathbf{D}_p^i, \mathbf{B}_a^j$ and $\mathbf{D}_c^k$ respectively depend linearly on the (lifted) image points, and are defined in equations (2), (3) and (6) respectively.

The matrix M of this equation system, in the following also called *joint matrix*, has $3(u + v + w)$ rows and $(u + v + w + 4)$ columns. Relations among the different projections of $\mathbf{Q}$ arise due to the fact that this matrix has a non-trivial null-vector (the vector containing the λ's and the coordinates of the 3D point $\mathbf{Q}$). Hence, M can not be of full column rank, i.e. its rank must be lower than $(u+v+w+4)$. This is equivalent to the statement that the determinants of all minors of size $(u+v+w+4)$ vanish. It is these determinants that give the multi-linear relations between matching image points in different views.

3

$$
\underbrace{\left[
\begin{array}{ccc|ccc|ccc|c}
\mathbf{t}_p^1 & \mathbf{D}_p^1 & & & & & & & & \mathtt{I} \\
\vdots & & \ddots & & 0 & & & 0 & & \vdots \\
\mathbf{t}_p^u & & \mathbf{D}_p^u & & & & & & & \mathtt{I} \\
\hline
\mathbf{B}_a^1 & & & \mathbf{r}_{a,3}^1 & & & & & & q_{a,3}^1\mathtt{I} \\
\vdots & & 0 & & \ddots & & & 0 & & \vdots \\
\mathbf{B}_a^v & & & & & \mathbf{r}_{a,3}^v & & & & q_{a,3}^v\mathtt{I} \\
\hline
\mathbf{t}_c^1 & & & & & & \mathbf{D}_c^1 & & & \mathtt{I} \\
\vdots & & 0 & & 0 & & & \ddots & & \vdots \\
\mathbf{t}_c^w & & & & & & & & \mathbf{D}_c^w & \mathtt{I}
\end{array}
\right]}_{\mathsf{M}}
\begin{bmatrix}
1 \\ \lambda_p^1 \\ \vdots \\ \lambda_p^u \\ \lambda_a^1 \\ \vdots \\ \lambda_a^v \\ \lambda_c^1 \\ \vdots \\ \lambda_c^w \\ -\mathbf{Q}
\end{bmatrix}
=
\begin{bmatrix}
\mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0}
\end{bmatrix}
\qquad (7)
$$

In the following, we make these relations explicit. We do this especially for the various two-view cases, which gives rise to different types of fundamental matrix. We then show that, like in the purely perspective case, trifocal and quadrifocal tensors exist, but no higher-order matching tensors.

3.2. Bifocal Constraints – Fundamental Matrices

With two views, of any mixture of camera types, the joint matrix M is of size 6×6. Consider for example the joint matrix for a perspective and a para-catadioptric view, shown here in detail:

$$
\underbrace{\left[
\begin{array}{ccc|ccc}
 & & 0 & 1 & 0 & 0 \\
\mathbf{t}_p & \mathsf{R}_p^{\mathsf{T}}\mathsf{K}_p^{-1}\mathbf{q}_p & 0 & 0 & 1 & 0 \\
 & & 0 & 0 & 0 & 1 \\
\hline
 & 0 & & 1 & 0 & 0 \\
\mathbf{t}_c & 0 & \mathsf{R}_c^{\mathsf{T}}\mathsf{B}_c\hat{\mathbf{q}}_c & 0 & 1 & 0 \\
 & 0 & & 0 & 0 & 1
\end{array}
\right]}_{\mathsf{M}}
\begin{bmatrix}
1 \\ \lambda_p \\ \lambda_c \\ \hline -Q_1 \\ -Q_2 \\ -Q_3
\end{bmatrix} = \mathbf{0} \; .
$$

This equation means that the 6×6 matrix M has a non-trivial null-vector, and hence must be of rank lower than 6. This in turn implies that all minors (submatrices) of size 6 are singular. The only minor of size 6 of M is the matrix itself. Hence, we obtain the bifocal matching constraint (the epipolar constraint) by developing its determinant. By doing so, one obtains an equation that is bilinear in the elements of $\mathbf{q}_p$ and $\hat{\mathbf{q}}_c$. This equation may thus be written in the following form:

$$
\mathbf{q}_p^{\mathsf{T}} \; \mathsf{F}_{pc} \; \hat{\mathbf{q}}_c = 0 \; ,
$$

where the matrix F_{pc} is of size 3×4 and its coefficients depend entirely on the entities defining the projections, i.e. the extrinsic parameters $\mathsf{R}_p, \mathbf{t}_p, \mathsf{R}_c, \mathbf{t}_c$ and the intrinsic parameters K_p and B_c.

One may recognize without difficulty in F_{pc} a fundamental matrix, which however relates here two views acquired with *different* camera types, and which does not have the usual dimensions, i.e. it is not even square as the fundamental matrix between two perspective views or between two para-catadioptric views [7].

This example concerned a perspective view, combined with a para-catadioptric one. The same findings hold for any mixture of the camera types considered in this paper:

- for two perspective views, the "traditional" fundamental matrix [11] is obtained. Any 3×3 matrix of rank 2 is a valid fundamental matrix.
- two affine views give a 3×3 affine fundamental matrix [14]. Affine fundamental matrices have a special form (upper left 2×2 submatrix is a null matrix).
- for two para-catadioptric views, a 4×4 fundamental matrix of rank 2 is obtained [7].
- mixtures of camera types lead to fundamental matrices of size 3×3 (perspective-affine) or 3×4 (perspective-catadioptric or affine-catadioptric). They can all be shown to be of rank 2.

A short comment is at order concerning affine cameras. In equation (7), image coordinates of affine views appear both in the first column (via the vectors $\mathbf{B}_a^j$) and in the last three columns (the identity matrices are multiplied by coordinates $q_{a,3}^j$). Thus, it is not obvious that a development of M's minors will lead to equations that are linear in the coordinates of each affine image point. Happily, it turns out that the equations can be factored such as to lead indeed to (multi-) linear equations.

In the following, we examine some properties of fundamental matrices of mixtures of a para-catadioptric with a perspective or an affine view.

3.3. Fundamental Matrices and Plane Homographies for Mixed View Pairs

These fundamental matrices are of size 3×4 (or 4×3, if we consider the transpose, which gives the "other direc-
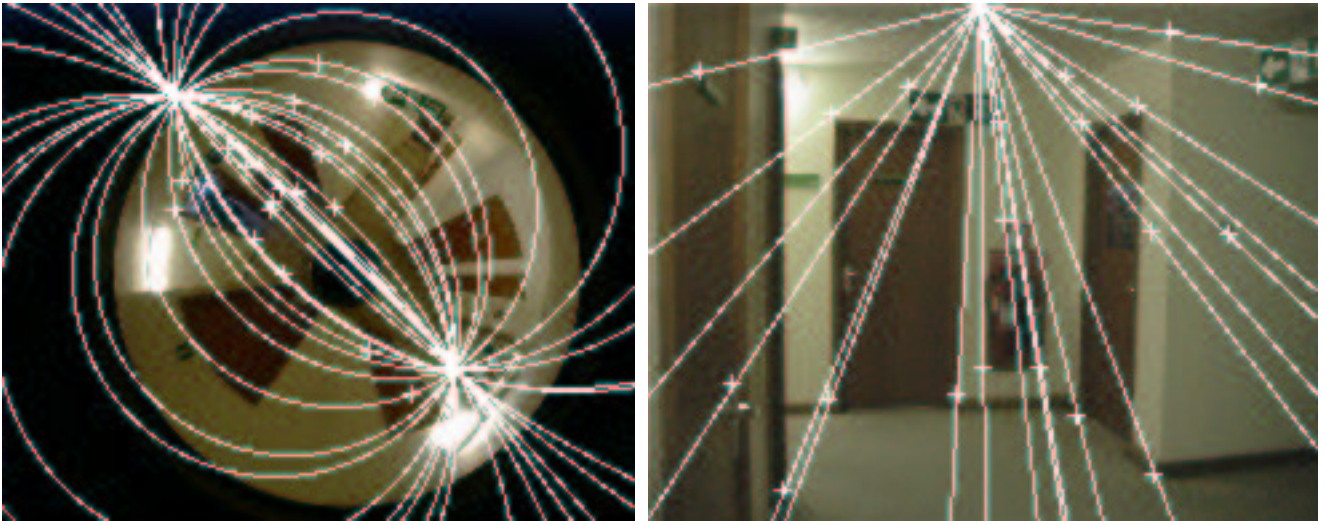
Figure 2. Estimated epipolar geometry for the stereo pair shown in figure 3. Points used to estimate F are shown by white crosses. For all 20 points, the epipolar circles in the catadioptric view and the epipolar lines in the perspective view, are shown. The two intersection points of the epipolar _circles_ are the two epipoles of the catadioptric view mentioned in the text, whereas the single intersection point of the epipolar _lines_ is the single epipole of the perspective view.

tion" of the epipolar constraint) and are of rank 2. The one-dimensional left null-space represents the epipole of the perspective (or affine) view, i.e. the image of the catadioptric view's effective viewpoint. The right null-space is two-dimensional. However, the fundamental matrix is only "valid" for 4-vectors of lifted coordinates, as defined in (4). There are exactly two right null-vectors of F (up to scale) that correspond to lifted coordinates. These are the two epipoles of the catadioptric view, i.e. the two projections of the perspective or affine camera's optical center (cf. [7]).

Products $F\hat{q}_c$ are 3-vectors, representing the usual epipolar lines in the perspective (or affine) view. As for products $F^T q_p$, these are 4-vectors, representing the epipolar conics of catadioptric views. Let $\mathbf{x} = F^T q_p$. The usual symmetric matrix of the associated epipolar conic is then given by:

$$\begin{pmatrix} 2x_1 & 0 & x_2 \\ 0 & 2x_1 & x_3 \\ x_2 & x_3 & 2x_4 \end{pmatrix} ,$$

which is a circle (the upper left 2×2 submatrix is a multiple of the identity matrix), which is in accordance with the known fact that epipolar conics of para-catadioptric systems are circles [17] (although this is only true for systems whose camera is perfectly orthographic).

Figure 2 shows the epipolar geometry (fundamental matrix), estimated by the analogon of the linear 8-point algorithm for the purely perspective case. Twenty manually selected points were used for the estimation. The estimated fundamental was also used for calibration transfer, see §4.3.

Figure 3. Stereo pair used in experiments.

Analogously to the purely perspective case, we may decompose the fundamental matrix to obtain the sometimes convenient epipole-homography form:

$$F \sim [\mathbf{e}_p]_\times H , \tag{8}$$

where $\mathbf{e}_p$ is the epipole in the perspective (or affine) view, and H a 3×4 *plane homography* matrix representing the mapping between the projections of points on some 3D scene plane. For example, the analogon to the *infinity homography* between two perspective views [9], for the case of a perspective and a para-catadioptric view, is given by the following 3×4 matrix:

$$H_\infty = K_p R_p R_c^T B_c , \tag{9}$$

with B_c defined as in equation (5). Using H_∞, we may derive the following expression for the fundamental matrix:

$$F \sim K_p^{-T} [R_p(\mathbf{t}_c - \mathbf{t}_p)]_\times K_p R_p R_c^T B_c .$$

Concerning the above *plane homographies* H, they can be derived for all 3D scene planes Π: let $\hat{\mathbf{q}}_c$ be the projection of any point on Π, then $\mathbf{q}_p \sim \mathsf{H}\hat{\mathbf{q}}_c$ is the projection of the same point in the perspective view, where H is a 3×4 matrix. Note however that there is an important difference to the purely perspective case. A plane homography, as given above, is only defined in one direction: the mapping of an image point in the para-catadioptric view, via the scene plane, and then onto the perspective view, is unique, whereas the reverse direction isn't. Indeed, the mapping of an image point in the perspective view, onto a scene plane, *is* unique, however the projection into the catadioptric camera, leads to two (theoretically possible) image points. It is possible to exclude the image point that is physically not possible, but the projection equation is still not linear in general, which prevents forming an homography matrix as for the other direction.

In section 4, we examine further properties of fundamental matrices and plane homographies and show their application for calibration.

3.4. Multifocal Constraints

Three views. Let us first consider the case of three views, with any mixture of camera types. The joint matrix M is of size 9×7 in this case. Its rank-deficiency implies that the determinants of all minors of size 7 vanish. In other words, the determinant of a submatrix of M, obtained by choosing any 7 rows, must be equal to zero. Since to each of the three views, three rows of M are associated, only the following two possibilities of choosing 7 rows exist:

$$(a) \quad 3 - 3 - 1$$
$$(b) \quad 3 - 2 - 2$$

where the figures refer to the number of rows chosen per one view. In case (a), it can be shown that the coordinates of the point in the view with a single contributed row, can be factored out from the resulting equation, and that we simply obtain the above bifocal relation for the two views with three contributed rows.

As for case (b), this gives rise to trilinear equations, which can be encoded via trifocal tensors. We identify tensors of size $4 \times 4 \times 4$ for the case of three para-catadioptric views, of size $4 \times 3 \times 3$ for a combination of one para-catadioptric and two perspective views, and so forth. Studying the properties of these tensors in more detail is beyond the scope of this paper though. As for trifocal tensors between triplets of cameras of the same type, the perspective case has been treated e.g. in [15] and the affine case in [18]. To our knowledge, no existing publication deals with the trifocal tensor for three para-catadioptric views or for the mixed configurations considered here.

Four views. In this case, the joint matrix is of size 12×8. Its rank-deficiency implies that the determinants of all minors of size 8 vanish. Analogously to the three-view case, we consider the different possibilities of choosing 8 rows of the joint matrix and their distribution among the four views:

$$(a) \quad 3 - 3 - 2 - 0$$
$$(b) \quad 3 - 3 - 1 - 1$$
$$(c) \quad 3 - 2 - 2 - 1$$
$$(d) \quad 2 - 2 - 2 - 2$$

Case (a) leads to trivial equations (always zero). Cases (b) and (c) lead to bifocal and trifocal relations respectively, whereas case (d) gives quadrifocal relations. Quadrifocal tensors for perspective views are dealt with e.g. in [8, 16].

More than four views. With five views, the joint matrix is of size 15×9. Obviously, there is no minor of size 9 that contains at least two rows per image. Hence, there are no multi-linear matching constraints between five views (or more), that can not be represented using bifocal, trifocal or quadrifocal ones. The same holds for the purely perspective case of course.

4. Calibration using Fundamental Matrices and Plane Homographies

4.1. Self-Calibration from Plane Homographies

Let H be the 3×4 homography between a catadioptric and a perspective view, associated with a 3D plane. It can be shown (proof omitted due to lack of space) that the null-vector of any such plane homography is:

$$\begin{pmatrix} r^2 + x_0^2 + y_0^2 \\ x_0 \\ y_0 \\ 1 \end{pmatrix}. \tag{10}$$

Hence, self-calibration of the para-catadioptric camera is possible from a single plane homography, defined with respect to a perspective camera, by computing its null-vector and extracting the three intrinsic parameters r, x_0 and y_0 from it in a straightforward manner.

This might also be explained intuitively as follows. A para-catadioptric camera can be calibrated by identifying line images (circles in the image plane, that constitute images of 3D lines). If we know a plane homography with respect to a perspective view, we may virtually create all possible lines images, by mapping all lines of the perspective view via the homography, to the catadioptric view. Calibration could then be done as e.g. shown in [6], or, better, via the above solution using H's null-vector.

This self-calibration approach was tested using the image pair of section 3. Seven manually selected points lying on the wall in the background of the right hand part of figure 3, were used to estimate the associated plane homography

H, using a straightforward linear algorithm. The catadioptric view's intrinsic parameters, extracted from H, were 2% (x_0), 0.6% (y_0) respectively 5% (r) off the ground truth values, obtained as the center of the circle circumscribing the image (x_0, y_0) or via constructor-provided values (r).

4.2. Self-Calibration from Fundamental Matrices

It has been shown in [7], that the vector given in (10) is a null-vector of any fundamental matrix that a para-catadioptric camera shares with another camera of the same type. Hence, fundamental matrices between catadioptric cameras are useful for self-calibration [7, 10].

This observation can be generalized to self-calibration from fundamental matrices between a para-catadioptric view and e.g. a perspective one: the above vector can actually be identified as the single null-vector (up to scale) of the 3×4 backprojection matrix B_c defined in equation (5). Since $\mathsf{F} \sim [\mathbf{e}_p]_\times \mathsf{K}_p \mathsf{R}_p \mathsf{R}_c^\mathsf{T} \mathsf{B}_c$ (cf. equations (8) and (9)), it follows that the null-vector of B_c is also in the null-space of any fundamental matrix F. Hence, given several fundamental matrices, the null-vector of B_c can be found by "intersecting" all their right null-spaces, and intrinsic parameters can then be extracted from it.

4.3. Calibration Transfer by Fundamental Matrices

Consider the surveillance scenario sketched in the introduction. A typical configuration might consist of one static catadioptric camera, which in addition can usually be assumed to be pre-calibrated, and one or several traditional cameras, perspective or affine. It might be useful to estimate the position of a perspective camera, relative to the catadioptric one. Another task might be to calibrate the perspective camera (e.g. after zooming or focusing), using the fundamental matrix and the available calibration of the catadioptric camera, which is what we call *calibration transfer*.

The analogous task for two perspective cameras has been developed in [21]. The development for the mixed perspective-catadioptric case, is similar. Concretely, given a fully calibrated catadioptric view, a perspective view that is calibrated besides the unknown focal length, and the fundamental matrix between the two, a closed-form solution for the focal length, in terms of the SVD (singular value decomposition) of the fundamental matrix, is possible. We very briefly outline the algorithm (derivations are based on an analogon to the classical Kruppa equations for perspective cameras [22]).

Let F be the 3×4 fundamental matrix between a catadioptric camera and a perspective one. We assume that the catadioptric camera is calibrated, so we know e.g. its backprojection matrix B_c. As for the perspective camera, we know all its intrinsic parameters, besides the focal length. Let its calibration matrix K_p be decomposed in its known part K_k and a diagonal matrix with the unknown focal length:

$$\mathsf{K}_p = \mathsf{K}_k \, \mathrm{diag}(f, f, 1) \ .$$

1. Compute a "semi-calibrated" fundamental matrix:

$$\mathsf{G} \sim \mathsf{K}_k^\mathsf{T} \, \mathsf{F} \, \mathsf{B}_c^+ \ ,$$

where B_c^+ is the Moore-Penrose pseudo-inverse. It can be shown that G is of the form:

$$\mathsf{G} \sim \mathrm{diag}(1, 1, f) \, [\mathbf{t}]_\times \mathsf{R} \ ,$$

for a rotation matrix R. From this form, the following steps can be derived (cf. [21]).

2. Compute the SVD of G (remember that it is of rank 2):

$$\mathsf{G} = \mathsf{U} \, \mathrm{diag}(r, s, 0) \, \mathsf{V}^\mathsf{T} \ .$$

3. The focal length can be computed by the following closed-form solution:

$$f = \sqrt{\frac{s^2 u_{32}^2 - r^2 u_{31}^2}{r^2(u_{11}^2 + u_{21}^2) - s^2(u_{12}^2 + u_{22}^2)}} \ .$$

The algorithm was applied using the fundamental matrix estimated for the stereo pair shown in §3. The estimated focal length for the perspective camera was about 8% off the ground truth, which is reasonable, considered that no non-linear optimization was performed and that the points were specified with an accuracy of probably less than a pixel.

5. Conclusion and Perspectives

We have shown that it is possible to obtain multi-linear matching constraints, especially fundamental matrices and trifocal tensors, for any mixed configuration of perspective, affine or para-catadioptric cameras. Our approach unifies the development of the previously known multifocal tensors for pairs or triplets of cameras of the same type, and substantially generalizes the concept in that it allows a transparent combination of cameras of different types.

We are only partly satisfied, since our basic goal is to get a complete generalization that encompasses *all* central catadioptric systems. We have already established (not shown here) the existence of a 3×6 fundamental matrix between a perspective or affine view, and a general central catadioptric view, which however only "works in one direction" (there is a linear mapping from points in the perspective view to the corresponding epipolar conic in the catadioptric view; the reverse however is not available yet). Thus, we currently do not know if a complete generalization of our approach (in the multi-linear framework), is possible.

In this paper, we have also outlined the possibility of self-calibration and calibration transfer using "mixed fundamental matrices" and "mixed plane homographies".

Throughout the paper we have, for the sake of clarity, only presented formulas for para-catadioptric systems whose camera is an *orthographic* one. Note however, that all formulas have an analogon for the general case of affine cameras, the difference being that lifted image coordinates are 6-vectors, resulting e.g. in 6×6 fundamental matrices between two such catadioptric systems and similarly the dimension 4 is replaced by 6 for the other concepts.

As for our future work, we have several perspectives, some of which should be relatively straightforward to realize, others maybe not. Motion estimation for mixed camera configurations should be straightforward, but has to be developed and tested. In this paper, we have introduced plane homographies only for one direction: from catadioptric to perspective views. We want to clarify if and how the mapping in the inverse direction can be represented linearly. It should be relatively straightforward to develop trifocal tensors for line images, again for mixed camera configurations. A complete study of matching relations for mixed configurations should also list in detail the different types of essential matrices. A detailed study of algebraic properties of such essential matrices and trifocal tensors is possible, but is not central to our interests.

Besides the above mentioned generalization of our approach to general central catadioptric cameras, we are interested in the possibility of factorization-based methods for 3D reconstruction from multiple catadioptric views. For practical applications, it might for example be fruitful to develop methods similar to "reconstruction from N views having one view in common" [20], for the case of several perspective views, overlapping with a single catadioptric camera.

References

[1] S. Baker, S. Nayar, "A Theory of Catadioptric Image Formation," *Proceedings of the International Conference on Computer Vision, Bombay, India*, pp. 35–42, 1998.

[2] J.P. Barreto, H. Araujo, "Issues on the Geometry of Central Catadioptric Image Formation," *Proceedings of the* IEEE *Conference on Computer Vision and Pattern Recognition, Kauai, USA*, pp. 422–427, Vol. II, 2001.

[3] R. Benosman, S.B. Kang, (Editors), *Panoramic Vision: Sensors, Theory, and Applications*, Springer Verlag, 2001.

[4] O. Faugeras, B. Mourrain, "On the Geometry and Algebra of the Point and Line Correspondences Between N Images," *Proceedings of the International Conference on Computer Vision, Boston, USA*, pp. 951–956, 1995.

[5] O. Faugeras, Q.-T. Luong, T. Papadopoulo, *The Geometry of Multiple Images*, MIT Press, 2001.

[6] C. Geyer, K. Daniilidis, "Catadioptric Camera Calibration," *Proceedings of the International Conference on Computer Vision, Kerkyra, Greece*, pp. 398–404, 1999.

[7] C. Geyer, K. Daniilidis, "Structure and Motion from Uncalibrated Catadioptric Views," *Proceedings of the* IEEE *Conference on Computer Vision and Pattern Recognition, Kauai, USA*, pp. 279–286, Vol. I, 2001.

[8] R.I. Hartley, "Computation of the quadrifocal tensor," *Proceedings of the European Conference on Computer Vision, Freiburg, Germany*, pp. 20–35, Vol. I, 1998.

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

[10] S.B. Kang, "Catadioptric Self-Calibration," *Proceedings of the* IEEE *Conference on Computer Vision and Pattern Recognition, Hilton Head Island, USA*, pp. 201–207, Vol. I, 2000.

[11] Q.-T. Luong, O. Faugeras, "The Fundamental Matrix: Theory, Algorithms and Stability Analysis," *International Journal of Computer Vision*, pp. 43–76, Vol. 17, No. 1, 1996.

[12] *Proceedings of the* IEEE *Workshop on Omnidirectional Vision, Hilton Head Island, USA*, IEEE Computer Society Press, 2000.

[13] *Proceedings of the* ICAR *Workshop on Omnidirectional Vision Applied to Robotic Orientation and Nondestructive Testing (NDT), Budapest, Hungary*, 2001.

[14] L.S. Shapiro, A. Zisserman, M. Brady, "3D Motion Recovery via Affine Epipolar Geometry," *International Journal of Computer Vision*, pp. 147–182, Vol. 16, No. 2, 1995.

[15] A. Shashua, "Algebraic Functions for Recognition," IEEE *Transactions on Pattern Analysis and Machine Intelligence*, pp. 779–789, Vol. 17, No. 8, 1995.

[16] A. Shashua, L. Wolf, "On the Structure and Properties of the Quadrifocal Tensor," *Proceedings of the European Conference on Computer Vision, Dublin, Ireland*, pp. 710–724, Vol. I, 2000.

[17] T. Svoboda, T. Pajdla, V. Hlavác, "Epipolar Geometry for Panoramic Cameras," *Proceedings of the European Conference on Computer Vision, Freiburg, Germany*, pp. 218–232, Vol. I, 1998.

[18] T. Thorhallsson, D. Murray, "The Tensors of Three Affine Views," *Proceedings of the* IEEE *Conference on Computer Vision and Pattern Recognition, Fort Collins, USA*, pp. 450–456, 1999.

[19] B. Triggs, "Matching Constraints and the Joint Image," *Proceedings of the International Conference on Computer Vision, Boston, USA*, pp. 338–343, 1995.

[20] M. Urban, T. Pajdla, V. Hlavac, "Projective reconstruction from N views having one view in common," *Proceedings of the ICCV Workshop on Vision Algorithms: Theory and Practice*, pp. 116-131, Springer Verlag, 2000.

[21] M. Urbanek, R. Horaud, P. Sturm, "Combining Off- and On-line Calibration of a Digital Camera," *Proceedings of the Third International Conference on 3D Digital Imaging and Modeling, Québec City, Canada*, pp. 99–106, 2001.

[22] C. Zeller, O. Faugeras, "Camera Self-Calibration from Video Sequences: the Kruppa Equations Revisited," Rapport de Recherche 2793, INRIA, France, 1996.