Canonic Representations for the Geometries of Multiple Projective Views

Q.-T. Luong[1,2] and T. Viéville[2]

[1] EECS, Cory Hall 211-215, University of California, Berkeley, CA 94720, USA
[2] INRIA, 2004 Route des Lucioles, 06561 Sophia-Antipolis, France

Abstract. We show how a special decomposition of general projection matrices, called *canonic* enables us to build geometric descriptions for a system of cameras which are invariant with respect to a given group of transformations. These representations are minimal and capture completely the properties of each level of description considered: Euclidean (in the context of calibration, and in the context of structure from motion, which we distinguish clearly), affine, and projective, that we also relate to each other. In the last case, a new decomposition of the well-known *fundamental matrix* is obtained. Dependencies, which appear when three or more views are available, are studied in the context of the canonic decomposition, and new composition formulas are established, as well as the link between local (ie for pairs of views) representations and global (ie for a sequence of images) representations.

1 Introduction and Background

1.1 Motivation

Three dimensional problems involving several views such as model-based recognition, stereovision or motion and structure from motion analysis have traditionally been studied under the assumption that the cameras are calibrated. The idea that several classical vision tasks could be performed without full calibration of the cameras, but only using some geometric information which can be obtained from mere point correspondences between uncalibrated images, has generated during the last few years an active research area, whose framework has been projective geometry [10]. More recently, affine geometry has been found to provide an interesting framework borrowing some nice characteristics from both Euclidean geometry and projective geometry.

However, one can remark that the representations adopted in the literature are of very disparate nature, and that often they are not even minimal. The relationships between different levels of representation has not been investigated thoroughly. Another important point which has not yet received much attention is the problem of dealing with multiple viewpoints to build a coherent representation in the case of uncalibrated cameras. Thus a unified representation is needed, to account in a single framework for the different geometric levels of representation, in the case of two, three, or more views. The principal aim of this paper is to describe such a framework, the *canonic decomposition*. The proofs and explicit formulas which could not be included because of lack of space can be found in [8]. In this section, some descriptions of the cameras

Lecture Notes in Computer Science, Vol. 800
Jan-Olof Eklundh (Ed.)
Computer Vision – ECCV '94

are presented. Section 2 gives the local canonic decomposition for two views. A global representation for the case of three and more views in Section 3. Section 4. discuss some relations between levels of representation.

1.2 The Projective Model

We consider the pinhole model. The main property of this camera model is that *the relationship between the world coordinates and the pixel coordinates is linear projective.* The consequence is that the relationship between 2-D pixel coordinates 3-D and any world coordinates can be described by a 3×4 matrix $\tilde{\mathbf{P}}$, called projection matrix, which maps points from $\mathcal{P}^3$ to $\mathcal{P}^2$:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \underbrace{[\mathbf{P}\ \ \mathbf{p}]}_{\tilde{\mathbf{P}}} \begin{bmatrix} \mathcal{X}_1 \\ \mathcal{X}_2 \\ \mathcal{X}_3 \\ \mathcal{X}_4 \end{bmatrix} \tag{1}$$

where the retinal projective coordinates x_1, x_2, x_3 are related to usual pixel coordinates by $(u, v) = (x_1/x_3, x_2/x_3)$ and the projective world coordinates $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{X}_3$, $\mathcal{X}_4$ are related to usual affine world coordinates by $(X, Y, Z) = (\mathcal{X}_1/\mathcal{X}_4, \mathcal{X}_2/\mathcal{X}_4, \mathcal{X}_3/\mathcal{X}_4)$. Note that since we assume a pinhole model, there is an optical center at finite distance. It is easy to see that it is uniquely defined if, and only if the 3×3 submatrix $\mathbf{P}$ is invertible, which is an assumption that we will use all the way through the paper. This is in opposition with another class of simplified models ranging from orthographic, weak perspective, to the affine camera [10, 11].

The goal of this paper is to exploit equation (1) to its fullest extend by deriving algebraic consequences (with geometric interpretations) of this equation in the case where several viewpoints are available. The generality of the approach comes from the fact that only projection matrices are manipulated in the paper, thus the results found do not depend on the different primitives one may be interested in, or the algorithms used for the estimation.

1.3 Calibrated and Uncalibrated Cameras

The projection matrix can be decomposed uniquely in the following way:

$$\tilde{\mathbf{P}} = \lambda_w \underbrace{\begin{bmatrix} \alpha_u & \gamma & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} 1\,0\,0\,0 \\ 0\,1\,0\,0 \\ 0\,0\,1\,0 \end{bmatrix}}_{\mathbf{P}_C} \underbrace{\begin{bmatrix} \mathbf{R}_w & \mathbf{T}_w \\ \mathbf{0}_3^T & 1 \end{bmatrix}}_{\mathbf{D}_w} \tag{2}$$

The 3×3 matrix $\mathbf{A}$, whose five entries are called *intrinsic parameters*, describes the change of retinal coordinate system. A camera is calibrated when the matrix $\mathbf{A}$ is known, so that one can use *normalized coordinates* which have an Euclidean meaning. The 4×4 matrix $\mathbf{D}_w$ describes the change of world coordinate system (the pose of the camera) called *extrinsic parameters.* The 5 intrinsic parameters and the 6 pose parameters together account for the 11 parameters of $\tilde{\mathbf{P}}$, which is a 3×4 matrix defined up to a scale factor. Here is a one-to-one mapping between the two representations, considering physically realizable systems.

2 A Local Canonic Decomposition

2.1 The Idea of the Canonic Decomposition

If two projective views are considered, the most complete description is given through the two projection matrices $\tilde{\mathbf{P}} = [\mathbf{P}, \mathbf{p}]$ and $\tilde{\mathbf{P}}' = [\mathbf{P}', \mathbf{p}']$. Since each matrix is defined up to a scale factor, this representation is not unique and the total number of parameters is 22. However, a total determination of these matrices cannot be done except in the case where a calibration object and its associated coordinate system are known. This total determination is not necessary: for example, in the Euclidean case, the choice of a particular world coordinate system is arbitrary, which means that the representation is defined up to a displacement. One is generally interested only in descriptions of the geometric relationship between the two images that are invariant by some group $\mathcal{G}$ of transformation of the projective space $\mathcal{P}^3$, which will be referred to as *descriptions of level* $\mathcal{G}$ or $\mathcal{G}$-invariant descriptions. The properties which can be recovered from these descriptions are those which are left invariant by the transformations of $\mathcal{G}$.

The idea, simple and powerful, is to consider the action of the group $\mathcal{G}$ on pairs of projection matrices ($\tilde{\mathbf{P}}$, $\tilde{\mathbf{P}}'$). It defines the equivalence relation:

$$\{(\tilde{\mathbf{P}}_1, \tilde{\mathbf{P}}'_1)\ \Re(\mathcal{G})\ (\tilde{\mathbf{P}}_2, \tilde{\mathbf{P}}'_2)\} \quad \Leftrightarrow \quad \{\exists \mathcal{T} \in \mathcal{G}\ ,\ \tilde{\mathbf{P}}_2 = \tilde{\mathbf{P}}_1 \mathcal{T}\ \wedge\ \tilde{\mathbf{P}}'_2 = \tilde{\mathbf{P}}'_1 \mathcal{T}\}$$

In each orbit, we choose the simplest form for the first projection matrix:

- for calibrated cameras, the normalized coordinate system: $\mathcal{I} = \mathbf{A}\mathbf{P}_C = [\mathbf{A}, \mathbf{0}]$
- for uncalibrated cameras, the pixel coordinate system: $\mathcal{I} = \mathbf{P}_C = [\mathbf{I}_3, \mathbf{0}]$.

Just by taking into the structure of the elements of $\mathcal{G}$, this yields a particular second projection matrix $\mathcal{I}'$, which we have found to have a remarkable interpretation in terms of geometric quantities. Thus, the principle is:

> *The matrices $\mathcal{I}$, $\mathcal{I}'$, expressed as functions of a pair of generic projection matrices $\tilde{\mathbf{P}}$, $\tilde{\mathbf{P}}'$, such that there is a unique decomposition, called canonic:*
>
> $$\tilde{\mathbf{P}} = \mathcal{I}\mathcal{T} \qquad \tilde{\mathbf{P}}' = \mathcal{I}'\mathcal{T} \tag{3}$$
>
> *with $\mathcal{T}$ being an element of $\mathcal{G}$, provide a complete description of the geometric properties* of two projective views *which are left invariant by the group of transformation $\mathcal{G}$.*

Note that here the invariant is attached to the set of camera, and not a set of 3D objects observed by the cameras. Let us list some consequences of this construction:

- The sum of the number of parameters in the representation $\mathcal{I}, \mathcal{I}'$ and in the generic transformation $\mathcal{T}$ has to be 22.
- Every quantity which depends only on the projection matrices and is invariant with respect to $\mathcal{G}$ is also a function of $\mathcal{I}$ and $\mathcal{I}'$.
- The quantities which appear in matrix $\mathcal{T}$ are not measurable from two views using the representation of level $\mathcal{G}$. But they may be expressed using representations of the previous level, instead.
- The decomposition provides a tool for explicitly building a pair of projection matrices $\tilde{\mathbf{P}}, \tilde{\mathbf{P}}'$ from the invariants obtained with respect to $\mathcal{G}$, which captures all the properties of a pair of views up to a transformation of $\mathcal{G}$. For example, given a particular fundamental matrix $\mathbf{F}$, we can obtain a projective 3D reconstruction, like [1, 5].

2.2 Classical Groups

The most general group of transformations of $\mathcal{P}^3$ is the group of homographies $\mathcal{GL}_4$.

The Affine Group Since the direction of a line $\mathbf{l} = [\mathbf{d}, d_4]^T$ can be represented by its intersection with the plane at infinity Π_∞, defined by $\mathcal{X}_4 = 0$, the conservation of parallelism by a transformation of $\mathcal{P}^3$ is equivalent to the fact that it leaves the plane at infinity invariant. For any given plane Π of $\mathcal{P}^3$, and any pair of cameras, there is a 3×3 matrix $\mathbf{H}_\Pi$ such that images of points of the plane are related by the projective linear relation: $\mathbf{m}' = \mathbf{H}_\Pi \mathbf{m}$. This matrix is invertible in the general case, and has the expression [12]:

$$\mathbf{H}_\Pi = \mathbf{A}'(\mathbf{R} + \frac{1}{d}\mathbf{t}\mathbf{n}^T)\mathbf{A}^{-1} \tag{4}$$

where $\mathbf{n}$ is the normal vector of the plane and d the distance to the origin. The limit as $d \to \infty$ for this expression is $\mathbf{H}_\infty$, the homography of the plane Π_∞, the ***infinity homography***. The matrix $\mathbf{H}_\infty$ is proportional to $\mathbf{Q} = \mathbf{A}'\mathbf{R}\mathbf{A}^{-1}$, and hence depends only on the rotational component. The other component of the representation, the epipole $\mathbf{e}'$, depends only on the translational component, since it is proportional to $\mathbf{s} = \mathbf{A}'\mathbf{T}$. These two quantities together define the uncalibrated motion, called ***Qs-representation*** (associated to the affine unimodular group) in [13] which appears to be the simplest generalization of what is well known in the calibrated case, with identical laws of composition as will be seen in section 3.

The Similarity Group A transformation of $\mathcal{P}^3$ leaves the absolute conic Ω of Π_∞ ($\mathcal{X}_1^2 + \mathcal{X}_2^2 + \mathcal{X}_3^2 = 0$) invariant if, and only if it is a similarity, which is a rigid displacement multiplied by a scale factor [4, 2]. The knowledge of the conic ω, image of Ω by the camera, is equivalent to that of the intrinsic parameters. This conic determines the angle between optical rays, which is coherent with the fact that the similarities conserve angles.

The Special Euclidean Group By contrast with the structure from motion paradigm where there is an ambiguity between the amount of displacement, represented by $\|\mathbf{T}\|$, and the depth of objects (only the direction of translation can be determined), in the calibration, or reconstruction paradigm, the scale factor is known, and the relevant group is the group of displacements (a rotation followed by a translation).

2.3 A New Representation for Projective Structure

Factorizations of the Fundamental Matrix It has been shown in [6] that the matrix $\mathbf{H}_\Pi$ which relates the two images of a plane, is linked to the fundamental matrix by the following system of equations:

$$\mathbf{H}_\Pi^T\mathbf{F} + \mathbf{F}^T\mathbf{H}_\Pi = \mathbf{0} \tag{5}$$

which is equivalent to the condition (***compatibility*** of $\mathbf{F}$ and $\mathbf{H}$):

$$\mathbf{F} = [\mathbf{e}']_\times \mathbf{H}_\Pi \tag{6}$$

The decomposition (6) is not unique, since $\mathbf{H}_\Pi$ can defined by any plane of $\mathcal{P}^3$, and thus there remains three degrees of freedom.

The S-matrix Let us define a special matrix $\mathbf{S}$ compatible with $\mathbf{F}$, and *which is only a function of* $\mathbf{F}$:

$$\mathbf{F} = \underbrace{\frac{\mathbf{e}'}{\|\mathbf{e}'\|^2}\mathbf{e}'^T\mathbf{F}}_{\mathbf{0}} + [\mathbf{e}']_\times \underbrace{(-\frac{[\mathbf{e}']_\times}{\|\mathbf{e}'\|^2}\mathbf{F})}_{\mathbf{S}} \tag{7}$$

This relation shows that $\mathbf{S}$ is determined by $\mathbf{F}$ (since $\mathbf{e}'$ is determined by $\mathbf{F}$), and that $\mathbf{F}$ is determined by $\mathbf{S}$ and $\mathbf{e}'$. An analogy can be noted with the decomposition of the essential matrix $\mathbf{E} = [\mathbf{T}]_\times\mathbf{R}$ as the product of an antisymmetric matrix and a rotation matrix $\mathbf{R}$, since the fundamental matrix is decomposed as the product of an antisymmetric matrix and the singular special matrix $\mathbf{S}$, which we call *epipolar projection* matrix.

A Geometric Interpretation The matrix $\mathbf{S}$ is the correspondence defined by the plane $\Pi_{e'}$ which contains the optical center of the second camera, and whose image in the second camera is the line $\langle\mathbf{e}'\rangle$. We see that this is consistent with definition (7). The matrix $\mathbf{F}$ maps points to lines, the matrix $[\mathbf{e}']_\times$ either maps lines to points or points to lines, thus from (7) the matrix $\mathbf{S}$ maps points to points. More specifically, a point $\mathbf{m}$ is mapped to $\mathbf{m}'_1 = \mathbf{e}' \times \mathbf{F}\mathbf{m}$, which is the intersection of the epipolar line of $\mathbf{m}$ with the line $\langle\mathbf{e}'\rangle$. We can note that this point is always defined as soon as $\mathbf{m} \neq \mathbf{e}$ since the distinctive property of the line $\langle\mathbf{e}'\rangle$ is that it does not contain the point $\mathbf{e}'$, as we always have $\mathbf{e}'^T\mathbf{e}' = \|\mathbf{e}'\|^2 \neq 0$. The interpretation of (7) is that the epipolar line $\mathbf{l}'_m = \mathbf{F}\mathbf{m}$ is defined by joining the epipole $\mathbf{e}'$ and the point $\mathbf{m}'_1 = \mathbf{S}\mathbf{m}$ (intersection of the epipolar line and the epipole), thus the transformation $\mathbf{S}$ and the epipole $\mathbf{e}'$ completely define the epipolar geometry.

The epipole $\mathbf{e}'$ depends on two independent parameters, since it is defined only up to a scale factor. The transformation $\mathbf{S}$ is a linear projection of a projective plane (the first retina) on a projective line $\langle\mathbf{e}'\rangle$, thus it is defined by a 2×3 matrix defined up to a scale factor. Since the line $\langle\mathbf{e}'\rangle$ is also defined by the same parameters as the epipole, we see that the knowledge of the linear projection (5 parameters) and the epipole (2 parameters) completely define the 3×3 matrices $\mathbf{S}$ and $\mathbf{F}$, which is consistent with the result that the fundamental matrix depends on 7 parameters [3].

Planes Once the fundamental matrix is known, any plane Π can be characterized by the vector $\mathbf{r}_\Pi$ such that:

$$\mathbf{H}_\Pi = \mathbf{S} + \mathbf{e}'\mathbf{r}_\Pi^T \tag{8}$$

In particular, we obtain, for the plane at infinity:

$$\mathbf{H}_\infty = \mathbf{S} + \mathbf{e}'\underbrace{\mathbf{e}'^T\mathbf{H}_\infty/\|\mathbf{e}'\|^2}_{\mathbf{r}_\infty^T} \tag{9}$$

We also see that $\mathbf{r}_\Pi = \mathbf{r}_\infty + \mathbf{A}^T\mathbf{n}/d$, and thus we can interpret the vector $\mathbf{r}_\Pi$ as the projective characterization of the plane Π. An affine characterization would be the vector $\nu = \mathbf{A}^T\mathbf{n}/d$, whereas a Euclidean characterization would be $\mathbf{n}/d$.

2.4 The Canonic Decomposition for Two Views

The canonic decomposition for two views is summarized in table 1, in which we mention:

- the characteristic properties and generic decomposition of a member of each of these group of transformations,
- a canonic decomposition of the form (3) of two projection matrices. The quantities above the horizontal line are the elements of the invariant description, the quantities under that line are non-measurable,
- indication of links with the previous level,
- the number of parameters, whose sum is exactly 22.

It should be noted that the invariants $\mathbf{e}'$, $\mathbf{H}_\infty$, $\mathbf{S}$ are projective, thus defined only up to a scale factor, as well as the matrices $\mathcal{A}$ and $\mathcal{H}$. It can be verified that this reflects coherently the fact that the projection matrices $\tilde{\mathbf{P}}, \tilde{\mathbf{P}}'$ are also projective quantities.

Using the projective epipole $\mathbf{e}'$ as an invariant would have been perfectly adequate in this two-view analysis, because the norm of this quantity is not constrained in any way by two mere views. However, we will see in the next section that it is constrained if three views are considered, and thus, in order to be coherent with the sequel, we have taken a scale-invariant representation for the epipole, the normalized epipole $\mathbf{e}'_N = \mathbf{e}/\|\mathbf{e}'\|$. The vectors $\mathbf{r}_\infty$, $\mathcal{L}$ and the scalar ν have to be scaled accordingly, resulting in the quantities:

$$\mathbf{r}_{\infty N} = \mathbf{H}^T\mathbf{e}'_N \quad \mathbf{L}_N = \mathbf{P}'^T\mathbf{e}'_N \quad \nu_N = \mathbf{p}'^T\mathbf{e}'_N \tag{10}$$

3 A Canonic Representation for Multiple Views

Whereas the composition relations in the displacement-invariant case are straightforward, in the similarity-invariant case [7] (resp. homography-invariant one [13]), it has been found that 11 (resp. 18) parameters are needed to represent the relations between three cameras. This is more than 2×5 (resp. 2×7), but less than 3×5 (resp. 3×7).

Three projective views are now considered, and their most complete description is given through the three projection matrices $\tilde{\mathbf{P}}_i = [\mathbf{P}_i, \mathbf{p}_i]$, $i = 1, 2, 3$, which depend on the total number of 33 independent parameters. The canonic decomposition for three views is defined as the unique representation:

$$\tilde{\mathbf{P}}_1 = \mathcal{I}_1\mathcal{T} \quad \tilde{\mathbf{P}}_2 = \mathcal{I}_2\mathcal{T} \quad \tilde{\mathbf{P}}_3 = \mathcal{I}_3\mathcal{T} \tag{11}$$

where $\mathcal{I}_1$ and $\mathcal{I}_2$ have the same form as in the canonic decomposition for two views (3). Of course, the form of $\mathcal{I}_3$ is expected to be in general different from the form of $\mathcal{I}_2$. Let us list the consequences of this construction:

- The two-view canonic decomposition and its properties are extended.
- There are two descriptions for the invariants of three views, one built upon the pair of descriptions 1-2, 1-3, the other upon the pair of descriptions 1-2, 2-3. They are more than the two descriptions for two views, including some additional parameters, which cannot be determined from them. Rather, these parameters are functions of descriptions of the previous level.
- The equivalence of the two forms of the alternative descriptions for three views gives the dependency of the composed description 1-3 (resp. 2-3) over the descriptions for 1-2 and 2-3 (resp. 1-3), and the additional parameters.
- The additional parameters can be determined from the knowledge of the *three* descriptions 1-2, 2-3, and 1-3. It means that knowing all the triples of descriptions for two views is equivalent to a description for three views. There are some subtleties with scale factors, discussed in [8]. From the count of parameters, it is seen that the triple of descriptions for two views is not a minimal representation.

We have summarized in table 2 the results specific to the canonic decomposition of a triple of projection matrices:

- the nature of the two equivalent invariant descriptions, the quantities above the horizontal line being the elements of the invariant description for two views, the quantities under that line being the additional parameters, which are measurable from three views but not from two pairs of invariant description for two views,
- the two alternative expressions for $\tilde{\mathbf{P}}_3$, as a function of the description 1-2, 1-3, or of the descriptions 1-2, 2-3,
- the definition of the additional elements, as a function of the description of previous level,
- the number of parameters, whose sum is exactly 33.

One advantage of the previous formalism is that the generalization of the canonic decomposition to the case of N views is straightforward. Thus the elements of the description are exactly the same than for three views, and can be summarized in the table bellow where it can be verified that the total number of parameters is $11N$:

EUCLIDEAN	displacement	6
(calibration)	intrinsic parameters	5N
	rotations	3 (N-1)
	translations	3 (N-1)
EUCLIDEAN	similarity	7
(motion)	intrinsic parameters	5N
	rotations	3 (N-1)
	directions of translations	2 (N-1)
	ratio of translation norms	N-2
AFFINE	affine transformation	12
	infinity homographies	8 (N-1)
	normalized epipoles	2 (N-1)
	ratios of epipole norms	N-2
PROJECTIVE	homography	15
	epipolar projections	5 (N-1)
	normalized epipoles	2 (N-1)
	differences of r_∞ vectors	3 (N-2)
	ratios of epipole norms	N-2

4 Relations between the Levels of Representation

From tables 1 and 2 we remark that each invariant description of a given level is formulated in terms of descriptions of the previous level. In this section, we present some more relations which are not explicitly described by these tables.

Projective Representation and Affine Representation Let us examine the case of three views. There are the following relations between fundamental matrices and infinity homography matrices:

$$\begin{aligned}\mathbf{F}_{13} &= \mathbf{H}^{*}_{\infty 23}\mathbf{F}_{12} + \mathbf{F}_{23}\mathbf{H}_{\infty 12} = (\det \mathbf{H}_{\infty 23})(\mathbf{H}^{T}_{\infty 32}\mathbf{F}_{12} + \mathbf{F}^{T}_{32}\mathbf{H}_{\infty 12}) \\ &= (\det \mathbf{H}_{\infty 23})(\mathbf{H}^{T}_{\infty 32}([\mathbf{e}_{21}]_\times - [\mathbf{e}_{23}]_\times)\mathbf{H}_{\infty 12})\end{aligned} \tag{12}$$

It can be noted that the system of equations obtained by writing (12) between the three images can not determine the infinity homography matrices from the knowledge of the three fundamental matrices, because there are 21 parameters in the affine representation, versus only 18 in the projective representation. The additional knowledge needed corresponds to one of the three vectors $\mathbf{r}_\infty$ defined in (9), which identify the plane at infinity. However, if this quantity is identified *locally* between any two views of an image sequence, it can be *propagated* along the whole sequence, thanks to (12).

Euclidean Representation and Affine Representation It is easy to see that the relation:

$$\mathbf{H}_\infty = \mathbf{A}'\mathbf{R}\mathbf{A}^{-1} \tag{13}$$

together with the relation $\mathbf{e}' = \mathbf{A}'\mathbf{t}$ allows one to determine directly the motion parameters, the rotation $\mathbf{R}$ and the direction of translation $\mathbf{t}$ from the affine representation, if the intrinsic parameters are determined.

We examine now the relation with the intrinsic parameters. The fact that $\mathbf{R}$ is a rotation matrix is equivalent to:

$$\mathbf{R}\mathbf{R}^T = \mathbf{I}_3 \tag{14}$$

Substituting $\mathbf{R} = \mathbf{A}'^{-1}\mathbf{H}_\infty\mathbf{A}$ obtained from equation (13) into (14),

$$\mathbf{K}' = \mathbf{H}_\infty\mathbf{K}\mathbf{H}_\infty^T \tag{15}$$

where the matrices $\mathbf{K} = \mathbf{A}\mathbf{A}^T$ and $\mathbf{K}' = \mathbf{A}'\mathbf{A}'^T$ represent the dual of the image of the absolute conic ω in each camera coordinate system. Each of these matrices is symmetric and defined only up to a scale factor, thus they depend on five independent parameters. It can be seen that relation (15) allows us to *update* camera calibration through a sequence of images where they do not remain constant, once the initial camera parameters are known. It can also be used for self-calibration.

The five constraints (15) on the intrinsic parameters are linear, whereas from the projective invariants, only two quadratic constraints were obtained [9, 3]. These last constraints are in fact implied by the former ones.

5 Conclusion

This paper lays the ground for further studies about problems involving 3D information, multiple viewpoints, and uncalibrated cameras. It confirms the interest of the affine representation, which turns out to yield simple and powerful descriptions.

We have described the *canonic decomposition*, an idea to account in a single framework for the different geometric levels of representation, in the case of two views, three views, or more. The approach is very general, since it involves only reasoning about the projection matrices. We first presented new descriptions for the affine and projective geometries of two views, which are respectively the *infinity homography matrix* and the *epipolar projection matrix*, which have been described from both an algebraic and geometric viewpoint. Then, a coherent hierarchy of representations has been studied. In particular, we have exhibited minimal and complete representations for each level of description, and showed clearly which elements of representation change and which ones are conserved across two different levels. These representations are description of the geometry of the cameras which are invariant with respect to a given group of transformations. The relationships which occur between the different levels of representation have been described. In the case of multiple views, new representations and their associated composition formulas have been established. They allow to deal with the case of multiple viewpoints while working with uncalibrated cameras, by relating local and global descriptions. Some consequences of the representation have begun to be explored [13]: reconstructions from multiple views, theory and computational methods to recover the invariant descriptions studied in this paper, from points and lines in uncalibrated images.

References

1. O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In *Proc. European Conference on Computer Vision*, pages 563–578, 1992.
2. O.D. Faugeras. *Three-dimensional computer vision: a geometric viewpoint.* MIT Press, 1993.
3. O.D. Faugeras, Q.-T. Luong, and S.J. Maybank. Camera self-calibration: theory and experiments. In *Proc. European Conference on Computer Vision*, pages 321–334, Santa-Margerita, Italy, 1992.
4. O.D. Faugeras and S.J. Maybank. Motion from point matches: multiplicity of solutions. *The International Journal of Computer Vision*, 4(3):225–246, 1990. also INRIA Tech. Report 1157.
5. R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proc. of the conf. on Computer Vision and Pattern Recognition*, pages 761–764, Urbana, 1992.
6. Q.-T. Luong and O.D. Faugeras. Determining the Fundamental matrix with planes: unstability and new algorithms. In *Proc. Conference on Computer Vision and Pattern Recognition*, pages 489–494, New-York, 1993.
7. Q.-T. Luong and O.D. Faugeras. Self-calibration of a stereo rig from unknown camera motions and point correspondences. In A. Grun and T.S. Huang, editors, *Calibration and orientation of cameras in computer vision.* Springer-Verlag, 1993. To appear. Also presented at XVII ISPRS, Washington, and INRIA Tech. Report RR-2014.
8. Q.-T. Luong and T. Viéville. Canonic representations for the geometries of multiple projective views. Technical Report UCB/CSD-93-772, University of California at Berkeley, Sept 1993. A shorter version appeared at ECCV'94.
9. S.J. Maybank and O.D. Faugeras. A Theory of Self-Calibration of a Moving Camera. *The International Journal of Computer Vision*, 8(2):123–151, 1992.
10. J. L. Mundy and A. Zisserman, editors. *Geometric invariance in computer vision.* MIT Press, 1992.
11. L.S. Shapiro, A. Zisserman, and M. Brady. Motion from point matches using affine epipolar geometry. Technical Report OUEL 1994/93, Oxford University., June 1993.
12. R.Y. Tsai and T.S. Huang. Estimating Three-dimensional motion parameters of a rigid planar patch, II: singular value decomposition. *IEEE Transactions on Acoustic, Speech and Signal Processing*, 30, 1982.
13. T. Viéville and Q.T. Luong. Motion of points and lines in the uncalibrated case. Technical Report RR-2054, INRIA, Sept 1993.

Table 1. The geometries of two views: canonic representation

EUCLIDEAN (calibration)	**displacements** *preserve* **angles, distances** $\|\mathcal{D}(\mathrm{i})\| = \|\mathrm{i}\|$	
$\mathcal{SE}_3$	$\mathcal{D} = \begin{bmatrix} \mathcal{R} & \mathcal{T} \\ 0_3^T & 1 \end{bmatrix}$ $\mathcal{R}$: rotation matrix $\mathcal{T}$: translation vector	6
invariant description	$\mathbf{A},\mathbf{A}'$: intrinsic parameters of cameras	5+5
	$\mathbf{R}$: rotation from camera 1 to camera 2	3
	$\mathbf{T}$: translation from camera 1 to camera 2	3
canonic decomposition	$\begin{cases} \tilde{\mathbf{P}} = \mathbf{A}[\mathbf{I}_3, 0]\,\mathcal{D} \\ \tilde{\mathbf{P}}' = \mathbf{A}'[\mathbf{R}, \mathbf{T}]\,\mathcal{D} \end{cases}$ $\mathbf{R} = \mathbf{R}'_w\mathbf{R}_w^T$, $\mathbf{T} = \mathbf{T}'_w - \mathbf{R}'_w\mathbf{R}_w^T\mathbf{T}_w$; $\mathcal{R} = \mathbf{R}_w$, $\mathcal{T} = \mathbf{T}_w$ where $\tilde{\mathbf{P}} = \mathbf{A}[\mathbf{R}_w, \mathbf{T}_w]$, $\tilde{\mathbf{P}}' = \mathbf{A}'[\mathbf{R}'_w, \mathbf{T}'_w]$	
EUCLIDEAN (motion)	**similarities** *preserve* **angles, relative distances** $\mathcal{S}(\Omega) = \Omega$	
$\lambda\mathcal{SE}_3$	$\mathcal{S} = \begin{bmatrix} \mathcal{R} & \mathcal{T} \\ 0_3^T & \lambda \end{bmatrix}$ $\mathcal{R}$: rotation matrix $\mathcal{T}$: translation vector λ: non-null scalar	7
invariant description	$\mathbf{A},\mathbf{A}'$: intrinsic parameters of cameras	5+5
	$\mathbf{R}$: rotation from camera 1 to camera 2	3
	$\mathbf{t}$: direction of translation from camera 1 to camera 2	2
canonic decomposition	$\begin{cases} \tilde{\mathbf{P}} = \mathbf{A}[\mathbf{I}_3, 0]\,\mathcal{S} \\ \tilde{\mathbf{P}}' = \mathbf{A}'[\mathbf{R}, \mathbf{t}]\,\mathcal{S} \end{cases}$ $\mathbf{R} = \mathbf{R}'_w\mathbf{R}_w^T$, $\mathbf{t} = \mathbf{T}/\|\mathbf{T}\|$; $\mathcal{R} = \mathbf{R}_w$, $\mathcal{T} = \mathbf{T}_w$, $\lambda = \|\mathbf{T}\|$ where $\mathbf{T} = \mathbf{t}'_w - \mathbf{R}'_w\mathbf{R}_w^T\mathbf{T}_w$	
AFFINE	**affine transformations** *preserve* **parallelism, center of mass** $\mathcal{A}(\Pi_\infty) = \Pi_\infty$	
$\mathcal{GA}_3$	$\mathcal{A} = \begin{bmatrix} \mathcal{M} & \mathcal{V} \\ 0_3^T & \mu \end{bmatrix}$ $\mathcal{M}$: non-singular 3×3 matrix $\mathcal{V}$: 3D vector μ: non-null scalar $\}$ $\mathcal{A}$ defined up to a global scale factor	12
invariant description	$\mathbf{H}_\infty$: infinity homography from image 1 to image 2	8
	$\mathbf{e}'_N$: normalized epipole in image 2	2
canonic decomposition	$\begin{cases} \tilde{\mathbf{P}} = [\mathbf{I}_3, 0]\,\mathcal{A} \\ \tilde{\mathbf{P}}' = [\mathbf{H}_\infty, \mathbf{e}'_N]\,\mathcal{A} \end{cases}$ $\mathbf{H}_\infty = \mathbf{P}'\mathbf{P}^{-1}$, $\mathbf{e}'_N = \mathbf{e}/\|\mathbf{e}'\|$; $\mathcal{M} = \mathbf{P}$, $\mathcal{V} = \mathbf{p}$, $\mu = \|\mathbf{e}'\|$ where $\mathbf{H}_\infty \sim \mathbf{A}'\mathbf{R}\mathbf{A}^{-1}$, $\mathbf{e}' = \mathbf{p}' - \mathbf{H}_\infty\mathbf{p} \sim \mathbf{A}'\mathbf{T}$	
PROJECTIVE	**homographies** *preserve* **collinearity, cross-ratio** $\mathcal{H}(\mathcal{P}^3) = \mathcal{P}^3$	
$\mathcal{GL}_4$	$\mathcal{H} = \begin{bmatrix} \mathcal{M} & \mathcal{V} \\ \mathcal{L}_N^T & \nu_N \end{bmatrix}$ $\mathcal{M}$: 3×3 matrix $\mathcal{V}$, $\mathcal{L}_N$: 3D vectors ν_N: scalar $\}$ $\mathcal{H}$ non-singular defined up to a global scale factor	15
invariant description	$\mathbf{S}$: epipolar projection from image 1 to image 2	5
	$\mathbf{e}'_N$: normalized epipole in image 2	2
canonic decomposition	$\begin{cases} \tilde{\mathbf{P}} = [\mathbf{I}_3, 0]\,\mathcal{H} \\ \tilde{\mathbf{P}}' = [\mathbf{S}, \mathbf{e}'_N]\,\mathcal{H} \end{cases}$ $\mathbf{S} = -[\mathbf{e}'_N]_\times{}^2\mathbf{H}_\infty$, $\mathbf{e}'_N = \mathbf{e}/\|\mathbf{e}'\|$; $\mathcal{M} = \mathbf{P}$, $\mathcal{V} = \mathbf{p}$, $\mathcal{L}_N = \mathbf{P}'^T\mathbf{e}'_N$, $\nu_N = \mathbf{p}'^T\mathbf{e}'_N$ where $\mathbf{H}_\infty = \mathbf{P}'\mathbf{P}^{-1}$, $\mathbf{e}' = \mathbf{p}' - \mathbf{H}_\infty\mathbf{p}$	

Table 2. The geometries of three views: canonic representation

EUCLIDEAN (calibration)	$\mathcal{D} \in \mathcal{SE}_3$		**displacement**	6
	$\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$		intrinsic parameters	5+5+5
invariant	$\mathbf{R}_{12}, \mathbf{R}_{23}$	$\mathbf{R}_{12}, \mathbf{R}_{13}$	rotations	3+3
descriptions	$\mathbf{T}_{12}, \mathbf{T}_{23}$	$\mathbf{T}_{12}, \mathbf{T}_{13}$	translations	3+3
canonic decomposition	$\begin{cases} \tilde{\mathbf{P}}_1 = \mathbf{A}_1[\mathbf{I}_3, 0]\mathcal{D} \\ \tilde{\mathbf{P}}_2 = \mathbf{A}_2[\mathbf{R}_{12}, \mathbf{T}_{12}]\mathcal{D} \\ \tilde{\mathbf{P}}_3 = \begin{matrix} \mathbf{A}_3[\mathbf{R}_{13} \quad , \mathbf{T}_{13}]\mathcal{D} \\ \mathbf{A}_3[\mathbf{R}_{23}\mathbf{R}_{12}, \mathbf{R}_{23}\mathbf{T}_{12} + \mathbf{T}_{23}]\mathcal{D} \end{matrix} \end{cases}$			
EUCLIDEAN (motion)	$\mathcal{S} \in \lambda\mathcal{SE}_3$		**similarity**	7
	$\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$		intrinsic parameters	5+5+5
invariant	$\mathbf{R}_{12}, \mathbf{R}_{23}$	$\mathbf{R}_{12}, \mathbf{R}_{13}$	rotations	3+3
descriptions	$\mathbf{t}_{12}, \mathbf{t}_{23}$	$\mathbf{t}_{12}, \mathbf{t}_{13}$	directions of translations	2+2
	α_1	α_2	ratios of translation norms	1
canonic decomposition	$\begin{cases} \tilde{\mathbf{P}}_1 = \mathbf{A}_1[\mathbf{I}_3, 0]\mathcal{S} \\ \tilde{\mathbf{P}}_2 = \mathbf{A}_2[\mathbf{R}_{12}, \mathbf{t}_{12}]\mathcal{S} \\ \tilde{\mathbf{P}}_3 = \begin{matrix} \mathbf{A}_3[\mathbf{R}_{13} \quad , \alpha_1\mathbf{t}_{13}]\mathcal{S} \\ \mathbf{A}_3[\mathbf{R}_{23}\mathbf{R}_{12}, \mathbf{R}_{23}\mathbf{t}_{12} + \alpha_2\mathbf{t}_{23}]\mathcal{S} \end{matrix} \end{cases}$		$\alpha_1 = \lVert\mathbf{T}_{13}\rVert / \lVert\mathbf{T}_{12}\rVert$ $\alpha_2 = \lVert\mathbf{T}_{23}\rVert / \lVert\mathbf{T}_{12}\rVert$	
AFFINE	$\mathcal{A} \in \mathcal{GA}_3$		**affine transformation**	12
	$\mathbf{H}_{\infty 12}, \mathbf{H}_{\infty 23}$	$\mathbf{H}_{\infty 12}, \mathbf{H}_{\infty 13}$	infinity homographies	8+8
invariant	$\mathbf{e}_{N21}, \mathbf{e}_{N32}$	$\mathbf{e}_{N21}, \mathbf{e}_{N31}$	normalized epipoles	2+2
descriptions	β_1	β_2	ratios of epipole norms	1
canonic decomposition	$\begin{cases} \tilde{\mathbf{P}}_1 = [\mathbf{I}_3, 0]\mathcal{A} \\ \tilde{\mathbf{P}}_2 = [\mathbf{H}_{\infty 12}, \mathbf{e}_{N21}]\mathcal{A} \\ \tilde{\mathbf{P}}_3 = \begin{matrix} [\mathbf{H}_{\infty 13} \quad , \beta_1\mathbf{e}_{31N}]\mathcal{A} \\ [\mathbf{H}_{\infty 23}\mathbf{H}_{\infty 12}, \mathbf{H}_{\infty 23}\mathbf{e}_{N21} + \beta_2\mathbf{e}_{N32}]\mathcal{A} \end{matrix} \end{cases}$		$\beta_1 = \lVert\mathbf{e}_{31}\rVert / \lVert\mathbf{e}_{21}\rVert = \alpha_1\lVert\mathbf{A}_3\mathbf{t}_{13}\rVert / \lVert\mathbf{A}_2\mathbf{t}_{12}\rVert$ $\beta_2 = \lVert\mathbf{e}_{32}\rVert / \lVert\mathbf{e}_{21}\rVert = \alpha_2\lVert\mathbf{A}_3\mathbf{t}_{23}\rVert / \lVert\mathbf{A}_2\mathbf{t}_{12}\rVert$	
PROJECTIVE	$\mathcal{H} \in \mathcal{GL}_4$		**homography**	15
	$\mathbf{S}_{12}, \mathbf{S}_{23}$	$\mathbf{S}_{12}, \mathbf{S}_{13}$	epipolar projections	5+5
invariant	$\mathbf{e}_{N21}, \mathbf{e}_{N32}$	$\mathbf{e}_{N21}, \mathbf{e}_{N31}$	normalized epipoles	2+2
descriptions	$\mathbf{q}_{N1}$	$\mathbf{q}_{N2}$	differences of r_∞-vectors	3
	β_1	β_2	ratios of epipole norms	1
canonic decomposition	$\begin{cases} \tilde{\mathbf{P}}_1 = [\mathbf{I}_3, 0]\mathcal{H} \\ \tilde{\mathbf{P}}_2 = [\mathbf{S}_{12}, \mathbf{e}_{N21}]\mathcal{H} \\ \qquad [\mathbf{S}_{13} - \beta_1\mathbf{e}_{31N}\mathbf{q}_{N1}^T, \beta_1\mathbf{e}_{31N}]\mathcal{H} \\ \tilde{\mathbf{P}}_3 = [\mathbf{S}_{23}\mathbf{S}_{12} + \beta_2\mathbf{e}_{N32}(\mathbf{q}_{N2}^T\mathbf{S}_{12} + \gamma_1\mathbf{e}_{N12}^T)\ , \\ \qquad \mathbf{S}_{23}\mathbf{e}_{N21} + \beta_2\mathbf{e}_{N32}(\mathbf{q}_{N2}^T\mathbf{e}_{N21})]\mathcal{H} \end{cases}$		$\mathbf{q}_{N1} = \mathbf{H}_{\infty 12}^T\mathbf{e}_{N21} - \frac{1}{\beta_1}\mathbf{H}_{\infty 13}^T\mathbf{e}_{N31}$ $\mathbf{q}_{N2} = \frac{1}{\beta_2}\mathbf{H}_{\infty 23}^T\mathbf{e}_{N32} - \gamma_1\mathbf{H}_{\infty 21}^T\mathbf{e}_{N12}$ $\beta_1 = \lVert\mathbf{e}_{31}\rVert / \lVert\mathbf{e}_{21}\rVert$ $\beta_2 = \lVert\mathbf{e}_{32}\rVert / \lVert\mathbf{e}_{21}\rVert$ *with* $\gamma_1 = \lVert\mathbf{e}_{21}\rVert / \lVert\mathbf{e}_{12}\rVert$	