COMPUTER VISION AND IMAGE UNDERSTANDING
Vol. 64, No. 2, September, pp. 193–229, 1996
ARTICLE NO. 0055

Canonical Representations for the Geometries of Multiple Projective Views

Q.-T. LUONG

*Computer Science Division, University of California, Berkeley, California 94720**

AND

T. VIÉVILLE

I.N.R.I.A., 2004, route de Lucioles, 06902 Sophia-Antipolis, France

Received October 21, 1993; accepted November 18, 1994

This work is in the context of motion and stereo analysis. It presents a new unified representation which will be useful when dealing with multiple views in the case of uncalibrated cameras. Several levels of information might be considered, depending on the availability of information. Among other things, an algebraic description of the epipolar geometry of *N* views is introduced, as well as a framework for camera self-calibration, calibration updating, and structure from motion in an image sequence taken by a camera which is zooming and moving at the same time. We show how a special decomposition of a set of two or three general projection matrices, called *canonical*, 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. The theory is illustrated by tutorial examples with real images.

1. INTRODUCTION

Nonmetric Descriptions for 3D Vision. This paper is about a unified framework to account for the Euclidean, affine, and projective geometries of two, three, or more cameras. Three-dimensional problems involving several views have traditionally been studied under the assumption that the cameras are calibrated. Quite recently, Forsyth and his co-workers [19], in the context of object recognition, and Koenderink and Van Doorn [29] in the context of structure from motion, have shown the value of nonmetric representations, which have then become the subject of an active research area. There are two main reasons why such a line of research is attractive. First, projective geometry allows us to understand and express the geometry of the problem in a much simpler way. Second, it gives us a framework for 3D description even in the case when the cameras are not calibrated. Camera calibration is an off-line, model-based process which is impractical and unstable. Moreover, for applications such as image data base query processing, virtual reality model building, and video processing, the calibration data in most of the cases are just not available. The reason why camera calibration has been (and is still by many) considered to be a prerequisite for any visual task involving 3D information is that the 3D descriptions which are classically used are metric, and therefore, it is necessary to map the image measurement into the directions in 3D space of the lines of sight to obtain meaningful Euclidean measurements. Now, if we consider *projective* relations instead of Euclidean relations, we no longer need such requirements and thus can work directly from image measurements, even with uncalibrated images. It is clear, however, that the descriptions which could be recovered this way will be much more unconstrained. At a given level of description, there are a number of geometric properties such as collinearity, parallelism, perpendicularity, distance, etc. . . . , which may be recovered. Depending on the quantity of information which is available (about the camera calibration or motion, the scene), and on the quantity of information which is eventually needed, one level of description or another could be more appropriate. This leads to one of the key ideas of this

* Current address: SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025. E-mail: luong@ai.sri.com.

paper: stratification, which is the consideration of multiple *geometries* and their relationships. In contrast with the approaches described in [46] ours do not study configurations of 3D points, but the relations within the system of cameras instead, hence the term *views*. The knowledge of such relations is a prerequisite for the inference of properties of the 3D space, much the same way as camera calibration was considered to be a prerequisite for 3D vision.

Motivations for a New Representation. Representations for such relations in a purely projective framework can be found in papers such as [14, 21, 23, 25, 26, 44, 49, 57, 58]. However, the constraints provided by projective geometry have sometimes proven quite weak for some applications. Affine geometry has been found to provide an interesting framework (see, for instance, [28], where no less than six papers about affine structure can be found, and, for example, [4, 9, 42, 47, 52, 56, 60]), 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 (see the recent papers [22, 47] though), which is a consequence of the fact that as the mathematical language used was quite different, comparisons were difficult. 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 *canonical decomposition*.

On the theoretical side, the paper solves the problem of finding an algebraic/geometric representation of the relations between a system of cameras at different levels of description. It puts together a number of results. Some of them, while implied by the existing literature, are not yet interrelated in an organized manner. Some of them are new, for instance, the constructive count of the number of degrees of freedom in the projective and affine representation of the geometries of N cameras. There are two reasons why the new representation is important:

- First, it allows one to consider the world as a superposition of projective, affine, and Euclidean structures, and to deal with these structures simultaneously, whereas the existing formalisms allowed one only to consider one of these structures at a time, although the idea to express Euclidean constraints in a projective framework has already been exploited [5].
- Second, the new representation allows us to consider global representations within an image sequence, whereas only representations which involve pairs of images were available in the general case (In a paper such as [44], the representation is not worked out explicitly.) with the consequence that only binocular algorithms could be used.

On the practical side, the new representation is good for dealing with all the problems which involve multiple views and uncalibrated cameras, the two main benefits being those previously emphasized. Some applications include projective and affine reconstructions from a *sequence* of uncalibrated views, propagation of affine and Euclidean information in a sequence *where the camera parameters may vary*, self-calibration using affine information. The usefulness of the representation has been illustrated by [67] where it was used to design a theory and algorithms for recovery of structure and motion from points and lines in the uncalibrated case. Parts of the framework have also already been used by several other researchers.

How to Read This Paper: Organization and Shortcuts. The paper provides the reader with a comprehensive and self-contained description of the new conceptual framework. The organization of the paper is as follows. Section 2 contains background material. It presents classical transformation groups and applies them to the case of a single camera, to obtain projective, affine, and Euclidean interpretations. It might be skipped by a reader already familiar with these concepts. Section 3 deals with a set of two cameras and introduces new elements of representation in the uncalibrated case (some background material on fundamental matrices and homographies is also given). Section 4 gives the canonical decomposition for two views. It is extended to the case of three and more views in Section 5. Section 6 discusses the relations between levels of representation. Section 7 discusses the problem of their recovery from image measurements. Section 8 gives some examples. We conclude in Section 9 by outlining the novel features and the usefulness of the new representation.

The following entries should allow the reader to get a quick grasp of the main results of the paper. The stratification of the camera model is summarized in the table at the beginning of Section 2.2. The definition and basic properties of the main components of the representation can be found at the beginning of each subsection of Section 3, whereas the expression of these components as a function of the projection matrices is in Table 2. The explanation of the main symbols used in the paper can be found in Table 1. The idea behind the canonical decomposition is in Section 4.1. The main result of Section 4, the canonical decomposition for two views, is summarized in Table 2. The extension of the canonical decomposition to three views is summarized in Table 3. Two important stratified geometric descriptions are given by Eq. (50) and the table at the beginning of Section 6.2. The examples from real

TABLE 1
Index of the Main Symbols Used in the Paper

$\mathbf{A}$	Intrinsic parameters matrix: transforms 2D normalized coordinates to retinal coordinates	(6)
$\mathscr{A}$	Affine transformation of $\mathscr{P}^3$	(1)
$\mathbf{B}$	Image of the absolute conic: matrix of a plane conic depending only on $\mathbf{A}$	Section 2.2
$\mathbf{C}$	Optical center: center of projection of the camera	(4)
d	Distance of a plane to the origin	Section 6.2
$\mathscr{D}$	Rigid displacement of $\mathscr{P}^3$	Section 2.1
$\mathbf{e}'$	Epipole in second view: projection into $\mathscr{R}'$ of the optical center of the first camera	Section 3.1; (12)
$\mathbf{e}'_N$	Normalized epipole	(28)
$\mathbf{E}$	Essential matrix: encodes relative displacement between two cameras	(53)
$\mathbf{F}$	Fundamental matrix: maps point to epipolar line in retinal coordinates	(8; 13)
$\mathbf{G}$	Projective dual to the fundamental matrix	(26)
$\mathbf{H}_\Pi$	Homography matrix: Correspondence between image in $\mathscr{R}$ and in $\mathscr{R}'$ of points in Π	(18)
$\mathbf{H}_\infty$	Infinity homography matrix: correspondence between image in $\mathscr{R}$ and $\mathscr{R}'$ of points in Π_∞	(15)
$\mathscr{H}$	Homography of $\mathscr{P}^3$	Section 2.1
$\mathbf{I}$	Identity matrix	
$\mathbf{K}$	Kruppa matrix: matrix of the dual to the image of Ω	(7)
$\mathbf{l}$	A line of $\mathscr{R}$	
$\mathbf{m}$	A 2D point of $\mathscr{R}$	
$\mathbf{M}$	A 3D point of $\mathscr{P}^3$	
$\mathbf{n}$	Unit normal to a plane	Section 6.2
$\mathbf{P}$	Projection matrix: Transforms 3D world coordinates to retinal coordinates	(2)
$\mathbf{p}$	Right 3×1 column vector of $\tilde{\mathbf{P}}$: projection of the origin of the 3D world coordinate system	(3)
$\mathbf{P}$	Left 3×3 submatrix of $\tilde{\mathbf{P}}$: Homography between Π_∞ and $\mathscr{R}$	(3)
$\mathscr{P}^2$	Projective plane	Section 2.1
$\mathscr{P}^3$	Projective space	Section 2.1
$\mathbf{q}$	Difference of $\mathbf{r}_\infty$-vectors	(36)
$\mathbf{q}_N$	Difference of $\mathbf{r}_{\infty N}$-vectors	(41)
$\mathbf{Q}$	Uncalibrated rotation	Section 4.2
$\mathbf{r}_\Pi$	Relative representation vector of Π: projective characterization of Π	(25)
$\mathbf{r}_\infty$	Relative representation vector of Π_∞: projective characterization of Π_∞	(17)
$\mathbf{r}_{\infty N}$	Normalized r_∞-vector	(29)
$\mathbf{R}$	Rotational component of motion (between two cameras)	Section 4.2; (51)
$\mathbf{R}_w$	Rotational component of pose (between 3D world coordinates and 3D camera coordinates)	(6)
$\mathscr{R}$	Retinal plane	Section 2.2
$\mathbf{s}$	Uncalibrated translation	Section 4.2
$\mathbf{S}$	Epipolar projection: maps to a point of the line $\langle \mathbf{e}' \rangle$	(11, 14)
$\mathbf{t}$	Unit direction of translational component of motion	Section 4.2
$\mathbf{T}$	Translational component of motion (between two cameras)	Section 4.2; (51)
$\mathbf{T}_w$	Translational component of pose (between 3D world coordinates and 3D camera coordinates)	(6)
$\mathscr{X}_i$	Projective coordinates in $\mathscr{P}^3$	Section 2.1
α	Ratio of translation norms	Section 5.1
β	Ratio of epipole norms	(37)
γ	Ratio of norm of coupled epipoles $\mathbf{e}$ and $\mathbf{e}'$	(40)
ν_Π	Affine characterization of Π	(55)
Π	A plane	
Π_∞	The plane at infinity	Section 2.1
Ω	The absolute conic	Section 2.1

images, presented in Section 8, can be used as a guide before the reader has digested the main body of the paper, and also give an idea of the potential applications of the ideas presented in the paper.

2. PROJECTIVE, AFFINE, AND EUCLIDEAN INTERPRETATIONS OF THE CAMERA MODEL

Projective geometry [55] is the reference framework which will be used throughout this article, because it deals with elegance with the general case of the type of projection that a camera performs. Although the natural geometry which we use is Euclidean, it is much more simpler to consider the Euclidean and affine geometries as special cases of the projective geometry rather than the reverse. Depending on the visual task, a Euclidean description is not always necessary. Using a weaker description is more general, is easier to obtain, and can capture more precisely the properties which are relevant to a given task. This leads to the first key idea of this paper: *stratification*, which is the analysis of entities at different geometric levels which correspond to given geometric properties. The three levels of description which are studied in the paper are projective, affine, and Euclidean. The main goal of this section is to provide a basis for such a stratified description.

The necessary background on transformation groups is first given in Section 2.1. These groups are important because each of them leaves invariant certain properties of 3D space that we might be interested in. The remarkable thing is that they form a hierarchy, in the sense that they can be considered as subgroups of each other. Table 2 which summarizes the canonical decomposition for two views contains also the main properties of the transformation groups which are discussed in this section. We then apply in Section 2.2 the hierarchy to a single camera, resulting in the interpretations which are summarized in the table presented at the beginning of Section 2.2.

2.1. *The Hierarchy of Transformation Groups*

The Projective Level: Homographies. The *projective space* of dimension n, $\mathcal{P}^n$ is the quotient space of $\mathbb{R}^{\mathbf{n+1}} - \{[\mathbf{0}, \ldots, \mathbf{0}]\}$ by the equivalence relation

$$[x_1, \ldots, x_{n+1}] \sim [x'_1, \ldots, x'_{n+1}] \Leftrightarrow \\ \exists \lambda \neq 0, [x_1, \ldots, x_{n+1}] = \lambda [x'_1, \ldots, x'_{n+1}].$$

This means that proportional $(n+1)$-uples of coordinates $\mathbf{x} = [x_1, \ldots, x_{n+1}]$ and $\mathbf{x}' = [x'_1, \ldots, x'_{n+1}]$ represent the same point in projective space. The object space will be considered as $\mathcal{P}^3$. The image space (also called retinal plane and noted $\mathcal{R}$), as well as other planes, will be considered as $\mathcal{P}^2$. The notation $[\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3, \mathcal{X}_4]$ will be used for projective coordinates of $\mathcal{P}^3$.

An *homography* is any transformation $\mathcal{H}$ of $\mathcal{P}^n$ which is linear in projective coordinates (hence the terminology *linear projective*), and which is invertible, (thus $\mathcal{H}(\mathcal{P}^n) = \mathcal{P}^n$). It can be described by an $(n+1) \times (n+1)$ nonsingular matrix $\mathcal{H}$, such that the image of $\mathbf{x}$ is $\mathbf{x}'$:

$$\mathbf{x}' = \mathcal{H}\mathbf{x}.$$

The basic consequence of the linearity is that homographies map hyperplanes to hyperplanes. More generally, they map any projective subspace to a subspace of the same dimension, a property which is called *conservation of incidence*. Homographies form a group $\mathcal{GL}_n$ which is called the *projective group*. A basic measurement which is conserved by homographies are cross-ratios (*ratios of ratios of distances*).

The Affine Level: Affine and Unimodular Transformations. Any point $\mathbf{h}$ of $\mathcal{P}^n$ defines a *hyperplane*, which is the set of points $\mathbf{x}$ of $\mathcal{P}^n$ whose coordinates satisfy

$$\sum_{1 \leq i \leq n+1} h_i x_i = \mathbf{h}^T \mathbf{x} = 0.$$

A hyperplane of $\mathcal{P}^n$ can be considered as a projective subspace of dimension $\mathcal{P}^{n-1}$. Hyperplanes of $\mathcal{P}^3$ are planes, hyperplanes of $\mathcal{P}^2$ are lines.

Affine structure of $\mathcal{P}^3$ is characterized by the *plane at infinity* Π_∞, which is represented by the vector $[0, 0, 0, 1]$, and thus defined by $\mathcal{X}_4 = 0$. The projective space $\mathcal{P}^3$ can be described as the union of the usual affine space (points $[\mathbf{X}, \mathcal{X}_4]$ with $\mathcal{X}_4 \neq 0$) and the plane at infinity Π_∞ (points $[\mathbf{X}, 0]$). For points in the affine space, the usual space coordinates X, Y, and Z are related to the projective world coordinates by

$$X = \frac{\mathcal{X}_1}{\mathcal{X}_4} \quad Y = \frac{\mathcal{X}_2}{\mathcal{X}_4} \quad Z = \frac{\mathcal{X}_3}{\mathcal{X}_4}.$$

If we consider points of Π_∞ as the limit of $[X, Y, Z, \lambda]$ (or equivalently of $[X/\lambda, Y/\lambda, Z/\lambda, 1]$) when $\lambda \to 0$, then we see that they are the limit of a point of $\mathbb{R}^3$ going to infinity in the direction $[X, Y, Z]$, hence the appellation *point at infinity*. For this reason, the *direction* $[\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3]$ of any plane of the form $[\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3, \mathcal{X}_4]$ is defined by its intersection with the plane at infinity Π_∞.

The *affine transformations* of $\mathcal{P}^3$ are the subgroup $\mathcal{GA}_3$ of $\mathcal{GL}_3$ defined by the transformations $\mathcal{A}$ which conserve the plane at infinity, which means that $\mathcal{A}(\Pi_\infty) = \Pi_\infty$. Any two subspaces of $\mathcal{P}^3$ which are not contained in Π_∞ are *parallel* if their intersection is in Π_∞. This implies that affine transformations are those which preserve parallelism. It is easy to see that a transformation $\mathcal{A}$ conserves Π_∞ if and only if the last row of the matrix of $\mathcal{A}$ is of the form $[0, \ldots, 0, \mu]$, with $\mu \neq 0$. Since this matrix is defined only up to a scale factor, we can take $\mu = 1$, then the transformation $\mathcal{A}$ is fully described by its first 3×3 submatrix $\mathcal{M}$ and the three first coordinates of the last column vector $\mathcal{V}$

$$\mathcal{A} = \begin{bmatrix} \mathcal{M} & \mathcal{V} \\ \mathbf{0}_3^T & 1 \end{bmatrix}, \tag{1}$$

which yields the classical description of a transformation of the affine space $\mathbb{R}^3$: $\mathbf{x}' = \mathcal{M}\mathbf{x} + \mathcal{V}$.

The *ratio of distances of three collinear points* is invariant by an affine transformation. A consequence of this property is the fact that affine transformations leave invariant center of mass and convex hulls. Another quantity which is left invariant is the ratio of volumes defined by four points in 3D space. Moreover, if we add to (1) the constraint $\det(\mathcal{M}) = \pm 1$, then the volumes themselves are left invariant. In fact, this constraint alone leads to a subgroup of $\mathcal{GA}_3$, which is called the *affine unimodular group*.

The Euclidean Level: Similarities and Displacements. Any symmetric $(n+1) \times (n+1)$ matrix $\mathbf{Q}$ defines a *hyperquadric*, which is the set of points $\mathbf{x}$ of $\mathcal{P}^n$ whose coordinates satisfy

$$\sum_{1 \le i,j \le n+1} Q_{ij} x_i x_j = \mathbf{x}^T \mathbf{Q} \mathbf{x} = 0.$$

In $\mathcal{P}^3$, the hyperquadrics are quadric surfaces, in $\mathcal{P}^2$, they are conics, and in $\mathcal{P}^1$ they reduce to two points.

The Euclidean structure of $\mathcal{P}^3$ is characterized[1] by the *absolute conic* Ω which lies in the plane at infinity Π_∞ ($\mathcal{X}_4 = 0$) and has matrix identity, and thus equation

$$\mathcal{X}_1^2 + \mathcal{X}_2^2 + \mathcal{X}_3^2 = 0.$$

The *Euclidean* subgroup, $\mathcal{GE}_3$ of $\mathcal{GL}_3$, is defined by the transformations $\mathcal{S}$ which conserve the absolute conic, which means that $\mathcal{S}(\Omega) = \Omega$. Note that this implies that $\mathcal{S}(\Pi_\infty) = \Pi_\infty$, and therefore the Euclidean transformations form a subgroup of the affine group. Much the same way than Π_∞ can be used to define directions of planes, Ω can be used to define angles between two planes h_1 and h_2 by the Laguerre formula $\alpha = (1/2i)\log(\{h_1, h_2, h_i, h_j\})$ where h_i and h_j are the two planes of the pencil defined by h_1 and h_2, which are tangent to the absolute conic. This implies that Euclidean transformations are those which preserve angles. It is known that Euclidean transformations are the affine transformations (1) for which we have the additional constraint: $\mathcal{M}\mathcal{M}^T = s\mathbf{I}_n$, which means that the first 3×3 submatrix is proportional to an orthogonal matrix. The resulting transformations are the *similarities*. They are the product of a scale factor by a rigid displacement.

Euclidean transformations perserve the *relative distance*, which is the ratio of distance of any three points. Moreover, the additional constraint $\det(\mathcal{M}) = \pm 1$ is equivalent to the fact that the absolute distances of two points are preserved. The resulting group is the group of *isometries*, which is the intersection of the affine unimodular groups with the similarity group.

2.2. *Stratification of the Camera Model*

The projective, affine, and Euclidean properties of a single camera, are presented. The relevant decompositions are in the table below:

[1] The idea first appeared in the work of Cayley, and has been introduced in the computer vision literature by [17].

Level	Decomposition	Interpretation
Projective	$\tilde{\mathbf{P}}$	$\tilde{\mathbf{P}}$: projection matrix mapping object space $\mathcal{P}^3$ to retinal plane $\mathcal{R}$
Affine	$[\mathbf{P}, \mathbf{p}]$	$\mathbf{P}$: homography between the plane at infinity Π_∞ and $\mathcal{R}$ $\mathbf{p}$: projection of the origin of the world coordinate system
Euclidean	$[\mathbf{A}\mathbf{R}_w, \mathbf{A}\mathbf{T}_w]$	$\mathbf{A}$: change of coordinate system in $\mathcal{R}$ (five intrinsic parameters) $(\mathbf{R}_w, \mathbf{T}_w)$: pose displacement between world and camera coordinate systems

Projective Level: The General Perspective Projection Matrix. The camera model which we consider is the pinhole model. In this model, the camera performs a perspective projection of an object point M onto a pixel m in the retinal plane $\mathcal{R}$ through the optical center C. The main property of this camera model is thus that *the relationship between the world coordinates and the pixel coordinates is linear projective.* This property is independent of the choice of the coordinate systems in the retinal plane or in the three-dimensional space. The model takes into account the pose of the camera and a large class of transformations of the image plane. In particular, nonsquare and even nonorthogonal pixel grids, nonrigid cameras, and the process of taking pictures of pictures are adequately modeled (see also [57] for a discussion of this point). By contrast, the model does not take into account optical distortions due to lens or image plane imperfections. Therefore, by the terms *retinal* or *pixel* coordinates, we refer to measurements in images which are geometrically undistorted, but which might have undergone a projective transformation of the retinal plane $\mathcal{R}$. The consequence is that the relationship between 2D pixel coordinates and any 3D world coordinates can be described by a 3×4 matrix $\tilde{\mathbf{P}}$, called projection matrix, which maps points from $\mathcal{P}^3$ to $\mathcal{P}^2$:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \tilde{\mathbf{P}} \begin{bmatrix} \mathcal{X}_1 \\ \mathcal{X}_2 \\ \mathcal{X}_3 \\ \mathcal{X}_4 \end{bmatrix}. \tag{2}$$

Although this model has been used since the earliest days [54], its main characteristic is emphasized in the recent appellation *projective camera* [46]. The important point is that since the world, as well as the retina are both modeled as projective spaces, the use of this model involves only properties from projective geometry.

The goal of this paper is to exploit Eq. (2) *to its fullest extent 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.

The projection matrix $\tilde{\mathbf{P}}$ must be of rank 3; otherwise, its image would be a projective line instead of a projective plane. Since it has four columns, its kernel contains thus one unique projective point $\mathbf{C}$. This point, for which the projection is not defined, is the optical center. In the most general model, it can be either in the affine space (that is at finite distance) or in the plane at infinity.

Affine Interpretation: Optical Rays. Let us now decompose the projection matrix $\tilde{\mathbf{P}}$ as the concatenation of a 3×3 submatrix $\mathbf{P}$ and a 3×1 vector $\mathbf{p}$. Note that the decomposition has been already used by numerous authors (for instance, [4, 14, 26]) who did not make explicit its meaning. The optical center is also decomposed by separating its last component:

$$\tilde{\mathbf{P}} = [\mathbf{P}, \mathbf{p}], \tilde{\mathbf{C}} = \begin{bmatrix} \mathbf{C} \\ c \end{bmatrix}. \tag{3}$$

The equation determining the optical center is thus $\mathbf{PC} = -c\mathbf{p}$, and we can see, remarking that the solution must be unique, that the optical center $\tilde{\mathbf{C}}$ lies in the plane at infinity if and only if $\det(\mathbf{P}) = 0$. We can conclude that if $\det(P) \neq 0$, then the optical center is finite, and given by

$$\tilde{\mathbf{C}} = \begin{bmatrix} -\mathbf{P}^{-1}\mathbf{p} \\ 1 \end{bmatrix}. \tag{4}$$

It can be noted that any projection matrix arising from a physical system must satisfy $\det(\mathbf{P}) \neq 0$, since the optical center must lie in the affine space. We will make use of this assumption all the way through the paper. The alternative class of models can be considered as approximations to the pinhole model in some particular viewing situations. These include orthographic, weak perspective, and the affine camera [46], the most general form of these models, which was considered implicitly in [29] and studied extensively in [56]. The interest of such models is that they lend to *affine* descriptions of the 3D world, rather than projective descriptions which are obtained in the general case. Although from a purely projective standpoint the two cases are equivalent and can be switched by a change of coordinate system (used, for instance, in [18]), in practice the constraint $\det(\mathbf{P}) \neq 0$ is very important, as it allows us to obtain a standard affine interpretation of the projection matrix.

The projection giving 2D pixel coordinates from 3D world coordinates can be described as a composition of an affine transformation and the projection expressed in the final camera retinal system:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\tilde{\mathbf{P}}_c} \begin{bmatrix} \mathbf{P} & \mathbf{p} \\ \mathbf{0}_3^T & 1 \end{bmatrix} \begin{bmatrix} \mathscr{X}_1 \\ \mathscr{X}_2 \\ \mathscr{X}_3 \\ \mathscr{X}_4 \end{bmatrix}. \tag{5}$$

At this affine level of description, we can introduce *directions* of optical rays. Since the projection of each point at infinity $[\mathbf{d}, 0]$ is the *vanishing point* $\mathbf{v} = \mathbf{Pd}$, $\mathbf{P}$ can be considered as the homography between the plane at infinity Π_∞ and the retinal plane $\mathscr{R}$. Note that parallel lines have the same direction, hence the same point at infinity; thus, their projection is a set of lines of $\mathscr{R}$ which contains the image of this vanishing point.

The optical ray corresponding to the pixel $\mathbf{m}$ has thus the direction $\mathbf{P}^{-1}\mathbf{m}$. The vector $\mathbf{p}$ is just the projection of the origin of the world coordinate system.

Euclidean Interpretation: Intrinsic and Extrinsic Parameters. We apply to the matrix $\mathbf{P}$ the QR theorem, which states that a nonsingular matrix can be factored uniquely as the product of a triangular matrix and an orthogonal matrix (see, for instance, [20]). We then factor out the last coefficient of the triangular matrix, λ_w. It is seen that the projection matrix can thus be decomposed uniquely as

$$\tilde{\mathbf{P}} = \lambda_w \underbrace{\begin{bmatrix} \alpha_u & \gamma & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\tilde{\mathbf{P}}_c} \underbrace{\begin{bmatrix} \mathbf{R}_w & \mathbf{T}_w \\ \mathbf{0}_3^T & 1 \end{bmatrix}}_{\mathscr{D}_w}, \tag{6}$$

where $\mathbf{A}$ is a 3×3 matrix describing the change of retinal coordinate system, whose five entries are called *intrinsic parameters*, and $\mathscr{D}_w$ is a 4×4 displacement matrix describing the change of world coordinate system (the pose of the camera) called *extrinsic parameters.* Note that to obtain a unique decomposition (6), we must restrict the form of $\mathbf{A}$. Different formulations are possible, as long as the uniqueness of decomposition is preserved. Other models with direct physical interpretation can be found, for example, in [15, 62]. The one adopted here for simplicity appeared in [62] and is also discussed in [68], and used by [24]. It is worth noting that $\mathbf{A}$, being upper triangular, defines an affine transformation of the retinal plane, rather than a general projective transformation of this plane. It can be seen that 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.

A camera is said to be *calibrated* when $\mathbf{A}$ is known. The change of coordinates which is represented by $\mathbf{A}^{-1}$ transforms the pixel (or retinal) coordinates into *normalized* coordinates, in which the 2D affine coordinates of a pixel $\mathbf{m} = [x, y]$ are directly related to the direction in 3D space of the optical ray $\langle \mathbf{C}, \mathbf{m} \rangle = [x, y, 1]$.

There is an interesting and important relationship between the camera's intrinsic parameters and the absolute conic, already used in [17, 41]. Since the absolute conic is invariant under rigid displacements, its image ω from the camera, which is also a conic with only complex points, does not depend on the pose of the camera. Therefore, its equation in the retinal coordinate system does not depend on the extrinsic parameters and depends only on the intrinsic parameters. Its matrix is $\mathbf{B} = \mathbf{A}^{-1}\mathbf{A}^{-1T}$, whereas its dual conic (the set of its tangents) has the adjoint matrix of $\mathbf{B}$ [16]:

$$\mathbf{K} = \mathbf{B}^* = \det(\mathbf{B})\mathbf{B}^{-1T} \sim \mathbf{A}\mathbf{A}^T. \tag{7}$$

The matrix $\mathbf{K}$ is called the Kruppa matrix. It is symmetric and defined up to a scale factor, and therefore it depends on five independent parameters. It is a better description of camera calibration than the intrinsic parameters, since it does not depend on the choice of a particular model. There is a one-to-one correspondence between these matrices of Kruppa coefficients and the intrinsic parameters [16, 67]. It can be shown [15] that ω determines the angle between optical rays, which is consistent with the fact that the similarities conserve angles.

3. REPRESENTING THE GEOMETRIC PROPERTIES OF TWO UNCALIBRATED VIEWS

The previous section has considered the case of a single camera. In this section, we deal with a set of two cameras. Its main goal is to provide a new representation of the geometric relations between two uncalibrated cameras, which can serve as a basis for the canonical decomposition. The main facts about the new representation are summarized in the beginning of each subsection of this section.

We start by basic algebraic descriptions, which are the essential part of this section. As a consequence of the linear projective model, there is a bilinear constraint between images of corresponding points described by the fundamental matrix, which is reviewed in Section 3.1. The fundamental matrix is a description of the geometric relations between two views at the projective level. A new equivalent description, based on the epipolar projection matrix $\mathbf{S}$, which will prove to be more suitable for the purposes of this paper, is also introduced. The affine level is considered next, in Section 3.2, where $\mathbf{H}_\infty$, the infinity homography matrix and r_∞, the relative vector representation of the plane at infinity, are introduced. The two last subsections might be skipped by a reader only interested by the basic results associated with the new representation. To fully understand the new description, one must keep in mind its link with the projective structure of planes, and this elaboration is the subject of Section 3.3. We then give a geometric interpretation of the construction introduced so far by algebraic means in Section 3.4, and list additional properties.

In this section and the following, all the quantities with primes are obtained from the quantity without primes by exchanging the two images.

3.1. *The Projective Level: Epipolar Geometry*

MAIN RESULTS. *Epipolar geometry encodes the projective relations between two uncalibrated images taken by a pinhole camera. Its classical algebraic representation involves fundamental matrices: for a point* $\mathbf{m}$ *in the first image, its epipolar line is given by* $\mathbf{Fm}$. *Two corresponding points are related by the relation* $\mathbf{m}'^T\mathbf{Fm} = 0$. *The new representation* $\mathbf{S}, \mathbf{e}'$ *composed of the epipolar projection matrix* $\mathbf{S} = [\mathbf{e}']_\times/\|\mathbf{e}'\|^2\ \mathbf{F}$ *and the epipole in the second image* (*defined by* $\mathbf{F}^T\mathbf{e}' = 0$) *is equivalent to the fundamental matrix, since* $\mathbf{F} = [\mathbf{e}']_\times\mathbf{S}$.

Fundamental Matrices. When considering two projective views, the main geometric property is known in computer vision as the epipolar constraint. It can readily be understood by looking at the left part of Fig. 1. Let $\mathbf{C}$ (resp. $\mathbf{C}'$) be the optical center of the first camera (resp. the second). The line $\langle \mathbf{C}, \mathbf{C}' \rangle$ projects to a point $\mathbf{e}$ (resp. $\mathbf{e}'$) in the first retinal plane $\mathscr{R}$ (resp. in the second retinal plane $\mathscr{R}'$). The points $\mathbf{e}$, $\mathbf{e}'$ are the epipoles. The lines through $\mathbf{e}$ in the first image and the lines through $\mathbf{e}'$ in the second image are the epipolar lines. The epipolar constraint is well known in stereovision: for each point $\mathbf{m}$ in the first retina, its corresponding point $\mathbf{m}'$ lies on its epipolar line $\mathbf{l}'_m$, projection of $\langle \mathbf{C}, \mathbf{M} \rangle$ in the second retina.

Let us consider the one-parameter family of planes going through $\langle \mathbf{C}, \mathbf{C}' \rangle$. This family is a pencil of planes, shown at the bottom of Fig. 1. Let Π be any plane containing $\langle \mathbf{C}, \mathbf{C}' \rangle$. Then Π projects to an epipolar line l in the first image and to an epipolar line l' in the second image. The correspondences $\Pi \to l$ and $\Pi \to l'$ are homographies between the two pencils of epipolar lines and the pencil of planes containing $\langle \mathbf{C}, \mathbf{C}' \rangle$. It follows that the correspondence $l \to l'$ is a homography (which is a *projective linear* transformation, that is, linear in projective coordinates), called the *epipolar transformation*.

An algebraic formulation of these properties has been introduced in [16, 37] thanks to the key notion of *fundamental matrix*, or F-matrix. It can be shown only from the hypothesis (2) that the relationship between the projective retinal coordinates of a point $\mathbf{m}$ and the projective coordi-

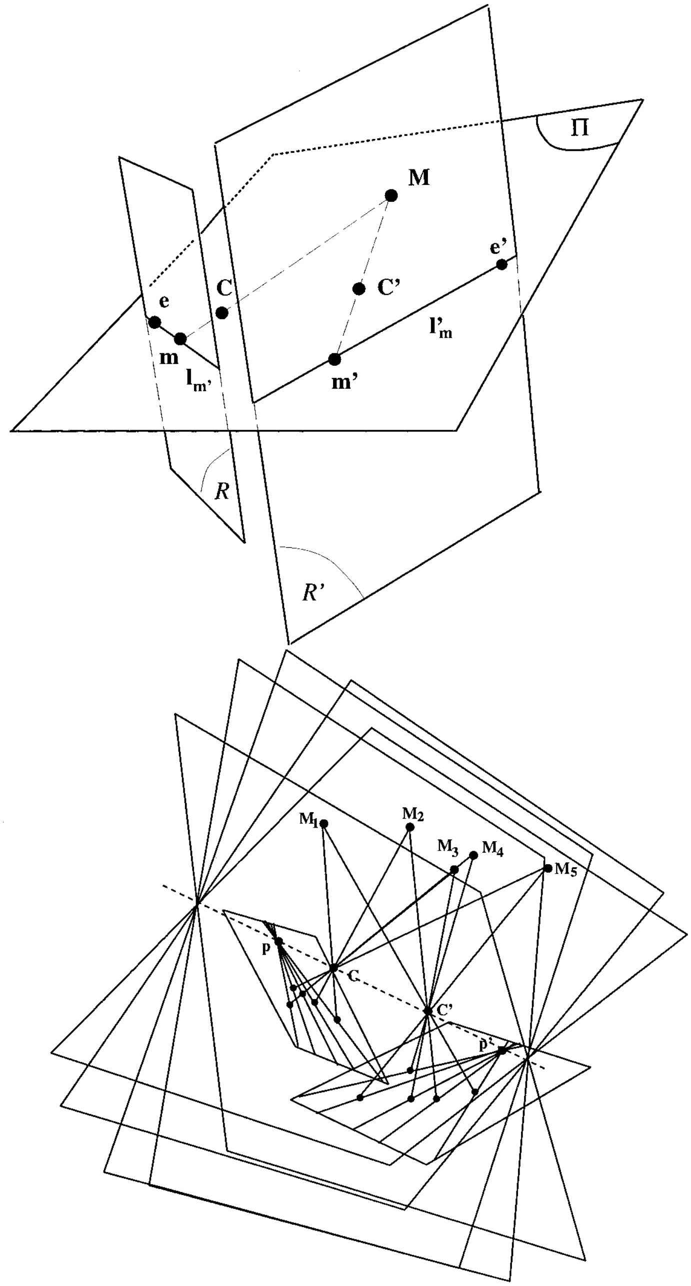


FIG. 1. The epipolar geometry and epipolar pencils.

nates of the corresponding epipolar line $\mathbf{l}'_m$ is linear. The *fundamental matrix* describes the correspondence

$$\begin{bmatrix} l'_1 \\ l'_2 \\ l'_3 \end{bmatrix} = \mathbf{l}'_m = \mathbf{Fm} = \mathbf{F} \begin{bmatrix} x_1 \\ x_2 \\ x_2 \end{bmatrix}. \tag{8}$$

The epipolar constraint has then a very simple expression: since the point $\mathbf{m}'$ corresponding to $\mathbf{m}$ belongs to the line $\mathbf{l}'_m$ by definition, it follows that

$$l'_1 x'_1 + l'_2 x'_2 + l'_3 x'_3 = \mathbf{m}'^T \mathbf{Fm} = 0. \tag{9}$$

From this relation, we remark that by switching the two images, $\mathbf{F}$ is transformed into $\mathbf{F}^T$. The epipolar transformation is characterized by the 2×2 projective coordinates of the epipoles $\mathbf{e}$ and $\mathbf{e}'$ (which are defined respectively by $\mathbf{Fe} = 0$ and $\mathbf{F}^T\mathbf{e}' = 0$), and by the three coefficients of the homography between the two pencils of epipolar lines. It follows that the epipolar transformation, like the fundamental matrix, depends on seven independent parameters, which represent the only generic information relating two uncalibrated views. Unless further hypotheses are made, there is no way to extract other geometric parameters from correspondences, since one in this case, must assume that the transformation between the two retinal frames is a general projective transformation, whereas the fundamental matrix contains the only geometric quantities which are invariant by any projective transformation. Thus, the fundamental matrix can be described as an *invariant of views*, which means that it is a function of the projection matrices which is left invariant by any projective transformation of the 3D space $\mathscr{P}^3$.

We want to extend this idea to show that the fundamental matrix and the classical description in terms of intrinsic parameters, rotations, and translations can be represented in a single framework, which includes also a new intermediate representation. We will see that the fundamental matrix represents indeed the minimal information (two views, no additional hypotheses), in a hierarchy of representations obtained by making further assumptions and adding views.

A New Representation: The **S**-*Matrix.* Using the relation

$$\|\mathbf{v}\|^2 \mathbf{I}_3 = \mathbf{v}\mathbf{v}^T - [\mathbf{v}]_\times^2 \tag{10}$$

where we have denoted by $[\mathbf{v}]_\times$ the antisymmetric matrix such that $[\mathbf{v}]_\times \mathbf{x} = \mathbf{v} \times \mathbf{x}$ for all vectors $\mathbf{x}$

$$[\mathbf{v}]_\times = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix},$$

it is seen that

$$\mathbf{F} = \underbrace{\frac{\mathbf{e}'}{\|\mathbf{e}'\|}\mathbf{e}'^T\mathbf{F}}_{\mathbf{0}} + [\mathbf{e}']_\times \underbrace{\left(-\frac{[\mathbf{e}']_\times}{\|\mathbf{e}'\|^2}\mathbf{F}\right)}_{\mathbf{S}}. \tag{11}$$

This relation defines a new matrix $\mathbf{S}$ which is determined by $\mathbf{F}$ (since $\mathbf{e}'$ is determined by $\mathbf{F}$) and shows at the same time that $\mathbf{F}$ is determined by $\mathbf{S}$ and $\mathbf{e}'$. A more natural way to introduce $\mathbf{S}$ is given in Section 3.3. An analogy can be noted with the decomposition of the essential matrix (see (53) for a definition) 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, for geometric reasons which will be explained in Section 3.4.

The two reasons why the $\mathbf{S}$-matrix is a more appropriate representation of projective information than the $\mathbf{F}$-matrix will be apparent below:

- it allows affine information to be naturally added to the representation (Section 3.2),
- it allows several views to be combined in a simple manner (Section 5.2).

Please note that although the scale of $\mathbf{S}$, like the scale of $\mathbf{F}$, cannot be recovered from two views, it will play an important role nevertheless, as stressed in Section 4.2.

Relation with Projection Matrices. We use the affine decomposition of the projection matrices

$$\tilde{\mathbf{P}} = [\mathbf{P}, \mathbf{p}], \quad \tilde{\mathbf{P}}' = [\mathbf{P}', \mathbf{p}'].$$

The epipole in the second image is the projection of the optical center of the first camera into the second camera, thus, using (4)

$$\mathbf{e}' = \tilde{\mathbf{P}}' \begin{bmatrix} -\mathbf{P}^{-1}\mathbf{p} \\ 1 \end{bmatrix} = \mathbf{p}' - \mathbf{P}'\mathbf{P}^{-1}\mathbf{p}. \tag{12}$$

The epipolar line of a point $\mathbf{m}$ of the first retina is defined by the image from the second camera of two particular points of the optical ray $\langle \mathbf{C}, \mathbf{M}\rangle$: the optical center $\mathbf{C}$ (which is projected to the epipole $\mathbf{e}'$) and the point of infinity of $\langle \mathbf{C}, \mathbf{M}\rangle$. This point is projected to

$$\tilde{\mathbf{P}}' \begin{bmatrix} \mathbf{P}^{-1}\mathbf{m} \\ 0 \end{bmatrix} = \mathbf{P}'\mathbf{P}^{-1}\mathbf{m}.$$

The projective representation of the epipolar line l'_m is obtained by taking the cross product of these two points, and it can be seen that this expression is linear in $\mathbf{m}$, as expected:

$$\mathbf{l}'_m = [\mathbf{p}' - \mathbf{P}'\mathbf{P}^{-1}\mathbf{p}] \times \mathbf{P}'\mathbf{P}^{-1}\mathbf{m}.$$

This gives us $\mathbf{F}$ as a function of projection matrices:

$$\mathbf{F} = [\mathbf{p}' - \mathbf{P}'\mathbf{P}^{-1}\mathbf{p}]_\times \mathbf{P}'\mathbf{P}^{-1}. \tag{13}$$

Furthermore, by combining with (11), we obtain the expression of $\mathbf{S}$ as a function of projection matrices:

$$\mathbf{S} = [p' - \mathbf{P}'\mathbf{P}^{-1}\mathbf{p}]_\times^2 \mathbf{P}'\mathbf{P}^{-1} / \|\mathbf{p}' - \mathbf{P}'\mathbf{P}^{-1}\mathbf{p}\|^2. \tag{14}$$

3.2. *The Affine Level: Plane at Infinity*

MAIN RESULTS. *Two most useful representations of affine information are the infinity homography $\mathbf{H}_\infty$, defined as the correspondence between images of points at infinity, and the $\mathbf{r}_\infty$-vector, relative representation of the plane at infinity such that $\mathbf{H}_\infty = \mathbf{S} + \mathbf{e}'\mathbf{r}_\infty^T$. The $\mathbf{r}_\infty$-vector represents the information which must be added to the projective representation to obtain the affine representation*; *thus*, *two alternative forms of the latter are $\mathbf{H}_\infty$, $\mathbf{e}'$, and $\mathbf{S}$, $\mathbf{e}'$, $\mathbf{r}_\infty$.*

The Infinity Homography. We have seen in Section 2.2 that if a projection matrix is $\tilde{\mathbf{P}} = [\mathbf{P}, \mathbf{p}]$, $\mathbf{P}$ can be considered as the homography between the plane at infinity Π_∞ and the retinal plane $\mathcal{R}$. If a second projection matrix $\tilde{\mathbf{P}}' = [\mathbf{P}', \mathbf{p}']$ is considered, then the transformation

$$\mathbf{H}_\infty = \mathbf{P}'\mathbf{P}^{-1} \tag{15}$$

is a homography from the first image to the second image, which maps vanishing points to vanishing points, as already remarked by [50]. It is called the *infinity homography,* and it will be seen in Section 6.2 that it is a limit of homographies generated by planes at finite distances. It allows us to determine whether two lines of $\mathcal{P}^3$ are parallel or not just by checking if their intersection in the first image is mapped to their intersection in the second image by $\mathbf{H}_\infty$. That fact is equivalent to whether the intersection of the two lines of $\mathcal{P}^3$ belong to Π_∞. Thus, the knowledge of $\mathbf{H}_\infty$ determines the parallelism of lines of $\mathcal{P}^3$, which is consistent with the fact that affine transformations preserve parallelism.

The Relative Representation Vector $\mathbf{r}_\infty$. We can now write the following equivalent relations between the infinity homography and the two matrices introduced so far to describe the epipolar geometry:

$$\mathbf{F} = [\mathbf{e}']_\times \mathbf{H}_\infty \quad \mathbf{S} = -\frac{[\mathbf{e}']_\times^2}{\|\mathbf{e}'\|^2}\mathbf{H}_\infty. \tag{16}$$

The first one is just (13) combined with (15), whereas the second one is obtained by combining also with (11). They enable us to relate the fundamental matrix to affine quantities. The first important consequence is that, if $\mathbf{H}_\infty$ is known, then the additional information contained in the $\mathbf{F}$-matrix, is equivalent to that in contained in $\mathbf{e}'$. The fact that $\mathbf{H}_\infty$ and $\mathbf{e}'$ determine $\mathbf{F}$ is obvious from (16), whereas we have already seen that $\mathbf{F}$ determines $\mathbf{e}'$. The second is that the relation puts constraints on $\mathbf{H}_\infty$ once the fundamental matrix is known (see Section 3.3 for an analysis of these constraints). However, there is no uniqueness of the decomposition. For instance, we have already seen that the relation also holds if $\mathbf{H}_\infty$ is replaced by $\mathbf{S}$. This means just that the knowledge of the $\mathbf{F}$-matrix does not allow one to recover completely $\mathbf{H}_\infty$. The missing parameters can be obtained by applying again (10)

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

The vector $\mathbf{r}_\infty$ (also discovered by [22]) which appears in this relation is very important. Once the fundamental matrix is known, its knowledge is equivalent to the knowledge of the infinity homography matrix, or, in other words, the plane at infinity. It is thus a representation of the plane at infinity *relative* to the fundamental matrix, hence its name. It is seen that exactly three parameters are needed to obtain the plane at infinity from the projective representation. This is consistent with the fact already mentioned in [14, 52, 57] that an affine structure still has three degrees of freedom when the only information available to relate the two images is the epipolar geometry.

3.3. *On the Structure of Planes*

MAIN RESULTS. *For a given plane* Π, *there exists a matrix* $\mathbf{H}_\Pi$ *such that the images* $\mathbf{m}$ *and* $\mathbf{m}'$ *of any point of* Π *are related by* $\mathbf{m}' = \mathbf{H}_\Pi\mathbf{m}$. *The fundamental matrix of the two views can be factored as* $\mathbf{F} = [\mathbf{e}']_\times\mathbf{H}_\Pi$. *The factorization is not unique, since* Π *can be any plane* (3 *degrees of freedom*). *Using the new representation* $\mathbf{S}$, $\mathbf{e}'$ *allows one to characterize projectively* Π *by the relative representation vector* $\mathbf{r}_\Pi$ *such that* $\mathbf{H}_\Pi = \mathbf{S} + \mathbf{e}'\mathbf{r}_\Pi^T$.

The Correspondence between the Two Images of a Plane. Let $\mathbf{M}_i$ be space points which happen to lie in the same plane Π and $\mathbf{m}_i$ be their images by a projective linear relation from $\mathscr{P}^3$ to $\mathscr{P}^2$. Its restriction to Π is a projective linear relation between points of $\mathscr{P}^2$, which is a homography h. This relation is invertible, in the generic case, where Π does not contain the optical center of the camera. If two images of the points $\mathbf{M}_i$ lying in a generic plane, $\mathbf{m}_i$ and $\mathbf{m}_i'$ are available, we can consider the relation $h' \circ h^{-1}$ between these two images. It is a homography; therefore, there is a 3×3 invertible matrix $\mathbf{H}_\Pi$, such that the following projective relation holds for each i:

$$\mathbf{m}_i' = \mathbf{H}_\Pi\mathbf{m}_i. \tag{18}$$

It is clear that by exchanging the roles of the two images, the homography matrix is transformed to its inverse $\mathbf{H}_\Pi^{-1}$. Let us also point out the composition relation between homographies generated by a plane Π between three images:

$$\mathbf{H}_{13} = \mathbf{H}_{23}\mathbf{H}_{12}.$$

It is easy to prove this result by decomposing each homography $\mathbf{H}_{ij}$ as a product of the homographies between retina $\mathscr{R}_j$ and plane Π, and between plane Π and retina $\mathscr{R}_i$. In the case where the plane Π contains the optical center of the first camera, $\mathbf{H}_\Pi$ is not defined. In the case where Π contains the optical center of the second camera, $\mathbf{H}_\Pi$ is of rank two. It still defines a projective linear relation, but this relation is no longer a homography, but rather a projection. The image of the epipole in the first image $\mathbf{e}$ is undefined, whereas the first retina is mapped only on the line which is the intersection of the second retina and the plane Π.

Factorizations of the Fundamental Matrix. It has been shown [35, 37] that any homography matrix $\mathbf{H}_\Pi$ generated by a plane between two views and the fundamental matrix between the two same views are related by the system of equations

$$\mathbf{H}_\Pi^T\mathbf{F} + \mathbf{F}^T\mathbf{H}_\Pi = 0. \tag{19}$$

A matrix $\mathbf{H}_\Pi$ is compatible with $\mathbf{F}$ (i.e., satisfies condition (19)) if and only if

$$\mathbf{F} = [\mathbf{e}']_\times\mathbf{H}_\Pi. \tag{20}$$

This last condition has been also presented in [25, 57]. It is straightforward to verify that (19) is satisfied if the substitution of (20) is made in that equation. Now let us suppose that $\mathbf{F}$ and $\mathbf{H}_\Pi$ satisfy (19) and consider a pair of corresponding points $\mathbf{m}$ and $\mathbf{m}'$. Equation (19) is equivalent to the fact that $\mathbf{F}^T\mathbf{H}_\Pi$ is an antisymmetric matrix; thus,

$$(\mathbf{Fm})^T\mathbf{H}_\Pi\mathbf{m} = 0. \tag{21}$$

Since the fundamental matrix maps points to corresponding epipolar lines, $\mathbf{Fm} = \mathbf{e}' \times \mathbf{m}'$. Using this relation, we see that (21) is equivalent to $\mathbf{m}'^T[\mathbf{e}']_\times\,\mathbf{H}_\Pi\mathbf{m} = 0$. If we identify this equation with the epipolar constraint $\mathbf{m}'^T\mathbf{Fm} = 0$, we obtain the expression (20).

Note that the proof does not depend on the fact that $\mathbf{H}_\Pi$ is an invertible (or homography) matrix. In the case where this matrix is singular, it also defines a mapping from the first plane to the second plane, but this mapping is not invertible. We also obtain as an easy consequence of (19) that if $\mathbf{H}_\Pi$ is compatible with $\mathbf{F}$, then

$$\mathbf{H}_\Pi\mathbf{e} \cong \mathbf{e}' \text{ or } \mathbf{H}_\Pi\mathbf{e} = 0. \tag{22}$$

The first case corresponds to the homography matrices, defined by a plane in general position. The second case corresponds to the degenerate case where the plane contains the optical center $\mathbf{C}'$, thus yielding a noninvertible correspondence.

It is obvious that the decomposition (20) is not unique, since $\mathbf{H}_\Pi$ can be any matrix defining a correspondence compatible with $\mathbf{F}$. More precisely, since the matrix equation (19) includes six homogeneous equations, such matrices form a three-dimensional set, each of them being defined by three corresponding points, as shown in [53]. This is consistent with the fact that a plane is defined by three points. If two matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ satisfy (20), then $[\mathbf{e}']_\times \mathbf{M} = 0$, where $\mathbf{M} = \lambda \mathbf{H}_1 - \mathbf{H}_2$. This implies that $\mathbf{M} = \mathbf{e}'\mathbf{r}^T$ for a certain vector $\mathbf{r}$. Thus, we find (as in [25]) that any two matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ satisfying the decomposition (20) are related by

$$\lambda \mathbf{H}_2 = \mathbf{H}_1 + \mathbf{e}'\mathbf{r}^T. \tag{23}$$

In Section 3.1, we defined a special matrix $\mathbf{S}$ compatible with $\mathbf{F}$, and *which is only a function of* $\mathbf{F}$. Although the matrix $\mathbf{S}$ is singular, the fact that it is compatible with the fundamental matrix allowed us to interpret this matrix as a correspondence between the two retinas induced by a plane. *An alternative way of defning* $\mathbf{S}$ *is to add an additional constraint to the condition* (20) *to ensure the uniqueness*:

$$\mathbf{F} = [\mathbf{e}']_\times \mathbf{S} \quad \mathbf{S}^T\mathbf{e}' = 0. \tag{24}$$

Let $\mathbf{H}$ be a homography matrix compatible with $\mathbf{F}$. By transposing (23) we get $\mathbf{H}^T = \mathbf{S}^T + \mathbf{r}\mathbf{e}'^T$. Since $\mathbf{S}^T\mathbf{e}' = 0$, we obtain $\mathbf{r} = \mathbf{H}^T\mathbf{e}'/\|\mathbf{e}'\|^2$. By using the relation (10), a substitution of this value back in (23) yields $\mathbf{S} = -[\mathbf{e}']_\times^2/\|\mathbf{e}'\|^2\ \mathbf{H}$. Since $\mathbf{H}$ is compatible with $\mathbf{F}$ the relation (20) applies, and the expression given in (11), independent of $\mathbf{H}$, is again found for $\mathbf{S}$.

Once the fundamental matrix is known, Eq. (23) can be used to characterize any plane Π relatively to this matrix by the vector $\mathbf{r}_\Pi$ such that

$$\mathbf{H}_\Pi = \mathbf{S} + \mathbf{e}'\mathbf{r}_\Pi^T. \tag{25}$$

In this equation, two types of parameters appear. The *projective* description of the structure of Π is the relative representation of Π, $\mathbf{r}_\Pi$. The motion parameters $\mathbf{S}$ and $\mathbf{e}'$ describe the projective relation between the two image frames.

3.4. *A Geometric Interpretation*

MAIN RESULTS. *The* $\mathbf{S}$*-matrix can be described as the correspondence induced by the plane containing the optical center of the second camera and the line* $\langle \mathbf{e}' \rangle$. *This correspondence is a projection from the first retinal plane to a line of the second retinal plane, and associates to a point the intersection of its epipolar line with* $\langle \mathbf{e}' \rangle$. *Dual relations are obtained, considering epipolar lines. The matrix* $\mathbf{S}^T$ *maps lines to lines*, *and the matrix* $\mathbf{G} = [\mathbf{e}']_\times \mathbf{F} [\mathbf{e}]_\times$ *maps lines to points*.

The Epipolar Projection. We drop the scale factors, since we are not seeking algebraic relations, but rather geometric properties.

It is shown in Appendix B that the matrix $\mathbf{S}$ describes the correspondence from image 1 to image 2 generated by the plane $\Pi_{\mathbf{e}'}$ of projective equation:

$$\Pi_{\mathbf{e}'} = \begin{bmatrix} \mathbf{P}'^T\mathbf{e}' \\ \mathbf{p}'^T\mathbf{e}' \end{bmatrix}.$$

It can be easily verified that this plane contains the optical center of the second camera

$$\mathbf{C}' = \begin{bmatrix} \mathbf{P}'^{-1}\mathbf{p}' \\ -1 \end{bmatrix}.$$

Another propriety which can also be verified in the canonical decomposition proposed in Section 4 is that the homography $\mathscr{H}$ from the first view to the second view maps the plane $\Pi_{\mathbf{e}'}$ to the plane at infinity Π_∞. This is why the plane $\Pi_{\mathbf{e}'}$ is a characterization of the *affine relations* between the two cameras. However, this plane *cannot* be recovered from the images, because the correspondence it generates is singular. The correspondence from the first image to the second image is thus a mapping to the line which is the image of $\Pi_{\mathbf{e}'}$ in the second retina, and this mapping is a projection and is not invertible, since the plane contains the optical center of the second camera.

A plane Π *of direction* $\mathbf{d}$ *containing the optical center* $\mathbf{C}$ *of a camera with projection matrix* $\tilde{\mathbf{P}} = [\mathbf{P}, \mathbf{p}]$ *is projected on the line* $\mathbf{P}^{-1T}\mathbf{d}$ *by this camera*. By writing the equivalence of the fact that Π contains $\mathbf{M}$, $\Pi^T\mathbf{M} = 0$ and of the fact that its image $\mathbf{l}_\Pi$ contains the projection of $\mathbf{M}$: $\mathbf{l}_\Pi^T\mathbf{P}\mathbf{M} = 0$, we obtain two equations. They are compatible because $\mathbf{C}$ belongs to Π, $\Pi^T\mathbf{C} = 0$, and eventually yield $\mathbf{l}_\Pi = \mathbf{P}^{-1T}\mathbf{d}$.

Using this result, we can identify matrix $\mathbf{S}$ as the correspondence defined by the plane $\Pi_{\mathbf{e}'}$ which contains the optical center of the second camera, and whose image in the second camera is the line $\langle \mathbf{e}' \rangle$, as seen in Fig. 2. We see that this is consistent with definition (11). The matrix $\mathbf{F}$ maps points to lines, the matrix $[\mathbf{e}']_\times$ either maps lines to points or points to lines; thus, from (11), 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

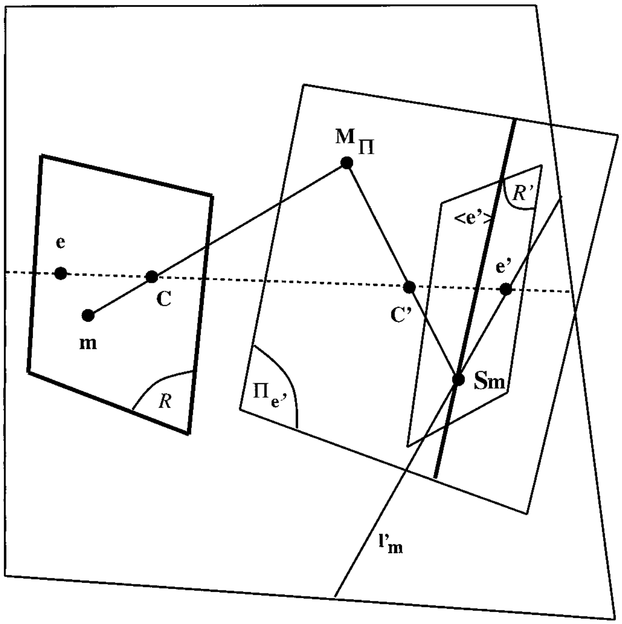


FIG. 2. The epipolar projection.

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 (11) is that the epipolar line $\mathbf{l}'_{\mathbf{m}} = \mathbf{FM}$ is defined by joining the epipole $\mathbf{e}'$ and the point $\mathbf{m}'_1 = \mathbf{Sm}$ (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 (five parameters) and the epipole (two parameters) completely define the 3×3 matrices $\mathbf{S}$ and $\mathbf{F}$, which is consistent with the result that the fundamental matrix depends on seven parameters. We have thus exhibited a new decomposition of this matrix in two subsets of independent parameters, which has a sound geometric interpretation in terms of epipolar mappings. It is illustrated in Fig. 2.

Dual Relations. Consider $\mathbf{S}^T = \mathbf{F}^T[\mathbf{e}']_\times$, and an epipolar line $\mathbf{l}'$ of the second retina. Then $\mathbf{m}'_1 = \mathbf{e}' \times \mathbf{l}'$ is a point of $\mathbf{l}'$ which is always defined if $\mathbf{l}' \neq \langle \mathbf{e}' \rangle$. Then $\mathbf{l} = \mathbf{F}^T\mathbf{m}'_1$ is the epipolar line of $\mathbf{m}'_1$, but since $\mathbf{m}'_1$ is a point of $\mathbf{l}'$, $\mathbf{l}$ and $\mathbf{l}'$ are corresponding epipolar lines. Thus, the matrix $\mathbf{S}^T$ maps epipolar lines to corresponding epipolar lines (as do the homography matrices [25, 58]). From relation (11), and the fact that $\mathbf{F}' = \mathbf{F}^T$, we obtain

$$\mathbf{S}' = [\mathbf{e}]_\times \mathbf{S}^T [\mathbf{e}']_\times$$

which can be interpreted as follows: a point $\mathbf{m}'$ is mapped to an epipolar line $\mathbf{l}'$ passing through it, which is then mapped into the corresponding epipolar line $\mathbf{l}$. The final result is the intersection of this line with the line $\langle \mathbf{e} \rangle$ which is what we expected.

Let us now introduce the matrix

$$\mathbf{G} = [\mathbf{e}']_\times \mathbf{F} [\mathbf{e}]_\times. \tag{26}$$

Using the relation (10), we find with some easy algebra that $\mathbf{F} = [\mathbf{e}']_\times \mathbf{G} [\mathbf{e}]_\times$. The matrix $\mathbf{G}$ plays a dual role with respect to the matrix $\mathbf{F}$, in the sense that if we write the relations

$$\mathbf{l} = \mathbf{F}^T\mathbf{m}' = \mathbf{e} \times \mathbf{m} = \mathbf{S}'\mathbf{l}'$$
$$\mathbf{l}' = \mathbf{Fm} = \mathbf{e}' \times \mathbf{m}' = \mathbf{Sl},$$

then we have the equivalence:

$$\mathbf{m}'^T\mathbf{Fm} = 0 \Leftrightarrow (\mathbf{e}' \times \mathbf{m}')^T\mathbf{G}(\mathbf{e} \times \mathbf{m}) = 0 \Leftrightarrow \mathbf{l}'^T\mathbf{Gl} = 0.$$

The matrix $\mathbf{G}$ maps lines to points, and, like $\mathbf{F}$, it satisfies

$$\mathbf{G}' = \mathbf{G}^T \quad \mathbf{Ge} = 0 \quad \mathbf{G}'\mathbf{e}' = 0.$$

The image of an epipolar line $\mathbf{l}$ passing through a point $\mathbf{m}$ is

$$\begin{aligned}\mathbf{Gl} &= [\mathbf{e}']_\times \mathbf{F} [\mathbf{e}]_\times (\mathbf{e} \times \mathbf{m}) = [\mathbf{e}']_\times \mathbf{F} [\mathbf{e}]^2_\times \mathbf{m} \\ &= \mathbf{e}' \times \mathbf{Fm} = \mathbf{e}' \times \mathbf{l}'.\end{aligned}$$

This point, which is nothing else than $\mathbf{Sm}$, does not depend on the choice of $\mathbf{m}$. The composition of $\mathbf{F}$, followed by $\mathbf{G}'$ gives thus the point obtained at the intersection of the original epipolar line, and the line $\langle \mathbf{e} \rangle$. The composition of $\mathbf{G}$ followed by $\mathbf{F}^T$ gives back the original epipolar line.

The Affine Level. To determine the infinity homography $\mathbf{H}_\infty$ three more parameters are needed. These parameters are, for example, the coordinates of the vector $\mathbf{r}_\infty = \mathbf{H}_\infty^T\mathbf{e}'/\|\mathbf{e}'\|^2$, introduced previously to characterize the infinity homography once the epipolar geometry is known. The direction of this vector defines the line in the first retina which is the image by $\mathbf{H}_\infty^T$ of the special line $\langle \mathbf{e}' \rangle$ of the second retina. This line is the projection of the plane parallel to $\Pi_{\mathbf{e}'}$ and containing the optical center $\mathbf{C}$ of the first camera. Note that a general property of the dual homography $\mathbf{H}_\infty^T$ is that two lines $\mathbf{l}$ and $\mathbf{l}'$, respective projections of the planes Π and Π' in each retina, are in correspondence by $\mathbf{H}_\infty^T$ if and only if the planes Π and Π' are parallel. This comes from the relation between planes and lines established at the beginning of the section. Taking

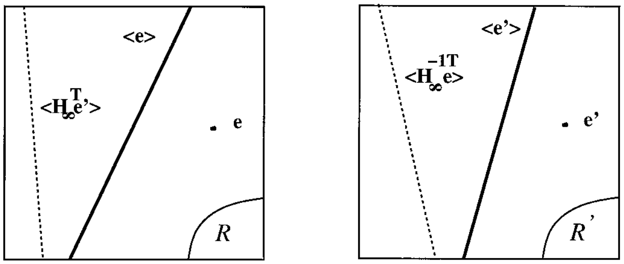


FIG. 3. Projective and affine information in the retinal planes.

only the direction of $\mathbf{r}_\infty$ drops one parameter, since we have lost the information corresponding to the norm of this vector. Thus, knowing the epipolar geometry and the line $\langle \mathbf{H}_\infty^T \mathbf{e}' \rangle$ is not sufficient to determine $\mathbf{H}_\infty^T$, and we need another piece of geometric information, for instance, the line $\langle \mathbf{H}_\infty^{-1T} \mathbf{e} \rangle$ in the second retina. The lines $\langle \mathbf{e} \rangle$ and $\mathbf{H}_\infty^T \mathbf{e}' \rangle$ are never identical, since we have $\mathbf{e}^T \mathbf{H}_\infty^T \mathbf{e}' = -\mathbf{e}'^T \mathbf{e}' \neq 0$, and thus these two line correspondences, together with the epipolar projection $\mathbf{S}$ are sufficient to characterize the infinity homography. The situation on the retinas is summarized in Fig. 3.

4. A LOCAL CANONICAL DECOMPOSITION

In this section, we introduce the canonical decomposition for two views. The basic insight of the canonical decomposition is first explained in Section 4.1. It allows one to obtain a unified representation invariant under different groups of transformation. A very important but somewhat tricky issue is that of scale. It is discussed in Section 4.2, which might be skipped in a first reading. The results of the section are summarized in Table 2. Please note that although the uncalibrated representations of two views were introduced and explained in the previous section, they are in fact entirely rederived in this table, using just the insight of the canonical decomposition.

4.1. *The Idea of the Canonical Decomposition*

If two projective views are considered, the most complete description is given through the two projection matrices $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{P}}'$. 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 $\mathscr{G}$ of transformation of the projective space $\mathscr{P}^3$, which will be referred to as *descriptions of level* $\mathscr{G}$ or $\mathscr{G}$-invariant descriptions. The properties which can be recovered from these descriptions are those which are left invariant by the transformations of $\mathscr{G}$.

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

$$\{(\tilde{\mathbf{P}}_1, \tilde{\mathbf{P}}_1') \,\Re(\mathscr{G})(\tilde{\mathbf{P}}_2, \tilde{\mathbf{P}}_2')\}$$
$$\Leftrightarrow \{\exists \mathscr{T} \in \mathscr{G}, \tilde{\mathbf{P}}_2 = \tilde{\mathbf{P}}_1 \mathscr{T} \quad \text{and} \quad \tilde{\mathbf{P}}_2' = \tilde{\mathbf{P}}_1' \mathscr{T}\}.$$

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

- for calibrated cameras, the normalized coordinate system $\mathscr{I} = \mathbf{A}\tilde{\mathbf{P}}_C = [\mathbf{A}, \mathbf{0}]$
- for uncalibrated cameras, the pixel coordinate system, as also done for instance by Hartley (see [26] and several other papers which therefore fit into our formalism): $\mathscr{I} = \tilde{\mathbf{P}}_C = [\mathbf{I}_3, \mathbf{0}]$.

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

> The matrices $\mathscr{I}, \mathscr{I}'$, expressed as functions of a pair of generic projection matrices $\tilde{\mathbf{P}}, \tilde{\mathbf{P}}'$, such that there is a unique decomposition, called canonical
>
> $$\tilde{\mathbf{P}} = \mathscr{I}\mathscr{T} \quad \tilde{\mathbf{P}}' = \mathscr{I}'\mathscr{T}, \tag{27}$$
>
> with $\mathscr{T}$ being an element of $\mathscr{G}$, provide a complete description of the geometric properties *of two projective views* which are left invariant by the group of transformation $\mathscr{G}$.

Note that here the invariant is attached to the set of cameras, and not a set of 3D objects observed by the cameras, unlike the invariants considered in [46]. Let us list some consequences of this construction:

- The sum of the number of parameters in the representation $\mathscr{I}, \mathscr{I}'$ and in the generic transformation $\mathscr{T}$ must be 22.
- Every quantity which depends only on the projection matrices and is invariant with respect to $\mathscr{G}$ is also a function of $\mathscr{I}$ and $\mathscr{I}'$.
- The quantities which appear in matrix $\mathscr{T}$ are not measurable from two views using the representation of level $\mathscr{G}$. But they may be expressed using representations of the previous level, instead.
- The decomposition provides a tool for explicitly

TABLE 2
The Geometries of Two Views: Canonical Representation

EUCLIDEAN (motion)	Similarities *preserve* angles, relative distances			
$\mathcal{GE}_3$	$\mathcal{S} = \begin{bmatrix} \mathcal{R} & \mathcal{T} \\ 0_3^T & \lambda \end{bmatrix}$	$\mathcal{R}$: rotation matrix $\mathcal{T}$: translation vector λ: nonnull scalar		7
Invariant description	$\mathbf{A}$, $\mathbf{A}'$: intrinsic parameters of cameras $\mathbf{R}$: rotation from camera 1 to camera 2 $\mathbf{t}$: direction of translation from camera 1 to camera 2			5 + 5 3 2
Canonic decomposition	$\begin{cases} \tilde{\mathbf{P}} = \mathbf{A}[\mathbf{I}_3, 0]\mathcal{S} \\ \tilde{\mathbf{P}}' = \mathbf{A}'[\mathbf{R}, \mathbf{t}]\mathcal{S} \end{cases}$	$\begin{array}{c} \mathbf{R} = \mathbf{R}'_w \mathbf{R}_w^T \\ \mathbf{t} = \mathbf{T}/\|\mathbf{T}\| \\ \hline \mathcal{R} = \mathbf{R}_w \\ \mathcal{T} = \mathbf{T}_w \\ \lambda = \|\mathbf{T}\| \end{array}$	where $\mathbf{T} = \mathbf{T}'_w - \mathbf{R}'_w \mathbf{R}_w^T \mathbf{T}_w$	
AFFINE	Affine transformations *preserve* parallelism, center of mass			
$\mathcal{GA}_3$	$\mathcal{A} = \begin{bmatrix} \mathcal{M} & \mathcal{V} \\ \mathbf{0}_3^T & \mu \end{bmatrix}$	$\left.\begin{array}{l} \mathcal{M}\text{: non-singular } 3 \times 3 \text{ matrix} \\ \mathcal{V}\text{: 3D vector} \\ \mu\text{: non-null scalar} \end{array}\right\}$	$\mathcal{A}$ defined up to a global scale factor	12
Invariant description	$\mathbf{H}_\infty$: infinity homography from image 1 to image 2 $\mathbf{e}'_N$: normalized epipole in image 2			8 2
Canonic decomposition	$\begin{cases} \tilde{\mathbf{P}} = [\mathbf{I}_3, 0]\mathcal{A} \\ \tilde{\mathbf{P}}' = [\mathbf{H}_\infty, \mathbf{e}'_N]\mathcal{A} \end{cases}$	$\begin{array}{c} \mathbf{H}_\infty = \mathbf{P}'\mathbf{P}^{-1} \\ \mathbf{e}'_N = \mathbf{e}/\|\mathbf{e}'\| \\ \hline \mathcal{M} = \mathbf{P} \\ \mathcal{V} = \mathbf{p} \\ \mu = \|\mathbf{e}'\| \end{array}$	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			
$\mathcal{GL}_4$	$\mathcal{H} = \begin{bmatrix} \mathcal{M} & \mathcal{V} \\ \mathcal{L}_N^T & \nu_N \end{bmatrix}$	$\left.\begin{array}{l} \mathcal{M}\text{: } 3 \times 3 \text{ matrix} \\ \mathcal{V}, \mathcal{L}_N\text{: 3 D vectors} \\ \nu_N\text{: scalar} \end{array}\right\}$	$\mathcal{H}$ non-singular defined up to a global scale factor	15
Invariant description	$\mathbf{S}$: Epipolar projection from image 1 to image 2 $\mathbf{e}'_N$: Normalized epipole in image 2			5 2
Canonic decomposition	$\begin{cases} \tilde{\mathbf{P}} = [\mathbf{I}_3, 0]\mathcal{H} \\ \tilde{\mathbf{P}}' = [\mathbf{S}, \mathbf{e}'_N]\mathcal{H} \end{cases}$	$\begin{array}{c} \mathbf{S} = -[\mathbf{e}'_N]_\times^2 \mathbf{H}_\infty \\ \mathbf{e}'_N = \mathbf{e}/\|\mathbf{e}'\| \\ \hline \mathcal{M} = \mathbf{P} \\ \mathcal{V} = \mathbf{p} \\ \mathcal{L}_N = \mathbf{P}'^T \mathbf{e}'_N \\ \nu_N = \mathbf{p}'^T \mathbf{e}'_N \end{array}$	where $\mathbf{H}_\infty = \mathbf{P}'\mathbf{P}^{-1}$ $\mathbf{e}' = \mathbf{p}' - \mathbf{H}_\infty \mathbf{p}$	

Note. Generic decomposition and properties of a member of each of these groups of transformations are mentioned. A canonical decomposition is given whereby quantities above the horizontal line are the elements of the invariant description, quantities under that line are nonmeasurable.

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 reconsruction, like [14, 26].

One point which is worth noting is that the choice of the transformation group and of the form of the first canonical element specifies completely the form of the second canonical element. Since the form that we have chosen for the first canonical element is the simplest possible, this gives a derivation of "natural" descriptions of the geometry of cameras. The relevant algebra is given in Appendices A and B. Therefore, although the uncalibrated representations for two views were introduced and explained in the previous section, they are in fact entirely rederived in this section.

The Canonical Decomposition for Two Views. Let us write the decompositions introduced in Section 2

$$\tilde{\mathbf{P}} = [\mathbf{P}, \mathbf{p}] = \lambda_w \mathbf{A}[\mathbf{R}_w, \mathbf{T}_w]$$

$$\tilde{\mathbf{P}}' = [\mathbf{P}', \mathbf{p}'] = \lambda'_w \mathbf{A}'[\mathbf{R}'_w, \mathbf{T}'_w].$$

The previous sections have laid the ground work for the results which are summarized in Table 2, in which we mention:

- The characteristic properties and generic decomposition of a member of each of these groups of transformations.
- a canonical decomposition of the form (27) of two projection matrices. The quantities above the horizontal line are the elements of the invariant description, the quantities under that line are nonmeasurable.
- Indication of links with the previous level.
- The number of parameters, whose sum is exactly 22.

4.2. *Considerations on Scale*

MAIN RESULTS. *There are two sources of undetermination for the scale. The first one is the depth–speed ambiguity, inherent in any structure from motion analysis. The second one comes from the fact that the elements of the uncalibrated representations are of a projective nature. These undeterminations can be dealt with in a multiple view analysis, by considering projection matrices. A way to account for the scale is to normalize the representation and to consider scale-preserving groups.*

The Depth–Speed Ambiguity. Let us now discuss the scale factor issue, beginning with the case of calibrated cameras. In the case of two views, the extrinsic parameters of the stereo system are classically described by a rotation $\mathbf{R}$ and a translation $\mathbf{T}$, such that if $\mathbf{M}$ (resp. $\mathbf{M}'$) represents the coordinates of a point in the first (resp. second) camera coordinate system, then $\mathbf{M}' = \mathbf{R}\mathbf{M} + \mathbf{T}$. When a calibration object and its associated coordinate system are known, the projection matrices can be fully recovered by model-based calibration and the transformation between the two camera coordinate systems is described as a rigid displacement $\mathbf{R}$, $\mathbf{T}$ which leaves absolute distances invariant. However, in the structure from motion paradigm where the data used are only image measurements, there is an ambiguity between the amount of displacement, represented by $\|\mathbf{T}\|$, and the depth of objects; thus, only the direction of translation can be determined, and we must consider a global scale factor, and only relative distances. The transformation between the two camera coordinate systems then must be described by a similarity. It can be noted that the observation of just one line segment of known length would be sufficient to eliminate the scale indetermination. In the affine case, the relation between the affine group and the affine unimodular group is much the same than that just described between the similarity group and the displacement group. In the projective case, there is no such notion to our knowledge.

The groups which are chosen for the two-view analysis are classical and can be divided into two categories. On one hand there are the groups which do not leave any kind of scale invariant:

- the group of similarities $\mathcal{GE}_3$,
- the group of affine transformations $\mathcal{GA}_3$,
- the group of homographies of $\mathcal{P}^3$, $\mathcal{GL}_4$.

On the other hand, there are the groups which conserve it:

- the group of displacements $\mathcal{SE}_3$,
- the group of direct affine unimodular transformations $\mathcal{SA}_3$.

Because of the depth–speed ambiguity, the last two ones are not relevant in the context of analysis from two views, and therefore they are mentioned in Table 3 but not in Table 2.

The Projective Indetermination of Scale in the Two-View Analysis. 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 the fact that the projection matrices $\tilde{\mathbf{P}}$, $\tilde{\mathbf{P}}'$ are projective quantities. In the following table, we show how the elements of the representation are transformed, when the projection matrices are multiplied by a scale factor. It can be seen that the elements of the representation are either invariant or simply multiplied by a scale factor, and thus the representation is consistent.

$\tilde{\mathbf{P}}$	$\tilde{\mathbf{P}}'$	$\mathbf{e}'$,	$\mathbf{H}_\infty$	$\mathbf{S}$	$\mathcal{A}$	$\mathcal{H}$
$\lambda\tilde{\mathbf{P}}$	$\tilde{\mathbf{P}}'$	$\mathbf{e}'$,	$\lambda^{-1}\mathbf{H}_\infty$	$\lambda^{-1}\mathbf{S}$	$\lambda\mathcal{A}$	$\lambda\mathcal{H}$
$\tilde{\mathbf{P}}$	$\mu\tilde{\mathbf{P}}'$	$\mu\mathbf{e}'$,	$\mu\mathbf{H}_\infty$	$\mu\mathbf{S}$	$\mu\mathcal{A}$	$\mu\mathcal{H}$

It is important to note that reciprocally, multiplying any of the elements $\mathbf{e}'$, $\mathbf{H}_\infty$, $\mathbf{S}$ in the canonical decomposition will only multiply the projection matrices $\tilde{\mathbf{P}}$, $\tilde{\mathbf{P}}'$ by a scale factor, which is not significant.

Normalized Representations. 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 views. However, we will see in the next section that it is constrained if three views are considered, and thus, in order to be consistent with the sequel, we have taken a fixed-scale representation for the epipole, the normalized epipole:

$$\mathbf{e}'_N = \mathbf{e}/\|\mathbf{e}'\|. \tag{28}$$

The vectors $\mathbf{r}_\infty$, $\mathscr{L}$ and the scalar ν must be scaled accordingly, resulting in the quantities

$$\mathbf{r}_{\infty N} = \mathbf{H}^T\mathbf{e}'_N\, \mathscr{L}_N = \mathbf{P}'^T\mathbf{e}'_N\, \boldsymbol{\nu}_N = \mathbf{p}'^T\mathbf{e}'_N. \tag{29}$$

Let us now examine the fixed-scale representations.

- The displacement-invariant representation is obtained from the similarity-invariant one by setting $\lambda = 1$. Then, instead of $\mathbf{t}$, we obtain $\mathbf{T}$, and the norm of this vector becomes completely fixed. This gives the $(\mathbf{A}, \mathbf{R}, \mathbf{T})$ representation.
- The unimodular invariant representation is obtained from the affine invariant by setting $\mu = \det(\mathscr{M})$. This is equivalent to setting $\mu = 1$ and imposing the constraint $\det(\mathscr{M}) = 1$. To distinguish this representation from the affine-invariant one, we call it $(\mathbf{Q}, \mathbf{s})$.

The first case just means that the absolute scale factor is resolved. Let us now discuss the second one, which is less obvious. As done previously, we can keep the decomposition by transforming $(\mathbf{Q}, \mathbf{s})$ into $(\lambda\mathbf{Q}, \lambda\mathbf{s})$, which results just in a global scaling which does not affect the transformation matrix. In contrast, because of the constraint $\det(\mathscr{M}) = \mu$, we cannot transform $(\mathbf{Q}, \mathbf{s})$ into $(\lambda\mathbf{Q}, \mathbf{s})$. Thus, if we impose a normalization constraint on $\mathbf{s}$, say $\|\mathbf{s}\| = 1$, then the scale of $\mathbf{Q}$ is entirely determined by $\mathbf{Q} = \mathbf{P}'\mathbf{P}^{-1}/\|\mathbf{e}'\|$. Reciprocally, if we impose a normalization constraint on $\mathbf{Q}$, say $\det(\mathbf{Q}) = 1$, which is equivalent to $\mathbf{Q} = [\det(\mathbf{A})/\det(\mathbf{A}')]\ \mathbf{A}'\mathbf{R}\mathbf{A}^{-1}$, then the scale of $\mathbf{s}$ is entirely determined by $\mathbf{s} = (1/\det(\mathbf{A}'))\mathbf{A}'\mathbf{T}$ and this is why the representation depends on 11 parameters rather than 10. The first normalization is the same as that adopted for the rest of the canonical representations. The second has the advantage that in the case when two cameras have the same intrinsic parameters, the normalization constraint is *exactly* satisfied [67]. It can be noted that the pair $(\mathbf{Q}, \mathbf{s})$ is derived by an affine transform from the pair $(\mathbf{R}, \mathbf{T})$, whereas the pair $(\mathbf{H}_\infty, \mathbf{e}'_N)$ is derived from $(\mathbf{R}, \mathbf{t})$.

Determining the Scale. Although the quantities in formulas such as (12) and (13) are projective, the scale factors can be kept, and this ensures that even if we have to *add* these quantities (see, for instance, (34) in next section) consistent results are still obtained. This will be particularly important in the three-view analysis. The use of the projection matrices enables us to determine all the scale factors which are undetermined in a projective context, and to obtain, for example,

$$\mathbf{He} = -\mathbf{e}' \quad \mathbf{F}' = (\det \mathbf{H}_\infty)\mathbf{F}^T \tag{30}$$

instead of the mere projective relations (22) and $\mathbf{F}' \sim \mathbf{F}^T$. A technique which has been found useful is to express all the quantities as functions of elements of the projection matrices, which ensures that the scale factors are consistent, and to account for the fact that these scale factors are not always observable in a latter stage. This treatment, which is one of the novel features of the paper, will allow us in particular to determine scale factors wich are consistent across several views, resolving this way the undetermination which is usually associated with projective quantities. This is parallel to the work of Hartley on "cheirality" invariants [21] where the use of the visibility constraint (the fact that cameras actually observe only a half-space) made it possible to go beyond a mere projective representation, to determine signs.

5. A CANONICAL REPRESENTATION FOR MULTIPLE VIEWS

Let us consider a sequence of views. They might be viewed as a set of pairs of views, to each of which we could apply the canonical decomposition which was considered in the last section. This is the local approach. The problem is that it does not tell us anything about the relationships between the different representations. On one hand, if we consider only the consecutive pairs of views, then describing the relationship between nonconsecutive pairs of views is not obvious. It requires composition formulas, and also that the amount of information contained in consecutive pairs of views is sufficient, which in fact turns out not to be the case. On the other hand, if we consider the lattice of all pairs of views, then we have a set of parameters which is far too large, and are subjected to a number of complex constraints. Thus, a global representation is needed to account for all the geometric relationships between any camera of the sequence. The goal of this section is to describe such a representation, and to relate it to the local representations already studied.

As previously, we first lay the ground work by considering the different levels. In Section 5.1, we first consider the Euclidean and affine case, which turn out to be rather similar. The projective case, which is then studied in Section 5.2, is more tricky. The canonical decomposition for three views, and then N views is presented in Section 5.3, as a generalization of the canonical decomposition for two views. As previously, the basic insight is simple and natural and leads automatically to the results which are summarized in Table 3. When recovering the global representation from the local representations, one needs to recover some scale factors from a set of representations which are each defined only up to a scale factor. Surprisingly enough, this can be done thanks to exploiting global constraints, and explicit formulas for performing this task are given in Section 5.4, which might be skipped in a first reading.

We turn now first to the case of three views, and will use the subscripts 1, 2, 3, to designate them. A transformation noted $\mathbf{M}_{ij}$ will always map quantities of view i to quantities of view j. A quantity present in view i in relation with view j will be noted $\mathbf{v}_{ij}$. We first describe the relations between each of the three pairs of views.

5.1. *Calibrated and Uncalibrated Motion*

MAIN RESULTS. *The Euclidean and affine descriptions of motion both include a rotational part* (*calibrated*, $\mathbf{R}$, *uncalibrated*, $\mathbf{H}_\infty$, *or* $\mathbf{Q}$) *and a translational part. If the representations* 1–2 *and* 2–3 *are known, the rotational part in the representation* 1–3 *is always defined. The translational part is defined when the scale of the translational parts is defined* (*descriptions* $\mathbf{T}, \mathbf{s}$). *If this is not the case* (*descriptions* $\mathbf{t}$, $\mathbf{e}'$), *then the ratio of the scale of the translational parts of* 1–2 *and* 2–3 *is equivalent to the additional information contained in the translational part of* 1–3.

Euclidean Representation. The knowledge of the intrinsic parameters ensures that in the case where the translations are completely determined, the composition relations are obviously obtained as composition of displacements. If the displacements $\mathscr{D}_{12}$ and $\mathscr{D}_{23}$ are known only up to a scale factor (by convention their translational components are taken to be unit vectors: $\mathbf{t}_{12} = \mathbf{T}_{12}/\|\mathbf{T}_{12}\|$ and $\mathbf{t}_{23} = \mathbf{T}_{23}/\|\mathbf{T}_{23}\|$), it is still possible to determine the rotation

$$\mathbf{R}_{13} = \mathbf{T}_{23}\mathbf{R}_{12}, \tag{31}$$

but not the direction of the translation of $\mathscr{D}_{23}\mathscr{D}_{12}$, since there are two additional unknowns in the relation

$$\alpha_1\mathbf{t}_{13} = \mathbf{R}_{23}\mathbf{t}_{12} + \alpha_2\mathbf{t}_{23}, \tag{32}$$

which are $\alpha_1 = \|\mathbf{T}_{13}\|/\|\mathbf{T}_{12}\|$ and $\alpha_2 = \|\mathbf{T}_{23}\|/\|\mathbf{T}_{12}\|$. But if $\mathscr{D}_{12}$, $\mathscr{D}_{23}$, and the ratio α_2 are known, then it is possible to determine $\mathbf{t}_{13}$ and the other ratio. This means that $\mathscr{D}_{12}$ and $\mathscr{D}_{23}$ (resp. $\mathscr{D}_{13}$) being known, the knowledge of $\mathscr{D}_{13}$ (resp. $\mathscr{D}_{23}$) is equivalent to that of α_2 (resp. α_1). When we have three views, there is only a global scale indetermination, but the ratio of the local scale factors is completely determined. This is the basis for a self-calibration approach to recover trinocular structure from three views [34, 38]. Thus, the description is complete if, in addition to the direction of translations, the ratio of their norm is known.

Affine Representation. Let us first point out a general result for infinity homographies between three images, which is easily derived from (15)

$$\mathbf{H}_{\infty 13} = \mathbf{H}_{\infty 23}\mathbf{H}_{\infty 12} \quad \mathbf{Q}_{13} = \mathbf{Q}_{23}\mathbf{Q}_{12}. \tag{33}$$

If the epipoles are determined by the formula (12), then we have, using this formula and Eq. (15),

$$\mathbf{e}_{31} = \mathbf{H}_{\infty 23}\mathbf{e}_{21} + \mathbf{e}_{32}.$$

From what has been seen about the scale factors for the $\mathbf{Qs}$-representation, we can conclude that

$$\mathbf{s}_{31} = \mathbf{Q}_{23}\mathbf{s}_{21} + \mathbf{s}_{32}. \tag{34}$$

The reader may wonder why these projective quantities can be treated as Euclidean values. The reason, already mentioned in Section 4.2 is that when starting from the formulas (15) and (12), scale factors consistent with the three projection matrices are automatically obtained. But if not starting from the projection matrices, the scale factors are undetermined, as expected. That means that we have to normalize the description, if we are to use more than two views, as was done in the case of the similarity-invariant description, as opposed to the displacement-invariant one. Thus, in the description, we must replace $\mathbf{e}_{ij}$ by the normalized epipole $\mathbf{e}_{Nij} = \mathbf{e}_{ij}/\|\mathbf{e}_{ij}\|$. We then see that the reasoning introduced in the preceding paragraph still holds. Thus, a complete description for the affine parameters of three views includes the descriptions between views 1 and 2, 2 and 3, and the ratio $\beta_2 = \|\mathbf{e}_{21}\|/\|\mathbf{e}_{32}\|$, these quantities being defined by (12). It jcan be noted that β_2 can be computed from α_2 and the elements of the similarity invariant description.

It can be seen that the formulas for composition of affine unimodular descriptions[2] $\mathbf{Q}, \mathbf{s}$ (resp. affine descriptions $\mathbf{H}_\infty$, $\mathbf{e}'$) have the same form as formulas for composition of displacements (resp. displacements up to a scale factor). In this sense, they can be thought of as "uncalibrated motion." In particular, $\mathbf{Q}$ and $\mathbf{H}_\infty$ can be interpreted as uncalibrated rotations, $\mathbf{s}$ and $\mathbf{e}'$ as uncalibrated translations (hence, the notation $\mathbf{Q}$, $\mathbf{s}$ which was defined in [67]). The relations $\mathbf{H}_\infty = \mathbf{A}'\mathbf{R}\mathbf{A}^{-1}$ and $\mathbf{e}' = \mathbf{A}'\mathbf{t}$ allows one to determine directly the motion parameters, the rotation $\mathbf{R}$, and the direction of translation $\mathbf{t}$ from the affine description, if the intrinsic parameters are determined. The matrix $\mathbf{H}_\infty$

[2] Note that (34), unlike (33), does not hold with general homographies considered in Section 3.3.

does not depend on the translational component of the displacement, a propriety used by [7, 69] to obtain the rotational component of the displacement between two cameras once the intrinsic parameters are known. The other component of the description, $\mathbf{e}'$, depends only on the translational component.

5.2. *Projective Composition Relations and Trifocal Geometry*

MAIN RESULTS. *The determination of the projective representation* $(\mathbf{S}, \mathbf{e}')$ 1–3 *from the representations* 1–2 *and* 2–3 *requires, in addition to these representations, four parameters, which are a ratio of scale factors and a 3D vector called* $\mathbf{q}$. *In the case of three views, this gives a total of* 18 *parameters of the epipolar geometry of three views, which is described by a geometric construction called trifocal geometry. This construction puts precisely three constraints on the set of fundamental matrices* (21 *parameters*). *In contrast with the representation based on fundamental matrices, the representation* $\mathbf{S}$, $\mathbf{e}'$ *is appropriate to obtain a global representation.*

Some Algebra. The reason why the similarity-invariant description yields more complicated composition relations than the displacement-invariant description is that we have in the first case an indetermination in the description, which is represented by the last element of the canonical similarity matrix, a nonmeasurable quantity. Only the ratios of these quantities can be obtained, using the three descriptions together. We have just seen that the behavior of the affine description is the same as the similarity-invariant one. Now the projective-invariant description in turn has more nonmeasurable variables, which are the last row of this matrix; thus, the composition relations are even more complicated. If we start from the projection matrices, the following relations hold

$$\begin{aligned} \mathbf{e}_{31} &= \mathbf{S}_{23}\mathbf{e}_{21} + \mathbf{e}_{32}(\mathbf{q}_2^T\mathbf{e}_{21}) \\ \mathbf{S}_{13} &= \mathbf{S}_{23}\mathbf{S}_{12} + \mathbf{e}_{32}\left(\mathbf{q}_2^T\mathbf{S}_{12} + \frac{\mathbf{e}_{12}^T}{\|\mathbf{e}_{12}\|^2}\right) + \mathbf{e}_{31}\mathbf{q}_1^T, \end{aligned} \tag{35}$$

where

$$\mathbf{q}_2 = \mathbf{r}_{\infty 23} - \mathbf{r}_{\infty 21}, \mathbf{q}_1 = \mathbf{r}_{\infty 12} - \mathbf{r}_{\infty 13}, \tag{36}$$

the vectors $\mathbf{r}_{\infty ij}$ being the nonmeasurable quantities defined in (17). These relations are proved in Appendix B.

Taking into account the scale factors

$$\beta_1 = \|\mathbf{e}_{31}\|/\|\mathbf{e}_{21}\|, \beta_2 = \|\mathbf{e}_{32}\|/\|\mathbf{e}_{21}\|, \tag{37}$$

we can write these relations

$$\beta_1\mathbf{e}_{N31} = \mathbf{S}_{23}\mathbf{e}_{N21} + \beta_2\mathbf{e}_{N32}(\mathbf{q}_{N2}^T\mathbf{e}_{N21}) \tag{38}$$

$$\begin{aligned} \mathbf{S}_{13} = \mathbf{S}_{23}\mathbf{S}_{12} + \beta_2\mathbf{e}_{N32}(\mathbf{q}_{N2}^T\mathbf{S}_{12} + \gamma_1\mathbf{e}_{N12}^T) \\ + \beta_1\mathbf{e}_{N31}\mathbf{q}_{N1}^T, \end{aligned} \tag{39}$$

where

$$\gamma_1 = \|\mathbf{e}_{21}\|/\|\mathbf{e}_{12}\| \tag{40}$$

$$\mathbf{q}_{N1} = \|\mathbf{e}_{21}\|\mathbf{q}_1 = \mathbf{r}_{\infty N12} - \frac{1}{\beta_1}\mathbf{r}_{\infty N13}$$

$$\mathbf{q}_{N2} = \|\mathbf{e}_{21}\|\mathbf{q}_2 = \frac{1}{\beta_2}\mathbf{r}_{\infty N23} - \gamma_1\mathbf{r}_{\infty N21}, \tag{41}$$

all the normalized vectors (denoted with a subscript N) being defined in (29). Note that the quantities γ_1 and $\mathbf{e}_{12}$ appearing in these equations are not part of the description 1–2; however, they can be derived from it, although there is no simple closed-form formula, by noting that by definition $\mathbf{S}_{12}\mathbf{e}_{12} = 0$, which gives $\mathbf{e}_{N12}$.The ratio γ_1 is computed by assigning an identical value to an identical component in the kernals of $\mathbf{S}_{12}^T$ and of $\mathbf{S}_{12}$. By analogy with the previous geometries where only a ratio of the nonmeasurable quantities (the norms of translations or epipoles) was available, we see that only a difference of nonmeasurable vectors is accessible.

These algebraic considerations show that to describe the epipolar geometry of three views, we need, in addition to $\mathbf{S}_{12}$, $\mathbf{e}_{N21}$ (seven parameters) and $\mathbf{S}_{23}$, $\mathbf{e}_{N32}$ (seven parameters), four additional parameters, a vector $\mathbf{q}$ (3) and a ratio β (1).

Trifocal Geometry. Let us give a geometric interpretation of these results, by considering first the constraints satisfied by the set of fundamental matrices $\mathbf{F}_{12}$, $\mathbf{F}_{23}$, and $\mathbf{F}_{13}$ (7×3 parameters). The fact that the description previously found depends on 18 parameters show that the three fundamental matrices are not independent, but are linked by three constraints. Let us consider $\mathbf{F}_{31}\mathbf{e}_{32}$, the epipolarline of $\mathbf{e}_{32}$ in the first image. This line is the projection of the line $\langle\mathbf{C}_2, \mathbf{C}_3\rangle$ in the first retina, through the optical center $\mathbf{C}_1$. Thus, it is the intersection of the *trifocal plane* $\langle\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3\rangle$ with the first retina. Now the epipoles $\mathbf{e}_{12}$ and $\mathbf{e}_{13}$ also belong to this plane and to the first retina; thus, the epipolar line $\mathbf{e}_{12} \times \mathbf{e}_{13}$ is identical to $\mathbf{F}_{31}\mathbf{e}_{32}$. By a circular permutation of indices follow the three equations, which basically express the fact that the six epipoles must belong to the trifocal plane:

$$\begin{aligned} &\mathbf{F}_{31}\mathbf{e}_{32} = \mathbf{e}_{21} \times \mathbf{e}_{13} \quad \mathbf{F}_{12}\mathbf{e}_{13} = \mathbf{e}_{23} \times \mathbf{e}_{21} \\ &\mathbf{F}_{23}\mathbf{e}_{21} = \mathbf{e}_{31} \times \mathbf{e}_{32}. \end{aligned} \tag{42}$$

The geometric argument presented above, and illustrated

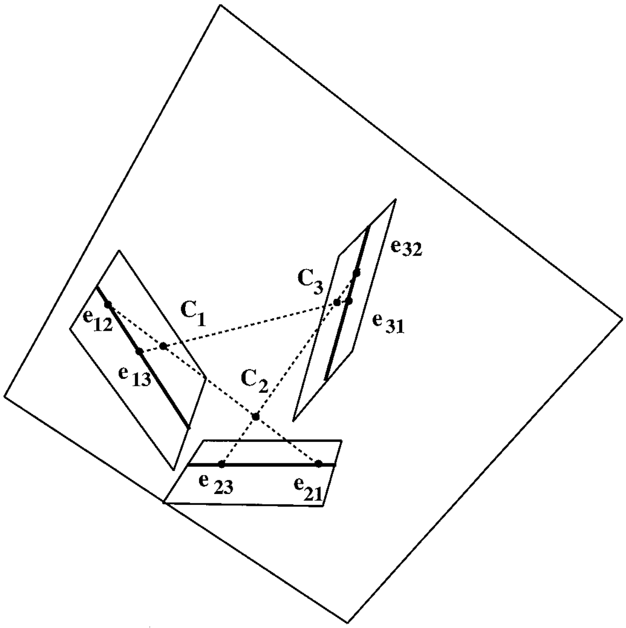


FIG. 4. The trifocal geometry.

in Fig. 4, proves that there are *at most* 18 parameters, as also remarked in [18].

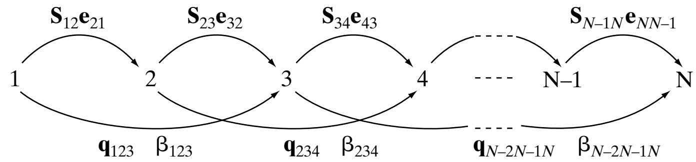
To prove that there are exactly 18 parameters, an explicit construction is required. Such a construction has been done algebraically in this section. A beautiful geometric construction due to Faugeras can be found in [67]. This construction is rather complicated because it also takes into account the geometry of line correspondences. Let us now describe in simpler geometric terms the degrees of freedom of the epipolar geometry of three views. As it was said in Section 3.1, the epipolar geometry of two cameras is determined by the positions of each epipoles and the three coefficients of the homography between epipolar lines. It is clear that the 14 parameters of $\mathbf{F}_{12}$ and $\mathbf{F}_{23}$ are independent. Let us suppose that these two matrices are given and figure out the parameters necessary to represent $\mathbf{F}_{13}$. Because of the trifocal construction detailed previously, the lines $\mathbf{l}_1 = \mathbf{F}_{12}^T\mathbf{e}_{23}$ and $\mathbf{l}_3 = \mathbf{F}_{23}\mathbf{e}_{21}$ are corresponding epipolar lines. It means that:

- The epipole $\mathbf{e}_{13}$ is constrained to lie on $\mathbf{l}_1$. The degree of freedom left is the position on this line.
- The epipole $\mathbf{e}_{31}$ is constrained to lie on $\mathbf{l}_3$. The degree of freedom left is the position on this line.
- $\mathbf{l}_1$ and $\mathbf{l}_3$ are in correspondence through the homography. The two degrees of freedom left are the choice of two other pairs of corresponding epipolar lines.

Therefore, we find again that four additional parameters are needed to account for the epipolar geometry of three views.

Generalization to N Views. Extending the previous algebraic and geometric construction to the case of N views is easily done by induction, and we find thus that the epipolar geometry of N cameras can be described by the $11N - 15$ parameters of the canonic decomposition:

- $\mathbf{S}$-matrices and epipoles between $N-1$ consecutive pairs of views
- $\mathbf{q}$-vectors and ratios of epipole norms β between $N-2$ consecutive triplets of views



It can be seen that on one hand, the lattice of $N(N-1)/2$ $\mathbf{F}$-matrices are *not* independent, and there is, to our knowledge, *no simple formula* to combine them. Moreover, for $N > 3$, the constraints of the form (42) are not independent. There are $(N-2)C_N^2$ of them, and it is easy to see that this number is larger than the number of parameters in the lattice of $\mathbf{F}$-matrices except for $N = 3$.

On the other hand, the $11N - 15$ parameters of the canonic decomposition are independent, and there are simple closed-form formulas to compute any $\mathbf{S}$, $\mathbf{e}$, $\mathbf{q}$, β from the $11N - 15$ parameters.

5.3. *The Canonical Decomposition for Three and More Views*

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 canonical decomposition for three views is defined as the unique representation

$$\tilde{\mathbf{P}}_1 = \mathcal{I}_1\mathcal{T} \quad \tilde{\mathbf{P}}_2 = \mathcal{I}_2\mathcal{T} \quad \tilde{\mathbf{P}}_3 = \mathcal{I}_3\mathcal{T} \tag{43}$$

where $\mathcal{I}_1$ and $\mathcal{I}_2$ have the same form as in the canonic decomposition for two views (27). 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 canonical 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. From the count of parameters, it is seen that the triple of descriptions for two views is not a minimal representation.

In order to make a purely algebraic derivation for the last points, we notice that simultaneously to (43), the following canonical representations of two views must hold

$$\tilde{\mathbf{P}}_1 = \mathscr{I}_1'\mathscr{T}' \quad \tilde{\mathbf{P}}_3 = \mathscr{I}_2'\mathscr{T}' \tag{44}$$

$$\tilde{\mathbf{P}}_2 = \mathscr{I}_1''\mathscr{T}'' \quad \tilde{\mathbf{P}}_3 = \mathscr{I}_2''\mathscr{T}'', \tag{45}$$

where all the quantities with the same subscript must be of the same form. It is again quite easy to make the derivation for the affine case (the form of the Euclidean case is similar), and this is done in Appendix A. The work of [23] for instance fits into our formalism in the projective case and illustrates its computational advantages.

We have summarized in Table 3 the results specific to the canonical 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 canonical decomposition to the case of N views is straightforward. Thus, the elements of the description are exactly the same as for three views and can be summarized in Table 4 where it can be verified that the total number of parameters is always $11N$.

5.4. *From Local to Global Representations*

Main Result. *The three-view invariant (global) description which has been described previously, can be computed from triples of two views invariant (local) descriptions using closed forms solutions, which resolve scale factors indeterminations.*

Euclidean Representation. First, there is no ambiguity for the rotational part. For the translational part, if we take by definition $\mathbf{t}_{ij} = \mathbf{T}_{ij}/\|\mathbf{T}_{ij}\|$, as previously done, then there is no sign ambiguity. But in the structure from motion paradigm, the sign information contained in $\mathbf{T}_{ij}$ is also lost in the direction of translation $\mathbf{t}_{ij}$, a fact we take into account by writing that the recovered quantities are $\mathbf{t}_{12} = \varepsilon_{12}\mathbf{T}_{12}/\|\mathbf{T}_{12}\|$ and $\mathbf{t}_{23} = \varepsilon_{23}\mathbf{T}_{23}/\|\mathbf{T}_{23}\|$. In the relation (32), the definition of the scale factors is thus $\alpha_1 = \varepsilon_{12}\varepsilon_{13}\|\mathbf{T}_{13}\|/\|\mathbf{T}_{12}\|$ and $\alpha_2 = \varepsilon_{12}\varepsilon_{23}\|\mathbf{T}_{23}\|/\|\mathbf{T}_{12}\|$.

However, the relation (32) constraints $\mathbf{t}_{13}$ to lie in the plane $\langle \mathbf{R}_{23}\mathbf{t}_{12}, \mathbf{t}_{23}\rangle$, and thus if $\mathbf{t}_{12}$, $\mathbf{t}_{23}$, and $\mathbf{t}_{13}$ are all known, the ratio α_2 can be computed by expressing the proportionality constraint

$$\mathbf{t}_{13} \times (\mathbf{R}_{23}\mathbf{t}_{12} + \alpha_2\mathbf{t}_{23}) = 0. \tag{46}$$

Once α_2 is determined, α_1 can also be computed. We obtain

$$\begin{aligned} \alpha_2 &= -\frac{\mathbf{t}_{13} \times \mathbf{R}_{23}\mathbf{t}_{12}}{\mathbf{t}_{13} \times \mathbf{t}_{23}} \quad \alpha_1 = \frac{\mathbf{R}_{23}\mathbf{t}_{12} + \alpha_2\mathbf{t}_{23}}{\mathbf{t}_{13}} \\ &= -\frac{\mathbf{t}_{23} \times \mathbf{R}_{13}\mathbf{R}_{21}^{-1}\mathbf{t}_{12}}{\mathbf{t}_{13} \times \mathbf{t}_{23}} \quad \text{for } i = 1, 2, 3, \end{aligned} \tag{47}$$

where we have used the usual division symbol for the vector division of two proportional vectors: if $\mathbf{v}_1 = \lambda\mathbf{v}_2$, then $\mathbf{v}_1/\mathbf{v}_2 = \lambda$. In the presence of noise, the two vectors would not be exactly proportional, and thus the vector division would have to be replaced by the least-squares formulation: $\min_\lambda \|\mathbf{v}_1 - \lambda\mathbf{v}_2\|$. Please note that it would be incorrect to consider only the quotient of the norms, since the signs would be lost.

Affine Representation. Using the fact that the composition laws are similar, and applying the same technique, it is found that $\beta_1\mathbf{e}_{N31} = \mathbf{H}_{\infty 23}\mathbf{e}_{N21} + \beta_2\mathbf{e}_{N32}$ with

$$\begin{aligned} \beta_2 &= -\frac{\mathbf{e}_{N31} \times \mathbf{H}_{\infty 23}\mathbf{e}_{N21}}{\mathbf{e}_{N31} \times \mathbf{e}_{N32}} \quad \beta_1 = \frac{\mathbf{e}_{N32} \times \mathbf{H}_{\infty 13}\mathbf{H}_{\infty 12}^{-1}\mathbf{e}_{N21}}{\mathbf{e}_{N31} \times \mathbf{e}_{N32}} \\ &\text{for } i = 1, 2, 3. \end{aligned} \tag{48}$$

In addition to the indetermination on the epipole, we must take into account the fact that in the affine description for two views, the infinity homography is defined only up to a scale factor. It means that we know only the matrices $\lambda_{12}\mathbf{H}_{\infty 12}$, $\lambda_{23}\mathbf{H}_{\infty 23}$, $\lambda_{13}\mathbf{H}_{\infty 13}$. Expliciting also the dependencies on the scale factors which are in (48), the invariant descriptions become

$$\begin{cases} \mathscr{I}_1 = [\mathbf{I}_3, 0] \\ \mathscr{I}_2 = [\lambda_{12}\mathbf{H}_{\infty 12}, \mathbf{e}_{N21}] \\ \mathscr{I}_3 = [\lambda_{13}\mathbf{H}_{\infty 13}, \lambda_{13}\lambda_{12}^{-1}\beta_1\mathbf{e}_{31N}] \end{cases}$$

TABLE 3
The Geometries of Three Views: Canonical Representation

EUCLIDEAN	$\mathcal{D} \in \mathcal{SE}_3$	Displacement	6
	$\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$	Intrinsic parameters	5 + 5 + 5
Invariant descriptions	$\begin{array}{c\|c} \mathbf{R}_{12}, \mathbf{R}_{23} & \mathbf{R}_{12}, \mathbf{R}_{13} \\ \mathrm{T}_{12}, T_{23} & \mathbf{T}_{12}, \mathbf{T}_{13} \end{array}$	Rotations	3 + 3
		Translations	3 + 3
Canonic decomposition	$\begin{cases} \tilde{\mathbf{P}}_1 = \mathbf{A}_1[\mathbf{I}_3, 0]\mathcal{D} \\ \tilde{\mathbf{P}}_2 = \mathbf{A}_2[\mathbf{R}_{12}, \mathbf{T}_{12}]\mathcal{D} \\ \tilde{\mathbf{P}}_3 = \begin{array}{l} \mathbf{A}_3[\mathbf{R}_{13}, \mathbf{T}_{13}]\mathcal{D} \\ \mathbf{A}_3[\mathbf{R}_{23}\mathbf{R}_{12}, \mathbf{R}_{23}\mathbf{T}_{12} + \mathbf{T}_{23}]\mathcal{D} \end{array} \end{cases}$		
EUCLIDEAN	$\mathcal{S} \in \mathcal{GE}_3$	Similarity	7
	$\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$	Intrinsic parameters	5 + 5 + 5
Invariant descriptions	$\begin{array}{c\|c} \mathbf{R}_{12}, \mathbf{R}_{23} & \mathbf{R}_{12}, \mathbf{R}_{13} \\ \mathbf{t}_{12}, \mathbf{t}_{23} & \mathbf{t}_{12}, \mathbf{t}_{13} \\ \hline \alpha_1 & \alpha_2 \end{array}$	Rotations	3 + 3
		Directions of translations	2 + 2
		Ratios of translation norms	1
Canonic decomposition	$\begin{cases} \tilde{\mathbf{P}}_1 = \mathbf{A}_1[\mathbf{I}_3, 0]\mathcal{S} \\ \tilde{\mathbf{P}}_2 = \mathbf{A}_2[\mathbf{R}_{12}, \mathbf{T}_{12}]\mathcal{S} \\ \tilde{\mathbf{P}}_3 = \begin{array}{l} \mathbf{A}_3[\mathbf{R}_{13}, \alpha_1\mathbf{t}_{13}]\mathcal{S} \\ \mathbf{A}_3[\mathbf{R}_{23}\mathbf{R}_{12}, \mathbf{R}_{23}\mathbf{t}_{12} + \alpha_2\mathbf{t}_{23}]\mathcal{S} \end{array} \end{cases}$	$\alpha_1 = \|\mathbf{T}_{13}\|/\|\mathbf{T}_{12}\|$ $\alpha_2 = \|\mathbf{T}_{23}\|/\|\mathbf{T}_{12}\|$	
AFFINE	$\mathcal{U} \in \mathcal{SA}_3$	Unimodular transformation	11
Invariant descriptions	$\begin{array}{c\|c} \mathbf{Q}_{12}, \mathbf{Q}_{23} & \mathbf{Q}_{12}, \mathbf{R}_{13} \\ \mathbf{s}_{21}, \mathbf{s}_{32} & \mathbf{s}_{21}, \mathbf{s}_{31} \end{array}$	Scaled infinity homographies	9 + 9 (8 + 8)
		Normalized epipoles	2 + 2 (3 + 3)
Canonic decomposition	$\begin{cases} \tilde{\mathbf{P}}_1 = [\mathbf{I}_3, 0]\mathcal{U} \\ \tilde{\mathbf{P}}_2 = [\mathbf{Q}_{12}, \mathbf{s}_{21}]\mathcal{U} \\ \tilde{\mathbf{P}}_3 = \begin{array}{l} [\mathbf{Q}_{13}, \mathbf{s}_{31}]\mathcal{U} \\ [\mathbf{Q}_{23}\mathbf{Q}_{12}, \mathbf{Q}_{23}\mathbf{s}_{21} + \mathbf{s}_{32}]\mathcal{U} \end{array} \end{cases}$		
AFFINE	$\mathcal{A} \in \mathcal{GA}_3$	Affine transformation	12
Invariant descriptions	$\begin{array}{c\|c} \mathbf{H}_{\infty 12}, \mathbf{H}_{\infty 23} & \mathbf{H}_{\infty 12}, \mathbf{H}_{\infty 13} \\ \mathbf{e}_{N21}, \mathbf{e}_{N32} & \mathbf{e}_{N21}, \mathbf{e}_{N31} \\ \hline \beta_1 & \beta_2 \end{array}$	Infinity homographies	8 + 8
		Normalized epipoles	2 + 2
		Ratios of epipole norms	1
Canonic decomposition	$\begin{cases} \tilde{\mathbf{P}}_1 = [\mathbf{I}_3, 0]\mathcal{A} \\ \tilde{\mathbf{P}}_2 = [\mathbf{H}_{\infty 12}, \mathbf{e}_{N21}]\mathcal{A} \\ \tilde{\mathbf{P}}_3 = \begin{array}{l} [\mathbf{H}_{\infty 13}, \beta_1\mathbf{e}_{31N}]\mathcal{A} \\ [\mathbf{H}_{\infty 23}\mathbf{H}_{\infty 12}, \mathbf{H}_{\infty 23}\mathbf{e}_{N21} + \beta_2\mathbf{e}_{N32}]\mathcal{A} \end{array} \end{cases}$	$\beta_1 = \|\mathbf{e}_{31}\|/\|\mathbf{e}_{21}\| = \alpha_1\|\mathbf{A}_3\mathbf{t}_{13}\|/\|\mathbf{A}_2\mathbf{t}_{12}\|$ $\beta_2 = \|\mathbf{e}_{32}\|/\|\mathbf{e}_{21}\| = \alpha_2\|\mathbf{A}_3\mathbf{t}_{23}\|/\|\mathbf{A}_2\mathbf{t}_{12}\|$	
PROJECTIVE	$\mathcal{H} \in \mathcal{GL}_4$	Homography	15
Invariant descriptions	$\begin{array}{c\|c} \mathbf{S}_{12}, \mathbf{S}_{23} & \mathbf{S}_{12}, \mathbf{S}_{13} \\ \mathbf{e}_{N21}, \mathbf{e}_{N32} & \mathbf{e}_{N21}, \mathbf{e}_{N31} \\ \hline \mathbf{q}_{N1} & \mathbf{q}_{N2} \\ \beta_1 & \beta_2 \end{array}$	Epipolar projections	5 + 5
		Normalized epipoles	2 + 2
		Differences of $\mathbf{r}_\infty$-vectors	3
		Ratios of epipole norms	1
Canonic decomposition	$\begin{cases} \tilde{\mathbf{P}}_1 = [\mathbf{I}_3, 0]\mathcal{H} \\ \tilde{\mathbf{P}}_2 = [\mathbf{S}_{12}, \mathbf{e}_{N21}]\mathcal{H} \\ \qquad [\mathbf{S}_{13} - \beta_1\mathbf{e}_{31N}\mathbf{q}_{N1}^T, \beta_1\mathbf{e}_{31N}]\mathcal{H} \\ \tilde{\mathbf{P}}_3 = [\mathbf{S}_{23}\mathbf{S}_{12} + \beta_2\mathbf{e}_{N32}(\mathbf{q}_{N2}^T\mathbf{S}_{12} + \gamma_1\mathbf{e}_{N12}^T), \\ \qquad\qquad \mathbf{S}_{23}\mathbf{e}_{N21} + \beta_2\mathbf{e}_{N32}(\mathbf{q}_{N2}^T\mathbf{e}_{N21})]\mathcal{H} \end{cases}$	$\mathbf{q}_{N1} = \mathbf{H}_{\infty 12}^T\mathbf{e}_{N21} - \frac{1}{\beta_1}\mathbf{H}_{\infty 13}^T\mathbf{e}_{N31}$ $\mathbf{q}_{N2} = \frac{1}{\beta_2}\mathbf{H}_{\infty 23}^T\mathbf{e}_{N32} - \gamma_1\mathbf{H}_{\infty 21}^T\mathbf{e}_{N12}$ $\beta_1 = \|\mathbf{e}_{31}\|/\|\mathbf{e}_{21}\|$ $\beta_2 = \|\mathbf{e}_{32}\|/\|\mathbf{e}_{21}\|$ with $\gamma_1 = \|\mathbf{e}_{21}\|/\|\mathbf{e}_{12}\|$	

Note. 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 years, and 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 are mentioned.

$$\begin{cases} \mathcal{I}_1 = [\mathbf{I}_3, 0] \\ \mathcal{I}_2 = [\lambda_{12}\mathbf{H}_{\infty 12}, \mathbf{e}_{N21}] \\ \mathcal{I}_3 = [\lambda_{12}\lambda_{23}\mathbf{H}_{\infty 23}\mathbf{H}_{\infty 12}, \lambda_{23}\mathbf{H}_{\infty 23}\mathbf{e}_{N21} + \lambda_{23}\beta_2\mathbf{e}_{N32}]. \end{cases}$$

Let us remark that the pairs of projection matrices which yield the same $\mathbf{H}_\infty$, $\mathbf{e}'$ description as $\{[\mathbf{P}, \mathbf{p}], [\mathbf{P}', \mathbf{p}']\}$ are of the form $\{[\lambda\mu\mathbf{P}, \lambda\mathbf{p}], [\lambda'\mu\mathbf{P}', \lambda'\mathbf{p}']\}$. Then we see that both the invariant descriptions are not affected by a change of the unknown scale factors λ_{12}, λ_{23}, λ_{13}. Note that this

TABLE 4
The Geometries of N Views: Canonical Representation

Euclidean	Displacement	6
	Intrinsic parameters	$5N$
	Rotations	$3\ (N-1)$
	Translations	$3\ (N-1)$
	Similarity	7
	Intrinsic parameters	$5N$
	Rotations	$3\ (N-1)$
	Directions of translations	$2\ (N-1)$
	Ratio of translation norms	$N-2$
Affine	Affine unimodular transformation	11
	Qs-representation	$11\ (N-1)$
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 $\mathbf{r}_\infty$ vectors	$3\ (N-2)$
	Ratios of epipole norms	$N-2$

result depends on the way the ratios β_i are computed. If, for example, the ratio β_1 was computed from $\mathbf{H}_{\infty 23}$ instead of $\mathbf{H}_{\infty 13}\mathbf{H}_{\infty 12}^{-1}$, the result would not hold.

Projective Representation. Let us first deal with the first canonical decomposition, and compute β_1 and $\mathbf{q}_{N1}$ from the pairs of descriptions 1–2, 1–3, and 2–3. The following pairs of projection matrices must yield the same projective view invariants:

$$\begin{cases}[\mathbf{I}_3, 0] \\ [\mathbf{S}_{32}, \mathbf{e}_{N23}]\end{cases} \quad \begin{cases}\left[\dfrac{1}{\beta_1}\mathbf{S}_{13} - \mathbf{e}_{N31}\mathbf{q}_{N1}^T, \mathbf{e}_{N31}\right] \\ [\mathbf{S}_{12}, \mathbf{e}_{N21}].\end{cases}$$

Writing that the fundamental matrices and epipoles obtained from the two pairs are proportional yields the two equations

$$\begin{gathered}([\mathbf{e}_{N23}]_\times \mathbf{S}_{32})(\mathbf{S}_{13} - \mathbf{e}_{N31}\beta_1\mathbf{q}_{N1}^T) \sim ([\mathbf{e}_{N23}]_\times \mathbf{S}_{12}) \\ \mathbf{e}_{N23} \times (\mathbf{e}_{N21} - \beta_1\mathbf{S}_{12}(\mathbf{S}_{13} - \mathbf{e}_{N31}\beta_1\mathbf{q}_{N1}^T)^{-1}\mathbf{e}_{N31}) = 0.\end{gathered}$$

The solution is obtained by

$$\begin{aligned}\beta_1(\mathbf{q}_{N1})_k &= \frac{([\mathbf{e}_{N23}]_\times \mathbf{S}_{32}\mathbf{S}_{13})_k \times ([\mathbf{e}_{N23}]_\times \mathbf{S}_{12})_k}{([\mathbf{e}_{N23}]_\times \mathbf{S}_{32}\mathbf{e}_{N31}) \times ([\mathbf{e}_{N23}]_\times \mathbf{S}_{12})_k}, \\ \beta_1 &= \frac{\mathbf{e}_{N23} \times \mathbf{e}_{N21}}{\mathbf{e}_{N23} \times \mathbf{S}_{12}(\mathbf{S}_{13} - \mathbf{e}_{N31}(\beta_1\mathbf{q}_{N1})^T)^{-1}\mathbf{e}_{N31}},\end{aligned} \tag{49}$$

where the notation $(\mathbf{M})_k$ (resp. $(\mathbf{v})_k$) designates the kth column vector (resp. component) of a matrix (resp. vector). From the formulas given above, it is easy to see that the solution is not affected by a scaling of the matrices $\mathbf{S}_{ij}$, since β_1 is proportional to λ_{13}^{-1} and $\mathbf{q}_{N1}$ to λ_{13}^2.

The second canonical decomposition is dealt with in the same way, by expressing the projective equivalence of the pairs of projection matrices

$$\begin{cases}[\mathbf{I}_3, 0] \\ [\mathbf{S}_{31}, \mathbf{e}_{N13}]\end{cases} \quad \begin{cases}[\mathbf{S}_{23}\mathbf{S}_{12} + \beta_2\mathbf{e}_{N32}(\mathbf{q}_{N2}^T\mathbf{S}_{12} + \lambda_1\mathbf{e}_{N12}^T), \\ \qquad \mathbf{S}_{23}\mathbf{e}_{N21} + \beta_2\mathbf{e}_{N32}(\mathbf{q}_{N2}^T\mathbf{e}_{N21})] \\ [\mathbf{I}_3, 0].\end{cases}$$

In writing the proportionality of the fundamental matrices, the term with λ_1 cancels because of (42), and we are left with the equation

$$[\mathbf{e}_{N13}]_\times \mathbf{S}_{31}(\mathbf{S}_{23}\mathbf{S}_{12} + \mathbf{e}_{N32}(\beta_2\mathbf{q}_{N2}^T)\mathbf{S}_{12}) \sim [\mathbf{e}_{N13}]_\times,$$

which can be solved similarly for $\beta_2\mathbf{q}_{N2}$. Then β_2 is determined by solving

$$\begin{gathered}(\mathbf{S}_{23}\mathbf{S}_{12} + \mathbf{e}_{N32}((\beta_2\mathbf{q}_{N2})^T\mathbf{S}_{12} + \beta_2\gamma_1\mathbf{e}_{N12}^T))\mathbf{e}_{N13} \\ \times (\mathbf{S}_{23}\mathbf{e}_{N21} + \mathbf{e}_{N32}((\beta_2\mathbf{q}_{N2})^T\mathbf{e}_{N21})) = 0.\end{gathered}$$

We have just given a sketch of the solution, and have not explicited it, because in practice, one would prefer to deal with the first equivalent canonical decomposition, which is less complicated. It can be noted that the simpler formulas[3]

$$\begin{aligned}\beta_2 &= \frac{\mathbf{e}_{N31} \times \mathbf{S}_{13}\mathbf{e}_{N12}}{\gamma_1\mathbf{e}_{N31} \times \mathbf{e}_{N32}} \\ \beta_2\mathbf{q}_{N2} &= \gamma_1 \, \frac{\mathbf{S}_{21}^T\mathbf{S}_{13}^T\mathbf{e}_{N32} + \mathbf{e}_{N32}^T\mathbf{e}_{N31}\mathbf{S}_{23}^T\mathbf{e}_{N31}}{1 - (\mathbf{e}_{N32}^T\mathbf{e}_{N31})^2}\end{aligned}$$

do not give consistent solutions when the scale of the matrices $\mathbf{S}_{ij}$ is changed.

6. RELATIONS BETWEEN THE LEVELS OF DESCRIPTIONS

The canonical decomposition has given a local and a global description at different levels. This description is a hierarchy, in the sense that descriptions of the projective level are included in descriptions of the affine level, and descriptions of the affine level are included in descriptions of the Euclidean level. This fact was already apparent in Tables 2 and 3, and some of the relations have already

[3] The first one is obtained by equating the two first 3×3 submatrix of the two forms of the canonical decomposition, then taking the cross product with $\mathbf{e}_{31}$ and the dot product with $\mathbf{e}_{N12}$, the second is from [67].

been discussed previously. The goal of this section is to detail in a more systematic manner the relations between descriptions of different levels.

In Section 6.1, we first describe the hierarchy in terms of different ambiguities of reconstruction in *object space*, each of them corresponding to a given level. In Section 6.2, we illustrate the stratification by applying it to *geometric relations* described by the fundamental matrix and the homography matrices. We then discuss in more detail, using both the local context and the global context, the relations between the affine representation, the projective representation (Section 6.3), and the Euclidean representation (Section 6.4).

6.1. *The Action on Object Space*

From Tables 2 and 3 we remark that each invariant description of a given level is formulated in terms of descriptions of the previous level. This heritage of descriptors is quite natural, except for the link between the similarity-invariant representations and the affine representations where a different change of retinal coordinates occur. It can be seen that the simplest representation is the affine one.

This hierarchy is naturally described as a composition of transformations in object space, by remarking that the general homography matrix of $\mathscr{P}^3$ which appears in the projective description can be written

$$\mathscr{H} = \underbrace{\begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{r}_\infty^T & 1 \end{bmatrix}}_{\mathscr{T}_\mathscr{P}} \underbrace{\begin{bmatrix} \mathbf{A} & \mathbf{0}_3 \\ \mathbf{0}_3^T & 1 \end{bmatrix}}_{\mathscr{T}_\mathscr{A}} \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0}_3^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3^T & \lambda \end{bmatrix}. \tag{50}$$

The inverse homography is applied to object space during the change of projective coordinates used to obtain the canonical representation.

- The matrix $\mathscr{T}_\mathscr{P}$ represents the transformation which moves the plane at infinity, the preimage of which is the plane $[\mathbf{r}_\infty^T, 1]^T$. It is a projective transformation. If we apply it to the canonical projective description, we obtain the affine description. Note that this transformation is the action $\mathscr{T}$ which sends one affine canonical element to the canonical element of another orbit, and it is also found by writing that the image of the pair $\{[\mathbf{I}_3, 0], [\mathbf{H}_\infty, \mathbf{e}]\}$ by $\mathscr{T}$ has the form $\{[\mathbf{I}_3, 0], [\mathbf{H}'_\infty, \mathbf{e}']\}$.
- The matrix $\mathscr{T}_\mathscr{A}$ represents the transformation which moves the absolute conic, the preimage of which is the conic of matrix $\mathbf{A}\mathbf{A}^T$. It is an affine transformation. If we apply it to the canonical affine description, we obtain the Euclidean description up to a scale factor. The same remark about the orbits apply.
- The third and fourth matrices represent the change of Euclidean coordinate system and the scale factor.

Note that the canonical decomposition shows clearly that to obtain an Euclidean representation does *not* require the knowledge of both intrinsic parameters matrices $\mathbf{A}$ and $\mathbf{A}'$, as a naive analysis would lead to. Instead, we need only the eight parameters $\mathbf{A}$ and $\mathbf{r}_\infty$. The fact that the extra two parameters are not required is consistent with the fact that once the $\mathbf{F}$-matrix is known, there are two constraints involving the entries of $\mathbf{A}$ and $\mathbf{A}'$, the so-called two Kruppa equations [16, 41].

6.2. *Stratification of Geometric Descriptors*

As an application of the canonical decomposition for two views, all the geometric objects can be described at a projective, affine, or Euclidean level, in a stratified representation. The table below summarizes the discussion of this subsection by showing the stratification applied to fundamental matrices, homography matrices, and representations of plane structure.

Level	$\mathbf{F}$	$\mathbf{H}$	Π
Projective	$[\mathbf{e}']_\times \mathbf{S}$	$\mathbf{S} + \mathbf{e}'\mathbf{r}_\Pi^T$	$\mathbf{r}_\Pi$
Affine	$[\mathbf{e}']_\times \mathbf{H}_\infty$	$\mathbf{H}_\infty + \mathbf{e}'\nu_\Pi^T$	$\nu_\Pi = \mathbf{r}_\Pi - \mathbf{r}_\infty$
Euclidean	$\mathbf{A}'^{-1T}[\mathbf{T}]_\times \mathbf{R}\mathbf{A}^{-1}$	$\mathbf{A}'(\mathbf{R} + \mathbf{T}\mathbf{n}^T/d)\mathbf{A}^{-1}$	$\mathbf{n}/d = \mathbf{A}^T\nu_\Pi$

Descriptions for Fundamental and Homography Matrices. Let us describe the rigid displacement between the two cameras normalized coordinate system by the rotation $\mathbf{R}$ and the translation $\mathbf{T}$ such that

$$Z'\mathbf{m}' = Z\mathbf{R}\mathbf{m} + \mathbf{T}, \tag{51}$$

where $\mathbf{m}$ is the image of a point $\mathbf{M}$ in the first retina, and Z its depth such that $Z\mathbf{m} = \mathbf{M}$, and the same in the second retina with primes.

In [16, 37], it is pointed to that relation (9) is a generalization of the Longuet-Higgins equation [31], and thus that there is the following relation between the fundamental matrix and the essential matrix

$$\mathbf{F} = \mathbf{A}'^{-1T}\mathbf{E}\mathbf{A}^{-1}, \tag{52}$$

where the essential matrix, a Euclidean quantity, is defined as

$$\mathbf{E} = [\mathbf{T}]_\times \mathbf{R}. \tag{53}$$

A consequence of (52) is that the second epipole is such that $[\mathbf{T}]_\times \mathbf{A}'^{-1}\mathbf{e}' = 0$, and thus $\mathbf{e}' \sim \mathbf{A}'\mathbf{T}$.

The homography matrix can be similarly decomposed using intrinsic parameters. Let the equation of plane Π be $\mathbf{n}^T\mathbf{M} = d$, where $\mathbf{n}$ is the unit normal vector of the plane and d the distance of the plane to the origin. Combining this equation with (51) shows that, if $\mathbf{M}$ belongs to Π, we have the following classical relation [13, 64] using normalized coordinate systems:

$$Z'\mathbf{m}' = Z\left(\mathbf{R} + \frac{1}{d}\mathbf{T}\mathbf{n}^T\right)\mathbf{m}.$$

If pixel coordinates are used, we must introduce the intrinsic parameters, and we eventually find that the projective coordinates of the points $\mathbf{m}$ and $\mathbf{m}'$ are related by the homography

$$\mathbf{H} = \mathbf{A}'\left(\mathbf{R} + \frac{1}{d}\mathbf{t}\mathbf{n}^T\right)\mathbf{A}^{-1}. \tag{54}$$

The infinity homography matrix $\mathbf{H}_\infty = \mathbf{A}'\mathbf{R}\mathbf{A}^{-1}$, is obtained for $d \to \infty$. It is seen that this matrix is obtained as a limit of the homography matrix of any plane, when the distance of this plane to the origin becomes close to ∞.

Structural Parameters of a Plane. We have already noticed in (25) that if $\mathbf{S}$ and $\mathbf{e}'$ are considered as projective motion parameters, then the projective structure parameters of the plane Π is given by the relative representation vector $\mathbf{r}_\Pi$.

Let us now note that Eq. (54) can also be written

$$\mathbf{H} = \mathbf{H}_\infty + \mathbf{e}'\nu_\Pi^T \quad \text{where } \nu_\Pi = \frac{1}{d}\mathbf{A}^{-1T}\mathbf{n}. \tag{55}$$

This formulation separates again the *affine* description of the structure of Π, ν_Π, and the motion parameters $\mathbf{H}_\infty$ and $\mathbf{e}'$ which describe the affine relation between the two image frames [67, 68].

The same remark applies to Eq. (54), where the structure parameters, related to the plane, are $\mathbf{n}$ and d, and the motion parameters $\mathbf{R}$ and $\mathbf{T}$ describe the displacement between the two cameras in the first camera normalized coordinate system.

Moreover, by combining the previous equations, we obtain the hierarchy of representations of a plane

$$\mathbf{r}_\Pi = \mathbf{r}_\infty + \mathbf{A}^{-1T}\mathbf{n}/d.$$

It is seen that starting from the projective representation $\mathbf{r}_\Pi$, the knowledge of the $\mathbf{r}_\infty$ vector, containing affine information, allows us to get ν_Π, and that further knowledge of $\mathbf{A}$, containing Euclidean information, allows us to get $\mathbf{n}$ and d.

6.3. *Projective Description and Affine Description*

MAIN RESULT. *In an image sequence where the global projective representation is known, the knowledge of the global affine representation (and hence the affine information between all the pairs of images) is equivalent to the local identification of affine information in one pair of images.*

Let us examine the case of three views. The canonical decomposition shows that from the two fundamental matrices $\mathbf{F}_{12}$ and $\mathbf{F}_{23}$ and the homography matrices $\mathbf{H}_{\infty 12}$ and $\mathbf{H}_{\infty 23}$, it is not possible to obtain the fundamental matrix $\mathbf{F}_{13}$, since the scale ratio β_2 is also needed.

Let us now see what can be done when the three fundamental matrices (or equivalently, the global projective representation) are known. In order to cope with the problem of scale factors, let us work with the nonnormalized quantities, and establish the interesting relation

$$\begin{aligned}\mathbf{F}_{13} &= \mathbf{H}'_{\infty 23}\mathbf{F}_{12} + \mathbf{F}_{23}\mathbf{H}_{\infty 12} = (\det \mathbf{H}_{\infty 23})(\mathbf{H}_{\infty 32}^T\mathbf{F}_{12} + \mathbf{F}_{32}^T\mathbf{H}_{\infty 12}) \\ &= (\det \mathbf{H}_{\infty 23})(\mathbf{H}_{\infty 32}^T([\mathbf{e}_{21}]_\times - [\mathbf{e}_{23}]_\times)\mathbf{H}_{\infty 12}).\end{aligned} \tag{56}$$

Applying the property, proved for example in [18],

$$\forall \mathbf{M} \in \mathscr{GL}_3, \forall \mathbf{x}, \mathbf{y} \in \mathscr{R}^3, \mathbf{M}^*[\mathbf{x}]_\times\mathbf{y} = [\mathbf{M}\mathbf{x}]_\times\mathbf{M}\mathbf{y} \tag{57}$$

and then (34) and (33) yields the first equality of (56):

$$\begin{aligned}\mathbf{H}_{\infty 23}^*\mathbf{F}_{12} &= \mathbf{H}_{\infty 23}^*[\mathbf{e}_{21}]_\times\mathbf{H}_{\infty 12} = [\mathbf{e}_{31} - \mathbf{e}_{32}]_\times\mathbf{H}_{\infty 23}\mathbf{H}_{\infty 12} \\ &= \underbrace{[\mathbf{e}_{31}]_\times\mathbf{H}_{\infty 13}}_{\mathbf{F}_{13}} - \underbrace{[\mathbf{e}_{32}]_\times\mathbf{H}_{\infty 23}}_{\mathbf{F}_{23}}\mathbf{H}_{\infty 12}.\end{aligned}$$

To obtain the second equality, we also use (30). The last member is obtained by substitution of the decompositions of $\mathbf{F}_{12}$ and $\mathbf{F}_{32}$. It allows us to write the following expression using normalized quantities

$$\mathbf{F}_{13} \sim \mathbf{H}_{\infty 32}^T(\gamma_2[\mathbf{e}_{N21}]_\times - \beta_2[\mathbf{e}_{N23}]_\times)\mathbf{H}_{\infty 12}, \tag{58}$$

where $\gamma_2 = \|\mathbf{e}_{32}\|/\|\mathbf{e}_{23}\|$ is obtained in the same way as the ratio γ_1 appearing in (39).

It can be noted that the system of equations obtained by writing (58) between the three images cannot determine the infinity homography matrices from the knowledge of the three fundamental matrices, because there are 21 parameters in the affine description, versus only 18 in the projective description. However, it can be seen that this system determines the infinity homography matrices as soon as one of them is known. By substitution of $\mathbf{H}_{\infty 32} = \mathbf{S}_{32} + \mathbf{e}_{N23}\mathbf{r}_{\infty 32}^T$ into (58), and some easy algebra, the following formula is obtained,

$$(\mathbf{r}_{\infty 32})_k = \frac{(\mathbf{H}_{\infty 12}^T(\beta_2[\mathbf{e}_{N23}]_\times - \gamma_2[\mathbf{e}_{N21}]_\times)\mathbf{S}_{32})_k \times (\mathbf{F}_{31})_k}{\gamma_2(\mathbf{H}_{\infty 12}^T[\mathbf{e}_{N21}]_{i\times}\mathbf{e}_{N23}) \times (\mathbf{F}_{31})_k} \tag{59}$$

where the notations are the same as in (49). The additional knowledge needed corresponds to one of the three vectors

$\mathbf{r}_\infty$ defined in (17), which identify the plane at infinity. This means that if this quantity is identified anywhere in an image sequence, it can be *propagated* along the whole sequence.

6.4. *Euclidean Description and Affine Description*

MAIN RESULTS. *The knowledge of $\mathbf{H}_\infty$ between two views puts five linear constraints on the Kruppa matrices of each of the views. Therefore, in an image sequence where the global affine representation is known, the knowledge of the intrinsic parameters of all cameras (or the global Euclidean representation) is equivalent to the identification of the intrinsic parameters of one camera in the sequence.*

The relation between the affine description and the Euclidean motion has been discussed already in Section 5. In examining the relation with the intrinsic parameters, the key relation is

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

In the case of self-calibration in the projective framework, Eq. (52) was used to constrain the intrinsic parameters from the fundamental matrix by means of the rigidity constraint expressing that the matrix

$$\mathbf{E} = \mathbf{A}'^{-1T}\mathbf{F}\mathbf{A}^{-1} \tag{61}$$

is an essential matrix [31] (product of an antisymmetric matrix by a rotation matrix; for several characterizations, see [27, 40]), thus eliminating $\mathbf{R}$ and $\mathbf{t}$ [34]. Now, by analogy, if we eliminate $\mathbf{R}$ from (60), by expressing that it is a rotation matrix, we will get equations relating the intrinsic parameters and the infinity homography matrix. This can be done much more simply than in the projective framework. The fact that $\mathbf{R}$ is a rotation matrix is equivalent to

$$\mathbf{R}\mathbf{R}^T = \mathbf{I}_3. \tag{62}$$

Substituting $\mathbf{R} = \mathbf{A}'^{-1}\mathbf{H}_\infty\mathbf{A}$ obtained from Eq. (60) into (62) yields

$$\mathbf{K}' = \mathbf{H}_\infty\mathbf{K}\mathbf{H}_\infty^T, \tag{63}$$

where the Kruppa matrices $\mathbf{K} = \mathbf{A}\mathbf{A}^T$ and $\mathbf{K}' = \mathbf{A}'\mathbf{A}'^T$ were defined in (7). It can be seen that relation (63) allows us to *update* camera calibration through a sequence of images where they do not remain constant, once the initial camera parameters (or equivalently image of the absolute conic in the first camera) are known. Since we have five equations, and each of the matrices $\mathbf{K}$ and $\mathbf{K}'$ depends on five parameters, the knowledge of five more variables is needed to determine the Euclidean description from the affine description. The fact to add a third image yields 15 equations, but since $\mathbf{K}_1 = \mathbf{H}_{\infty 31}\mathbf{K}_3\mathbf{H}_{\infty 31}^T$ is obtained by composing $\mathbf{K}_2 = \mathbf{H}_{\infty 12}\mathbf{K}_1\mathbf{H}_{\infty 12}^T$ and $\mathbf{K}_3 = \mathbf{H}_{\infty 23}\mathbf{K}_2\mathbf{H}_{\infty 23}^T$, only 10 of them are independent; therefore, there are still five unknowns.

We have obtained from (63) *five* constraints on the intrinsic parameters, which are *linear*, whereas from the projective invariants, only *two quadratic constraints* were obtained [16, 41]. *These last constraints are implied by the former ones.* Let us suppose that $\mathbf{A}$ and $\mathbf{A}'$ satisfy the constraint (63), then $\mathbf{H}_\infty\mathbf{A} = \mathbf{A}'\mathbf{R}$. The matrix $\mathbf{H}_\infty$ also must be compatible with the epipolar geometry; thus, $\mathbf{F} = [\mathbf{e}']_\times\mathbf{H}_\infty$. By substitution of the second, and then the first relation into (61), we obtain $\mathbf{E} = \mathbf{A}'^T[\mathbf{e}']_\times\mathbf{A}'\mathbf{R}$ which is the product of an antisymmetric matrix and a rotation matrix. Thus, we have shown the rigidity constraint (and in particular the Kruppa equations of [16, 44] are indeed implied by (63), and do not yield additional equations.

7. RECOVERING THE DESCRIPTIONS

So far, we have introduced the canonical decomposition in an abstract manner, starting from the projection matrices. However, the data we are using in practice are measurements performed in images, and various pieces of knowledge about the camera, motion, and 3-D world. The goal of this section is to discuss which parts of the descriptions previously described could be recovered depending on the knowledge available.

In Section 7.1 we consider the information contained in images, which allows us only to recover projective descriptions. They can nevertheless be described as a *relative* affine structure. To get affine (Section 7.2) and Euclidean (Section 7.3) information requires some additional hypotheses which we review. In particular, we discuss the problem of self-calibration when affine information is taken into account.

7.1. *Computing from the Images*

The Information Accessible from Measurements. When working with uncalibrated cameras, the only information that is accessible are pixel coordinates. Let us suppose that we observe corresponding points in two images. They are the image of point $\tilde{\mathbf{M}}$ of $\mathscr{P}^3$:

$$\kappa\mathbf{m} = \tilde{\mathbf{P}}\tilde{\mathbf{M}} \quad \kappa'\mathbf{m}' = \tilde{\mathbf{P}}'\tilde{\mathbf{M}}.$$

From the canonical decomposition, we see that this can also be written

$$\kappa\mathbf{m} = [\mathbf{I}_3, 0]\tilde{\mathbf{M}}_1 \; \kappa'\mathbf{m}' = [\mathbf{P}_c, \mathbf{p}_c]\tilde{\mathbf{M}}_1,$$

where $[\mathbf{P}_c\mathbf{p}_c]$ are one of the canonical forms of the second projection matrix, either the projective or the affine one.

The substitution of the first relation into the second yields the relation

$$\kappa' \mathbf{m}' = \kappa \mathbf{P}_c \mathbf{m} + \mathbf{p}_c. \tag{64}$$

There are several consequences of this relation. The first one is that since the quantities κ and κ' are unknown, the only equation relating $\mathbf{m}$ and $\mathbf{m}'$ is obtained by writing that the vectors $\mathbf{m}'$, $\mathbf{P}_c\mathbf{m}$, and $\mathbf{p}_c$ are coplanar. This leads to the epipolar constraint, with fundamental matrix $\mathbf{F} = [\mathbf{p}_c]_\times \mathbf{P}_c$, as already explained in Section 3.1, and thus we find again that this fundamental matrix is the only information which can be obtained from image correspondences alone.

Let us review some approaches which can be used for the recovery. The problem of the robust determination of $\mathbf{F}$ from point correspondences $(\mathbf{m}_i, \mathbf{m}_i')$ has been already considered in [10, 33, 61]. The basic idea is to use the relation (9) which is linear in the entries of $\mathbf{F}$. Another approach in the case of two views [35] is to first compute a homography, and then to use the relation between fundamental matrices and homographies presented in Section 3.3. If we have a minimum of three views, then line correspondences can also be considered [23, 67], as well as point correspondences across three views [32]. In this case, it is possible to recover in a single stage the global representation, by searching directly for the parameters of the global representation, instead of solving for the individual local representations, as illustrated by a complete implementation described in [67] and also by the recent work of [30]. This latter work is a good example of how the constraints presented in the paper could be actually enforced in a sequence involving real measurements with errors.

A second consequence is that one can see Eq. (64) as a generalization of the relation: $Z'\mathbf{m}' = Z\mathbf{Rm} + \mathbf{t}$, which holds in the calibrated case, provided a normalized representation is taken for $\mathbf{m}$ and $\mathbf{m}'$. Thus, we can interpret the quantities κ and κ' as affine or projective depths and use Eq. (64) to recover 3D structure up to an affine or projective transformation. The Euclidean depth is obtained from the affine depth by applying the transformation $\mathcal{T}_A$ in (50), whereas the affine depth is obtained from the projective depth by applying the transformation $\mathcal{T}_P$. It can be noted that the canonical decomposition gives a simple algebraic account of the geometric construction described by Shashua [57, 58, 60]. Using the form of the first projection matrix of the invariant description $[\mathbf{I}_3\ 0]$, we see that given a measurement $\mathbf{m}$ in $\mathcal{R}$, the possible 3D points $\tilde{\mathbf{M}}$ can be written $[\mathbf{m}, \lambda]$. Depending on which canonical representation we choose, λ is respectively Shashua's projective, relative affine (see below), or affine invariant.

There is no way to obtain affine or Euclidean descriptions from the measurement of fundamental matrices alone, if no additional hypotheses are made, since, as it has been seen previously, we need to know three additional parameters to obtain the affine description, and five more parameters to obtain the Euclidean description. The additional hypotheses can be about

1. the camera,
2. the displacements,
3. the scene structure.

The first class of hypotheses cannot be exploited to get affine descriptions, since the relation projective-affine does not depend explicitly on the intrinsic parameters (in both case, retinal coordinates are used). It will be seen in Section 7.3 that it can be used to obtain Euclidean descriptions from fundamental matrices, and also from infinity homography matrices. This is the *self-calibration paradigm.* The second and the third class of hypotheses can be used for both the recovery of affine or Euclidean information.

Relative Affine Structure. Although when one wants to proceed from information contained only in images, the affine description cannot be used as such, it is nevertheless possible to assign the plane at infinity to an arbitrary plane, and to use representations which are affine *relative to this plane.* The idea was first presented in [14], and has been considerably developed in [47, 57, 60], and used implicitly in [67]. Since this plane is not the plane at infinity, the representations are in fact projective, which means that the recovered structure is related to the "true" affine structure of the physical world by a projective transformation. Although no additional information is generated, there is one big advantage: an affine representation is computationally much more simple than a projective representation, as seen previously. In the following, we briefly show that the canonical decomposition accounts very simply for the relative affine construction.

We can define a canonic decomposition with respect to a given plane Π, which is equivalent to the projective one:

$$\begin{cases} \tilde{\mathbf{P}}_1 = [\mathbf{I}_3, 0]\ \mathcal{T}_\Pi \mathcal{A} \\ \tilde{\mathbf{P}}_2 = [\mathbf{H}_{\Pi 12}, \mathbf{e}_{N21}]\ \mathcal{T}_\Pi \mathcal{A} \end{cases} \quad \text{with } \mathcal{T}_\Pi = \begin{bmatrix} \mathbf{I}_3 & 0_3 \\ \nu_\Pi^T & 1 \end{bmatrix}.$$

This relation is obtained from the canonical projective decomposition just by replacing the vector $\mathbf{r}_\infty$ by the vector $\nu_\Pi = \mathbf{r}_\infty - \mathbf{r}_\Pi$ defined in (55). Alternatively, it can be interpreted from an affine representation, where $\mathbf{H}_\infty$ has been replaced by $\mathbf{H}_\Pi$, and an additional transformation is applied.

If we consider a third view, and use the canonical representation for three views that has been introduced with respect to the same plane Π, then we can note that the canonical decomposition between view 1 and view 3 takes the form

$$\begin{cases} \tilde{\mathbf{P}}_1 = [\mathbf{I}_3, 0]\, \mathscr{T}_\Pi \mathscr{A} \\ \tilde{\mathbf{P}}_3 = [\mathbf{H}_{\Pi 13}, \mathbf{e}_{N31}]\, \mathscr{T}_\Pi \mathscr{A}. \end{cases}$$

The thing which is remarkable here is that the transformation $\mathscr{T}_\Pi$ remains the same. This means that the two reconstructions done from the canonical decompositions 1–2 and 1–3 are related by an *affine transformation*, instead of a general projective transformation. However, two representations obtained by chosing a different reference plane Π will be related by a general projective transformation.

7.2. *Affine Calibration*

To get true affine information, several approximations and heuristics can be used. These concerning the scene structure are related to the very definition of $\mathbf{H}_\infty$:

• Identifying parallel directions: if images of three sets of lines which are parallel in 3D space can be identified (a task addressed for example in [1, 6, 39, 48, 51]) then we obtain three vanishing points $\mathbf{v}_i$ and $\mathbf{v}'_i$ in each image, and the infinity homography matrix is the solution of the system $\{\mathbf{H}_\infty \mathbf{v}_i = \lambda_i \mathbf{v}'_i\}$, $\mathbf{H}_\infty \mathbf{e} = \lambda \mathbf{e}'$, as mentioned by [50]. This approach works for scenes containing manmade objects (indoor, outdoor, aerial).

• Identify infinity: the homography $\mathbf{H}_\infty$ is the limit of any homography induced by a plane, when the distance of the plane to the optical centers increases arbitrarily (see Section 6.2). Thus, observing corresponding points which are at the horizon, or at remote distances provide an approximative way to compute $\mathbf{H}_\infty$, used already in [66].

• The average homography computed from a random set of points is almost equal to the infinity homography [68].

Let us now see how special displacements can lead to affine information:

• If we write Eq. (64) in the affine case $\kappa'_a \mathbf{m}' = \kappa_a \mathbf{H}_\infty \mathbf{m} + \mathrm{e}'$, we also see, since $\mathbf{e}' = \mathbf{A}'\mathrm{T}$ that when the translational component of the motion can be neglected, the average homography (i.e., the homography computed from all the point correspondences) is an approximation of the infinity homography. This is the situation used in [24] to perform self-calibration. Of course, if the motion is a pure rotation then this result becomes exact.

• In the case where the two cameras are identical, if the motion is a pure translation, then $\mathbf{H}_\infty$ is the identity, as used in [4, 45] to perform affine reconstruction.

7.3. *Euclidean Calibration*

Calibration and Self-Calibration. Self-calibration was initially defined as the task of recovering camera parameters using only image information in [16, 41] which is recovering Euclidean information from projective structure. This approach requires the weakest assumptions, since the only thing which is needed is the assumption of constant intrinsic parameters. In this paper, we use the more general meaning of recovering Euclidean information from affine or projective structure.

Using knowledge on camera displacements to get camera parameters is the more easy *calibration from motion* paradigm, for which some references are [11, 12, 65]. Assumptions about scene structure sufficient to recover camera calibration must be quite strong, and we name these approaches *model-based calibration.* Classical approaches are reviewed in [63], whereas some recent approaches include [3, 7, 8]. Part of the work of Mohr and colleagues [5, 43] can also be considered as a very sophisticated form of model-based calibration, because, although the initial setting is projective, some metric information about the world is eventually introduced.

Self-calibration Using Affine Information. It has been seen in Section 6.4 that once affine information is available and all the relations are used the system (63) has still five unknowns. An interesting way to deal with this problem is to use a single uncalibrated camera, which means a camera whose intrinsic parameters remain constant.

If the two cameras are the same, then from $\mathbf{H}_\infty = \mathbf{A}\mathbf{R}\mathbf{A}^{-1}$ it is seen that the infinity homography matrix is similar to a rotation matrix. Considering that this matrix is recovered up to a scale factor, its eigenvalues λ_i are α, $\alpha e^{i\theta}$, $\alpha e^{-i\theta}$. They must thus satisfy two constraints, which are, for example, $\|\lambda_1\| = \|\lambda_2\|$ and $\|\lambda_3\|^3 = \lambda_1\lambda_2\lambda_3$. They result in two algebraic constraints on the entries of the infinity homography matrix which would be interesting to investigate further. We can conclude that in this case the matrix $\mathbf{H}_\infty$ depends only on six parameters. Note that these constraints should be incorporated in any algorithm to estimate the matrix $\mathbf{H}_\infty$ for a single camera. A consequence is that the natural approach to solve the linear system of equations $\mathbf{K} = \mathbf{H}_\infty \mathbf{K} \mathbf{H}_\infty^T$ for the Kruppa coefficients would fail, since this system would not provide five independent equations. Let us make this point more precise. Equation (63) becomes

$$\mathbf{K} = (\mathbf{A}\mathbf{R}\mathbf{A}^{-1})\mathbf{K}(\mathbf{A}\mathbf{R}\mathbf{A}^{-1})^T.$$

Using the fact that $\mathbf{R}\mathbf{R}^T = \mathbf{I}_3$, this equation can be transformed into

$$\mathbf{R}\, \underbrace{\mathbf{A}^{-1}\mathbf{K}\mathbf{A}^{-1T}}_{\mathbf{X}} = \underbrace{\mathbf{A}^{-1}\mathbf{K}\mathbf{A}^{-1T}}_{\mathbf{X}}\, \mathbf{R}. \tag{65}$$

The matrix $\mathbf{K}$ and the matrix $\mathbf{X}$ are obtained from each other in a unique way. Using $\mathbf{X}$ as unknown, we see that

(65) is a commutator equation, but with the additional condition that $\mathbf{X}$ must be symmetric. All the solutions to this matrix equation $\mathbf{RX} = \mathbf{XR}$ can be obtained by using a rational parameterization of $\mathbf{R}$ by quaternions[4] which accounts for the orthogonal structure of the matrix $\mathbf{R}$, and writing explicitly the linear system of equations in the entries of $\mathbf{X}$. Such an approach, which is not detailed here, enables us to prove that the rank of the system is at most 4. Let us instead give a less analytical solution. Obviously, the identity matrix $\mathbf{I}_3$ is a solution. If $\mathbf{u}$ is the rotation axis of the rotation $\mathbf{R}$, we have

$$(\mathbf{Ru})\mathbf{u}^T = \mathbf{uu}^T = \mathbf{u}(\mathbf{R}^T\mathbf{u})^T = \mathbf{uu}^T\mathbf{R};$$

thus, the symmetrical matrix $\mathbf{uu}^T$ is the other solution, and we can conclude that the general solution is $\mathbf{X} = \lambda\mathbf{uu}^T + \mu\mathbf{I}_3$. This means that by solving Eq. (63) for the Kruppa matrix, we recover $\mathbf{K} = \lambda\mathbf{Au}(\mathbf{Au})^T + \mu\mathbf{AA}^T$, and thus there is an indetermination on $\mathbf{K}$ in the direction $\mathbf{Au}(\mathbf{Au})^T$.

This result can be found also in a second way. Let us represent the rotations using the exponential of a skew-symmetric matrix, that is, $\mathbf{R} = e^{[\mathbf{u}]_\times}$ where $\mathbf{u}$ is a vector aligned with the rotation axis and which magnitude is equal to the angle of the rotation. In this case, we can write, using (57),

$$\mathbf{H}_\infty = \mathbf{A}e^{[\mathbf{u}]_\times}\mathbf{A}^{-1} = e^{\mathbf{A}[\mathbf{u}]_\times\mathbf{A}^{-1}} = e^{\mathbf{K}[\rho]_\times},$$

with $\rho = \mathbf{Au}/\det(\mathbf{A})$. This relation allows us to compute $\mathbf{L} = \mathbf{K}[\rho]_\times$ as the logarithm matrix of $\mathbf{H}_\infty$ considering only the unique real solution, since $\mathbf{L}$ is a real matrix. It is now clear that we cannot recover all components of $\mathbf{K}$ and $\mathbf{L}$ but only those in a direction orthogonal to ρ.

We can conclude that two displacements with nonparallel rotation axes, or further constraints on the parameters, such as a restricted model, are necessary to recover $\mathbf{K}$ unambiguously. This is consistent with what has been observed by [65]. A similar analysis has been recently conducted in [24], where cameras performing a rotation around their optical axis were considered. For this particular motion, the retinal displacement is entirely described by $\mathbf{H}_\infty$.

8. EXAMPLES

Two examples are provided to illustrate the theory. They are presented in such a way as they can be understood before the reader has digested the main body of the paper. The associated numerical values are given in Appendix C, in order to help the reader to replicate the calculations. They involve real images, and tasks that we shall hopefully be able to perform automatically in the near future. Since our goal here is not to address stability or robustness issues, we have computed the fundamental matrices and the infinity homography matrices using calibration or self-calibration methods, in order to obtain very consistent values for these matrices. Robust methods to obtain the fundamental matrix from point correspondences work already (see Section 7.1 for references), whereas a number of practical methods to estimate the infinity homography are also available (see Section 7.2 for references).

8.1. *Computation of the Canonical Representation for Three Views and Reconstruction*

In the first example, we show how to get the canonical representations starting from pairs of uncalibrated images. The projective representation (which require no extra information) is first considered, and then affine information is introduced to enable us to get the affine representation. Please note that the approach presented here is given for illustrative purposes. A more robust way to proced [30, 67] is to recover directly the global representation instead of computing in a first stage the local representations, and then combining them.

The Projective Level. Three images of the calibration grid are used. We start from the three fundamental matrices $\mathbf{F}_{12}$, $\mathbf{F}_{32}$, $\mathbf{F}_{13}$. Remember that these matrices, which encode the epipolar geometry of each pair of views, could be determined only from the knowledge of point correspondences across pairs of views, and thus only image information is needed to obtain them. Since we are considering pairs of views in an independent manner, we get geometric information which is local, as opposed to information consistent along the whole sequence of three views. Each of these matrices has 7 degrees of freedom, and thus we have a total of 21 parameters which are not independent. We then convert each fundamental matrix to the equivalent new projective representation $\mathbf{S}$, $\mathbf{e}'$. The normalized epipoles $\mathbf{e}_{Nij}$ are obtained by solving $\mathbf{F}_{ij}\mathbf{e}_{Nij} = 0$, and this yields the epipolar projection matrices $\mathbf{S}_{ij} = [\mathbf{e}_{Nij}]_\times \cdot \mathbf{F}_{ij}$. The next thing to do is to compute the global representation from the triple of local representations: we keep the local representations 1–2 and 2–3, and replace the 7 parameters of the representation 1–3 by the scalar β_1 and the 3D vector $\mathbf{q}_n1$. These four parameters are those which are needed to obtain the representation 1–3 from the representations 1–2 and 2—3. We obtain this way the 18 independent parameters which account for the epipolar geometry of three views. Applying the formulas (49), we obtain β_1 and $\mathbf{q}_{N1}$. From then, Table 3 enables us to compute the invariant description $\mathscr{P}_1, \mathscr{P}_2, \mathscr{P}_3$, which is a particular triplet of projection matrices (the first one fixed by definition) yielding the epipolar geometry described by the three initial fundamental matrices. Thus, we have started

[4] See, for example, [15, 72].

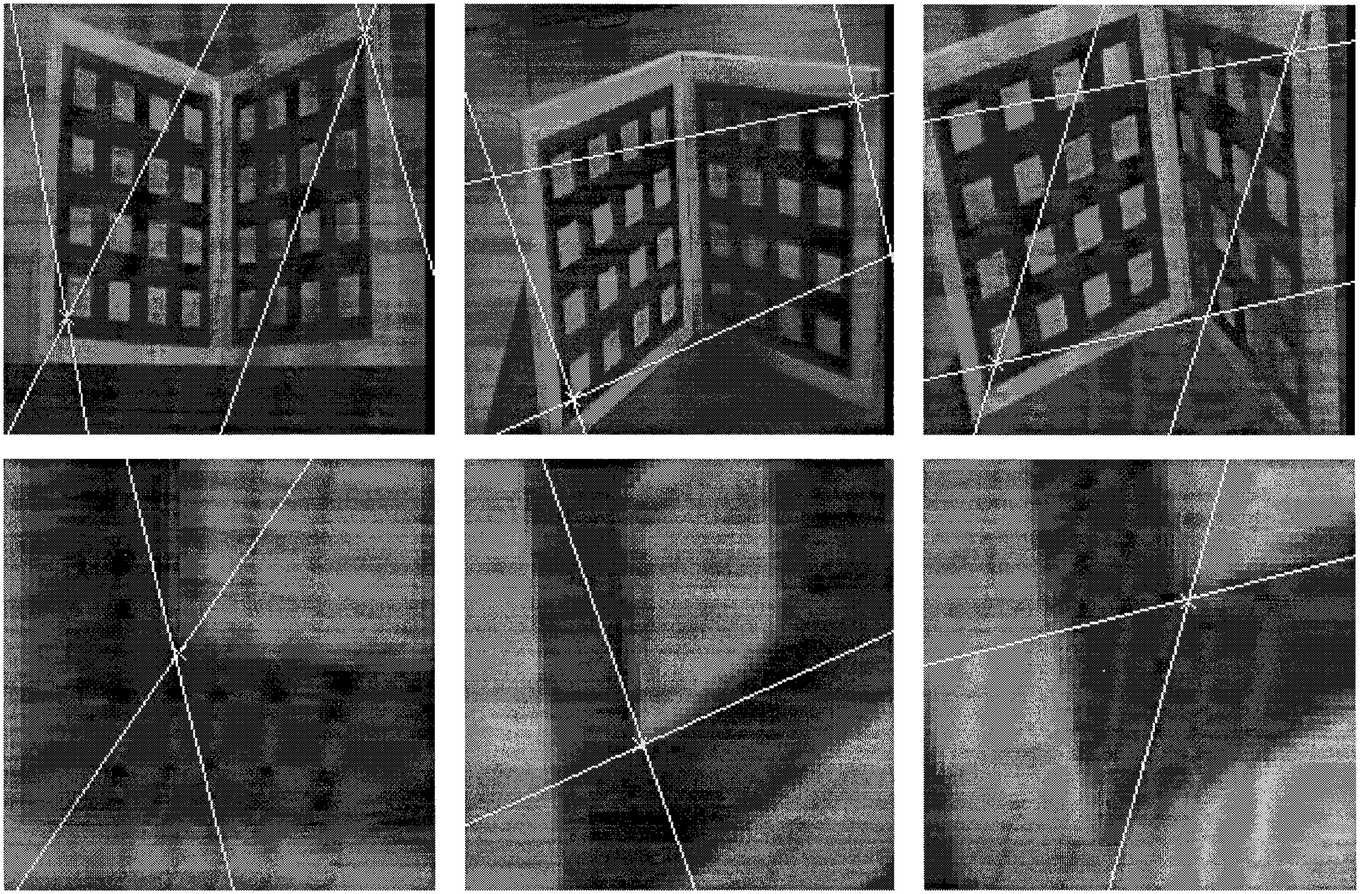


FIG. 5. Epipolar lines from the projection matrices obtained using only fundamental matrices. Control points are at the two opposite corners of the grid.

from fundamental matrices, which describe *relations* between cameras, and we have ended with projection matrices, which describe the cameras themselves. Since we started with partial information, we do not get a full determination of the projection matrices. More precisely, they are defined up to a projective transformation of space, which means that there is still an ambiguity in the reconstruction, since we cannot distinguish between two reconstructions if one is obtained from the other by a projective transformation of space. Nevertheless, the set of projection matrices which are obtained are:

- compatible with the fundamental matrices,
- uniquely defined.

The epipolar geometry computed from the matrices $\tilde{\mathbf{P}}_i$ is shown in Fig. 5. It can be seen that it is perfectly consistent. Thus, the classical *trinocular* algorithms can be used for matching and reconstruction. This is a progress over previous methods [14, 26] where only two cameras could be considered at the same time, since it is well known that trinocular methods are more precise, robust, and efficient.

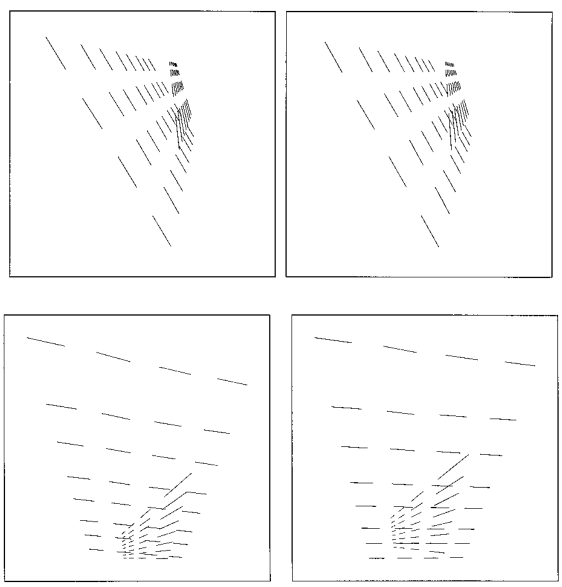


FIG. 6. Projective 3D reconstruction. (Top) Stereogram. (Bottom) Two orthographic projections.

A 3D reconstruction is shown in Fig. 6. General projective transformations can yield very strange deformations[5] which may make the shape of an object totally impossible to recognize, for displaying purposes, we have first applied a carefully chosen projective transformation of $\mathscr{P}_3$ to the three matrices above. It can be verified on the stereogram that collinearity and coplanarity are preserved. The two last views, taken under orthographic projections along the normal of each of the two planes, show clearly that parallelism is not preserved.

The Affine Level. To proceed further and obtain a reconstruction preserving the affine properties of the 3D object, some affine information is needed in addition to the fundamental matrices, which encode just the projective information. This information consists of *one* vector of dimensions 3, an $\mathbf{r}_\infty$-vector, in any of the three pairs of views. Its determination, called *affine calibration*, require some heuristics. Combining this $\mathbf{r}_\infty$-vector with the corresponding fundamental matrix gives a more intuitive piece of information, the infinity homography matrix which describes the correspondence between images of points at infinity (vanishing points). The *three* infinity homography matrices $\mathbf{H}_{\infty 12}, \mathbf{H}_{\infty 32}, \mathbf{H}_{\infty 13}$ can then be computed using (59), and since we have already epipoles from the projective level, this gives us the triple of local affine representations. As precedently, the global representation is obtained in two stages. First, the additional parameter is determined: applying (48), the value of β_1 is obtained. Then affine invariant descriptions are given by Table 3.

Unlike the projective reconstruction, the affine reconstruction limits deformations, since parallelism must be preserved, and thus we can use directly the invariant description $\mathscr{A}_1, \mathscr{A}_2, \mathscr{A}_3$ as projection matrices. Although the reconstructed points in the stereogram shown in Fig. 7 might seem all coplanar, they are not. It is just that the angle between the two planes in this reconstruction is very small. Applying a chosen affine transformation of $\mathscr{P}^3$ to the three matrices above makes the difference between the planes appear, as seen in the two last reprojected views. These views, taken under orthographic projection along the normal of each of the two planes, illustrate clearly that parallelism is preserved, but that distances and angles are not. This is to be compared with an Euclidean reconstruction, obtained from three views, given Fig. 8. Although some information is lost, the affine reconstruction remains attractive.

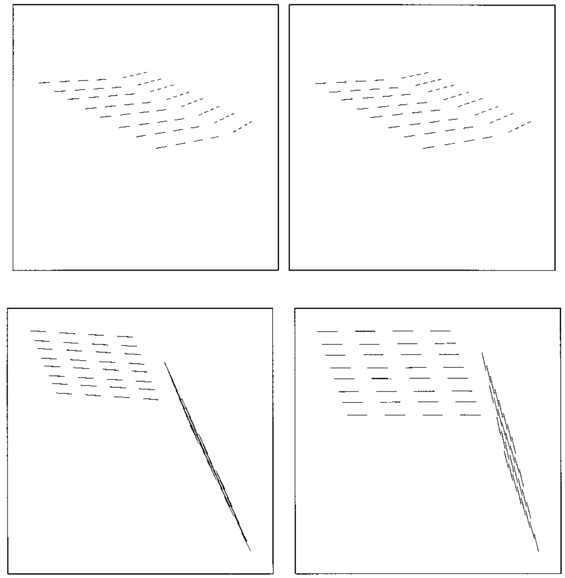


FIG. 7. Affine 3D reconstruction. (Top) Stereogram. (Bottom) Two orthographic projections.

8.2. *Using Affine Information to Perform Self-Calibration*

In the second example, we consider the problem of self-calibration, which is to get Euclidean information from uncalibrated images. In order to do this, we proceed to recover the calibration parameters of the cameras.

Three images of an indoor scene taken by a zooming camera are used. The focal has been changed from 8 to 12 mm (read on the lens barrel).

Partial Self-Calibration of a Camera from the Infinity Homography. We are first going to see that starting only from the infinity homography matrix, and provided that

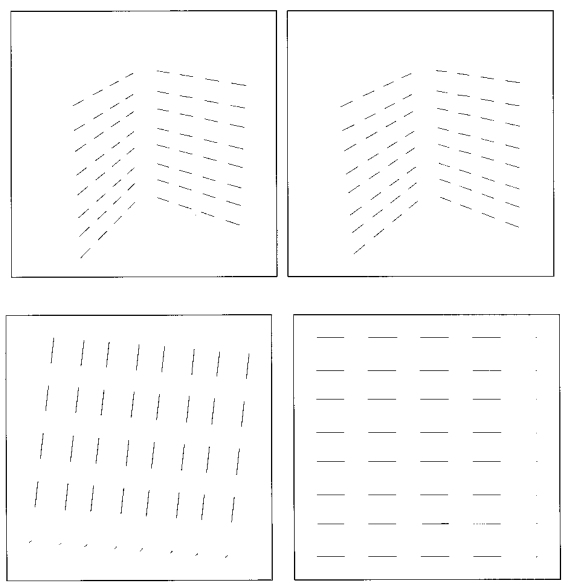


FIG. 8. Euclidean 3D reconstruction. (Top) Stereogram. (Bottom) Two orthographic projections.

[5] For instance, the inversion can map finite points to points at infinity.

the intrinsic parameters of the camera remain constant, they can be computed with an additional hypothesis, which is in practice almost always satisfied.

Let us suppose that the infinity homography matrix between the two first views is known:

$$\mathbf{H}_{\infty 12} = \begin{bmatrix} 1.72134 & -0.172125 & -218.805 \\ 0.758041 & 1.59733 & -620.388 \\ 0.00109138 & 0.000592603 & 1.00000 \end{bmatrix}.$$

The three eigenvalues of $\mathbf{H}_{\infty 12}$ have the same norm, 1.57575; thus, we conclude that the intrinsic parameters remain constant between image 1 and image 2. The system of equations (63), with $\mathbf{K} = \mathbf{K}'$, is linear. Let us use, in conformity with previous work [16], the notation for the Kruppa matrix:

$$\mathbf{K} = \begin{bmatrix} -\delta_{23} & \delta_3 & \delta_2 \\ \delta_3 & -\delta_{13} & \delta_1 \\ \delta_2 & \delta_1 & -\delta_{12} \end{bmatrix}.$$

The knowledge of this matrix is equivalent to those of the camera parameters. Normalizing by $\delta_{12} = -1$, we obtain a 5×5 matrix for the system. The singular values of this matrix are 957, 462, 0.19, 0.00055, 0.82×10^{-11}. The rank of this matrix is thus 4, as expected, and the solutions found are

$$\begin{aligned} &\delta_1 = 776.394 - 0.00115042t \\ &\delta_2 = -118.969 + 0.000817533t \qquad \delta_{13} = -731{,}756 + 0.353407t \\ &\delta_3 = 400{,}000 - 0.747472t \qquad \delta_{23} = 40{,}381.4 - 0.562483t. \end{aligned}$$

As mentioned in 7.3 we are left with a one-parameter family of solutions. A further assumption is needed to obtain a unique solution. Let us suppose that the two axes of the retinal coordinate system are orthogonal, which means that $\gamma = 0$ in Eq. (6). This is a condition satisfied by most of the cameras, and, in particular all the CCD cameras used in robotics. Then we have the additional constraint [16] $-\delta_3\delta_{12} = \delta_1\delta_2$, which yields the two solutions $t = 448{,}873$, $t = 0.116628\ 10^7$. The first solution leads to

$$\begin{aligned} &\delta_1 = 260, \delta_2 = 248, \\ &\delta_3 = 64{,}480, \delta_{13} = -573{,}121, \delta_{23} = -292{,}865 \end{aligned}$$

and then to the intrinsic parameters (scale factors in each axis, and coordinates of the principal point, formulas are in [16])

$$\alpha_u = 481, \alpha_v = 711, u_0 = 248, v_o = 260.$$

The second solution is to be discarded, because it leads to

$$\begin{aligned} &\delta_1 = 565.319, \delta_2 = 834.501, \delta_3 = -471{,}759, \\ &\delta_{13} = 319{,}585, \delta_{23} = 696{,}393. \end{aligned}$$

Thus, we see that $\delta_2^2 = \delta_{23}\delta_{12}$ and $\delta_1^2 = \delta_{13}\delta_{12}$, which corresponds to a degenerate case where $\mathbf{K}$ represents a line, and $\alpha_u = \alpha_v = 0$.

Even though we were not able to compute the full intrinsic parameters matrix $\mathbf{X}$ (five parameters), the condition $\gamma = 0$ allowed us to determine the four parameters which are significant in most applications, hence the term "partial self-calibration." Note that no information about the camera motion, or about the world, other than the three numbers provided by affine calibration (and the fundamental matrix is computed from point correspondence), was required.

Propagation of Affine and Euclidean Information. Having obtained somewhere in the image sequence Euclidean information, we show know how to propagate it on the whole sequence, thus updating the calibration of a *zooming* camera.

After determining the three fundamental matrices between the images, we are able to obtain, using (59), the other infinity homography matrices from $\mathbf{H}_{\infty 12}$. In particular, we obtain

$$\mathbf{H}_{\infty 23} = \begin{bmatrix} 1.07992 & 0.302292 & -35.2662 \\ -0.905154 & 0.959146 & 413.566 \\ -0.0000364269 & -0.000355526 & 1.00000 \end{bmatrix}.$$

The norms of the eigenvalues are 1.21, 1.21, 1.0; thus, we conclude that the intrinsic parameters have changed between image 2 and image 3. The zooming thus happened between these two images. Applying formula (63) gives immediately the new matrix $\mathbf{K}'$, from which the following parameters are computed:

$$\alpha_u = 642, \alpha_v = 950, u_0 = 248, v_0 = 263.$$

It can be verified that they correspond to the variation of focal length described previously. We have thus updated the calibration and are left with a classical (calibrated) structure from motion problem. One way to solve it is to use the intrinsic parameters and the fundamental matrices to obtain the essential matrices, and thus the motions, up to a scale factor. Applying (47) as done in [36], it is thus possible to recover three projection matrices which are equivalent to the initial ones, up to a global scale factor

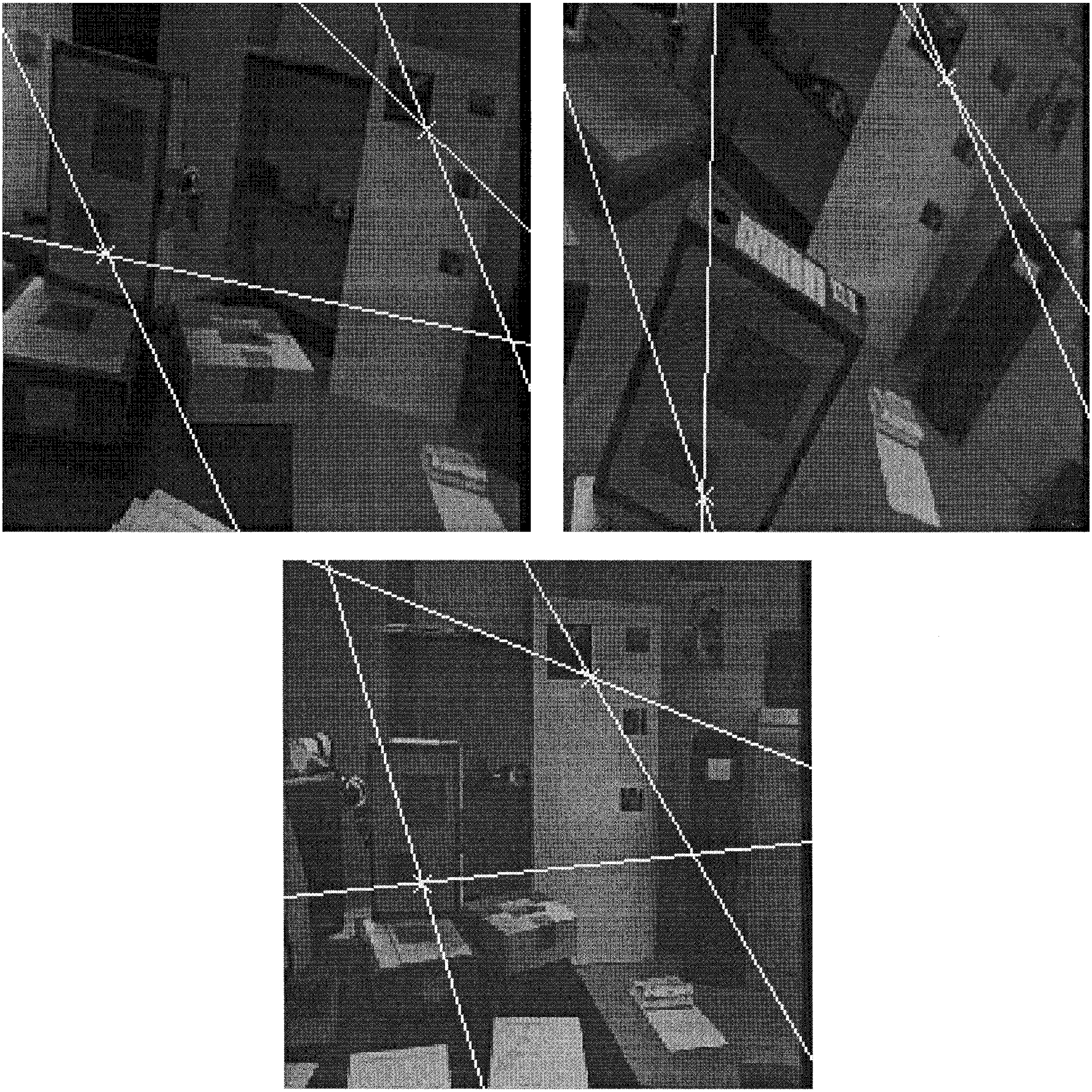


FIG. 9. Three images taken by a zooming camera, with epipolar lines.

and a change of coordinate system. This means that we are able to reconstruct the scene up to a similarity transform. Some epipolar lines obtained from these new projection matrices are shown in Fig. 9. Note that if we have a fourth view, the same method will allow us to obtain a fourth projection matrix with Euclidean information, and so on.

This example has illustrated the power of the theory and the usefulness of the proposed representations. Knowing only one infinity homography matrix enables to perform (partial) self-calibration from two views in the case of constant intrinsic parameters, but also to deal with the case of variable intrinsic parameters, by propagating affine and then Euclidean information.

9. 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 *canonical 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 the *infinity homograpy matrix* $\mathbf{H}_\infty$, *relative representation* of Π_∞, $\mathbf{r}_\infty$, and *epipolar projection matrix* $\mathbf{S}$. They 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. In the case of three views, new representations and their associated composition formulas have been established. They allow one to deal with the case of multiple viewpoints while working with uncalibrated cameras.

Some examples of specific vision tasks for which the new representation should be useful are:

Camera self-calibration. Once affine calibration is performed, camera parameters can be identified if they remain constant across images, and they can be updated in a sequence where the camera is zooming.

Token-tracking. The position of a given token can be predicted in an image sequence: given a local correspondence of a given token in two views and given the global canonical representation we can predict the location of this token in any other view, generalizing [2, 9, 18, 59]. It is important to note that this mechanism does not requires the knowledge of the Euclidean or affine geometry of the system of cameras but only its projective geometry.

3D reconstruction. It allows one to reconstruct the structure of the scene in the chosen geometry from a large number of views, generalizing [14, 26, 58]. The main idea here is that one must use the specific representation for the geometry he is concerned with: if the user is interested in detecting coplanar structures or getting the position of an object with respect to a plane, the projective geometry is suitable; if he is interested in detecting parallel lines or the middle of a line segment, the affine geometry is appropriate; etc. . . . .

Object recognition. It allows one to compute invariants which are to be used to identify a given object whatever its position in space is [46, 70, 71]. Again, depending on the kind of properties we want to use in order to characterize this object, the proposed stratified representation allows one to choose the right parametrization.

The paper makes explicit the number of parameters required for a given geometry and also tells *what* are the parameters to be used. The stability, efficiency, and some other advantages for this kind of representation are discussed in a companion paper [67], where it has been shown that we can easily implement this representation and obtain an effective parameterization of the motion in the uncalibrated case.

APPENDIX A

Algebraic Derivation of the Affine Invariants

In this section, we illustrate the principle of the derivation of invariants thanks to the simplest case, the affine one. It can be noted that using the QR decomposition, the reasoning for the Euclidean case is exactly similar, but in normalized coordinates. The projective case if more complicated, and thus we will skip algebraic details which are not enlightening. Verifications are nevertheless given in Appendix B. One choice must be made, the form of the invariant for the first camera, which we make to the simplest possible. Then the case of two views gives the nature of the invariant as a function of projection matrices, and the case of three views gives in addition the composition relations. All are done by applying the definition of the canonic decomposition.

Two Views. In the affine case, looking for matrices $\mathbf{X}$, $\mathbf{Y}$, vectors $\mathbf{x}$, $\mathbf{y}$ and a scalar μ such that

$$\begin{cases} \tilde{\mathbf{P}} = [\mathbf{I}_3, 0]\,\mathscr{A} \\ \tilde{\mathbf{P}}' = [\mathbf{X}, \mathbf{x}]\,\mathscr{A} \end{cases} \quad \text{with } \mathscr{A} = \begin{bmatrix} \mathbf{Y} & \mathbf{y} \\ \mathbf{0}_3^T & \mu \end{bmatrix} \tag{66}$$

gives the equations

$$\mathbf{P} = \mathbf{Y},\ \mathbf{p} = \mathbf{y},\ \mathbf{P}' = \mathbf{X}\mathbf{Y},\ \mathbf{p}' = \mathbf{X}\mathbf{y} + \mu\mathbf{x},$$

which immediately yield the representation $\mathbf{X} = \mathbf{P}'\mathbf{P}^{-1}$, $\mu\mathbf{x} = \mathbf{P}'\mathbf{P}^{-1}\mathbf{p} - \mathbf{p}'$. One arbitrary choice for $\|\mathbf{x}\|$ must be made, and we take the value 1; thus, $\mu = \|\mathbf{x}\|$.

Three Views. Starting from the canonic representation found for two views, we look for matrices $\mathbf{X}$, $\mathbf{Y}'$, $\mathbf{Y}''$, vectors $\mathbf{x}$, $\mathbf{y}'$, $\mathbf{y}''$, and scalars μ', μ'', such that

$$\begin{cases} \tilde{\mathbf{P}}_1 = [\mathbf{I}_3, 0]\,\mathscr{A} \\ \tilde{\mathbf{P}}_2 = [\mathbf{H}_{12}, \mathbf{e}_{N21}]\,\mathscr{A} \\ \tilde{\mathbf{P}}_3 = [\mathbf{X}, \mathbf{x}]\,\mathscr{A} \end{cases} \text{with } \mathscr{A} = \begin{bmatrix} \mathbf{Y} & \mathbf{y} \\ \mathbf{0}_3^T & \mu \end{bmatrix} \tag{67}$$

$$\begin{cases} \tilde{\mathbf{P}}_2 = [\mathbf{I}_3, \mathbf{0}]\,\mathscr{A} \\ \tilde{\mathbf{P}}_3 = [\mathbf{H}_{23}, \mathbf{e}_{N32}]\,\mathscr{A} \end{cases} \text{with } \mathscr{A} = \begin{bmatrix} \mathbf{Y}' & \mathbf{y}' \\ \mathbf{0}_3^T & \mu' \end{bmatrix} \tag{68}$$

$$\begin{cases} \tilde{\mathbf{P}}_1 = [\mathbf{I}_3, 0]\,\mathscr{A}'' \\ \tilde{\mathbf{P}}_3 = [\mathbf{H}_{13}, \mathbf{e}_{N31}]\,\mathscr{A}'' \end{cases} \text{with } \mathscr{A}'' = \begin{bmatrix} \mathbf{Y}'' & \mathbf{y}'' \\ \mathbf{0}_3^T & \mu'' \end{bmatrix}. \tag{69}$$

By equating expressions in (67) and in (68), we obtain

$$\begin{aligned} &\mathbf{Y}' = \mathbf{H}_{12}\mathbf{Y}, \mathbf{y}' = \mathbf{H}_{12}\mathbf{y} + \mu\mathbf{e}_{N21}, \mu' = \|\mathbf{e}_{32}\| \\ &\mathbf{X} = \mathbf{H}_{23}\mathbf{H}_{12}, \mathbf{x} = \mathbf{H}_{23}\mathbf{e}_{N21} + (\mu'/\mu)\mathbf{e}_{N32}, \end{aligned} \tag{70}$$

where $\mu = \|\mathbf{e}_{21}\|$, cannot be eliminated from the equations, and by equating expressions in (67) and in (69), we obtain

$$\mathbf{Y}'' = \mathbf{Y}, \mathbf{y}'' = \mathbf{y}, \mu' = \|\mathbf{e}_{31}\|, \mathbf{X} = \mathbf{H}_{13}, \mathbf{x} = (\mu''/\mu)\mathbf{e}_{N31}. \tag{71}$$

We thus obtain the two alternative representations. By equating the values of $\mathbf{X}$ and $\mathbf{x}$ found in (70) with those found in (71), the composition relations are obtained.

APPENDIX B

Proofs of Some Formulas

B.1. *The Epipolar Projection*

We show that the matrix $\mathbf{S}$ describes the correspondence from image 1 to image 2 generated by the plane $\mathbf{\Pi}_{e'}$.

First compute the intersection of the optical ray of a point $\mathbf{m}$ in the camera defined by the projection matrix $\tilde{\mathbf{P}} = [\mathbf{P}, \mathbf{p}]$ with a given plane $\mathbf{\Pi}$. Two points of this optical ray are the optical center $\mathbf{C}$ and the point of infinity representing the direction $\mathbf{P}^{-1}\mathbf{m}$ of this line. Thus, a point of this optical ray can be written

$$\mathbf{M} = \mathbf{C} + \lambda \begin{bmatrix} \mathbf{P}^{-1}\mathbf{m} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{P}^{-1}(\mathbf{p} + \lambda\mathbf{m}) \\ -1 \end{bmatrix}.$$

By writing that $\mathbf{\Pi}^T\mathbf{M} = 0$, a value for λ is found as a function of $\tilde{\mathbf{P}}$, $\mathbf{\Pi}$, and $\mathbf{m}$, and by back-substitution in the previous relation, we obtain, after some algebra (the double cross-product formula is used):

$$\mathbf{M}_{\Pi} = \begin{bmatrix} \mathbf{P}^{-1}([\mathbf{P}^{-1T}\mathbf{d}] \times [\mathbf{p}]_\times + \delta\mathbf{I}_3)\mathbf{m} \\ -\mathbf{d}^T\mathbf{P}^{-1}\mathbf{m} \end{bmatrix}, \quad \text{where } \mathbf{\Pi} = \begin{bmatrix} \mathbf{d} \\ \delta \end{bmatrix}.$$

Now let define $\mathbf{\Pi}_{e'}$ by $\mathbf{d} = \mathbf{P}'^T\mathbf{e}'$ and $\delta = \mathbf{p}'^T\mathbf{e}'$. The correspondence to $\mathbf{m}$ through plane $\mathbf{\Pi}_{e'}$ is $\mathbf{m}' = \tilde{\mathbf{P}}'\mathbf{M}_{\Pi}$, which can be written as

$$\mathbf{m}' = \mathbf{H}_\infty([\mathbf{H}_\infty^T\mathbf{e}'] \times [\mathbf{p}]_\times + \mathbf{p}'^T\mathbf{e}')\mathbf{m} - \mathbf{p}'(\mathbf{e}'^T\mathbf{H}_\infty\mathbf{m}).$$

Expanding the double cross product and using $\mathbf{H}_\infty\mathbf{p} = \mathbf{p}' - \mathbf{e}'$ yield

$$\begin{aligned} &\mathbf{H}_\infty[\mathbf{H}_\infty^T\mathbf{e}'] \times [\mathbf{p}] \times \mathbf{m} \\ &\quad = (\mathbf{e}'^T\mathbf{H}_\infty\mathbf{m})(\mathbf{p}' - \mathbf{e}') - (\mathbf{e}'^T(\mathbf{p}' - \mathbf{e}'))\mathbf{H}_\infty\mathbf{m}. \end{aligned}$$

Several terms cancel and it follows that

$$\mathbf{m}' = -(\mathbf{e}'^T\mathbf{H}_\infty\mathbf{m})\mathbf{e}' + (\mathbf{e}'^T\mathbf{e}')\mathbf{H}_\infty\mathbf{m} = -[\mathbf{e}'] \times [\mathbf{e}'] \times \mathbf{H}_\infty\mathbf{m}.$$

B.2. *The Projective Representation*

Let us verify that $\mathbf{S}$ relates the projection matrices through two relations

$$\mathbf{P}' = \mathbf{S}\mathbf{P} + \mathbf{e}'\mathscr{L}^T \tag{72}$$

$$\mathbf{p}' = \mathbf{S}\mathbf{p} + \mathbf{e}'v. \tag{73}$$

First, multiplying relation (17) by $\mathbf{P}$ yields Eq. (72), with $\mathscr{L} = \mathbf{P}^T\mathbf{r}_\infty = \mathbf{P}'^T\mathbf{e}'/\|\mathbf{e}'\|^2$. Using the expression (11) and expanding the double crossproduct leads to

$$\mathbf{S}\mathbf{p} = 1 \cdot \mathbf{H}_\infty\mathbf{p} - \left(\frac{\mathbf{e}'^T\mathbf{H}_\infty\mathbf{p}}{\|\mathbf{e}'\|^2}\right)\mathbf{e}'.$$

Since from (12), $\mathbf{H}_\infty\mathbf{p} = \mathbf{p}' - \mathbf{e}'$, we obtain Eq. (73), with $v = (1 + \mathbf{e}'^T\mathbf{H}_\infty\mathbf{p}/\|\mathbf{e}'\|^2) = \mathbf{p}'^T\mathbf{e}'/\|\mathbf{e}'\|^2$.

B.3. *Composition of Projective Representations*

The Third Epipole. Starting from the definition of $\mathbf{S}_{23}$, then using (42) and expending the double cross product yield

$$\begin{aligned} \mathbf{S}_{23}\mathbf{e}_{21} &= -\frac{[\mathbf{e}_{32}]_\times}{\|\mathbf{e}_{32}\|^2}\mathbf{F}_{23}\mathbf{e}_{21} = -\frac{\mathbf{e}_{32}}{\|\mathbf{e}_{32}\|^2} \times (\mathbf{e}_{32} \times \mathbf{e}_{31}) \\ &= \mathbf{e}_{31} - \frac{(\mathbf{e}_{32}^T\mathbf{e}_{31})}{\|\mathbf{e}_{32}\|^2}\mathbf{e}_{32}. \end{aligned} \tag{74}$$

Starting from the definition of the vectors $\mathbf{r}_\infty$ and using (34) and (30) yield

$$(\mathbf{r}_{\infty 23}-\mathbf{r}_{\infty 21})^T\mathbf{e}_{21} = -\frac{\mathbf{e}_{12}^T}{\|\mathbf{e}_{12}\|^2}\mathbf{H}_{\infty 21}\mathbf{e}_{21} + \frac{\mathbf{e}_{32}^T}{\|\mathbf{e}_{32}\|^2}\mathbf{H}_{\infty 23}\mathbf{e}_{21}$$
$$= 1 + \frac{\mathbf{e}_{32}^T}{\|\mathbf{e}_{32}\|^2}(\mathbf{e}_{31}-\mathbf{e}_{32}) = \frac{(\mathbf{e}_{32}^T\mathbf{e}_{31})}{\|\mathbf{e}_{32}\|^2}.$$

Thus,

$$\mathbf{S}_{23}\mathbf{e}_{21} + ((\mathbf{r}_{\infty 23}-\mathbf{r}_{\infty 21})^T\mathbf{e}_{21})\mathbf{e}_{32} = \mathbf{e}_{31}$$

The Third Epipolar Projection Matrix. The substitution of the decompositions (17) into (33) yields

$$\mathbf{S}_{13} - \mathbf{S}_{23}\mathbf{S}_{12} = \mathbf{S}_{23}\mathbf{e}_{21}\mathbf{r}_{\infty 12}^T + \mathbf{e}_{32}\mathbf{r}_{\infty 23}^T\mathbf{S}_{12} + \mathbf{e}_{32}\mathbf{r}_{\infty 23}^T\mathbf{e}_{21}\mathbf{r}_{\infty 12}^T - \mathbf{e}_{31}\mathbf{r}_{\infty 13}^T.$$

The first term of the right side is transformed using (74). The fourth term is transformed using $\mathbf{r}_{\infty 23}^T\mathbf{e}_{21} = \mathbf{r}_{\infty 23}^T(\mathbf{e}_{31}-\mathbf{e}_{32}) = (\mathbf{e}_{32}^T\mathbf{e}_{31}/\|\mathbf{e}_{32}\|^2) - 1$. This yields after simplifications

$$\mathbf{S}_{13} - \mathbf{S}_{23}\mathbf{S}_{12} = \mathbf{e}_{32}\mathbf{r}_{\infty 23}^T\mathbf{S}_{12} + (\mathbf{e}_{31}-\mathbf{e}_{32})\mathbf{r}_{\infty 12}^T - \mathbf{e}_{31}\mathbf{r}_{\infty 13}^T.$$

By expressing $\mathbf{S}_{12}$ and $\mathbf{r}_{\infty 21}$ as functions of infinite homographies and epipoles and using (10) and then (30), we obtain $\mathbf{r}_{\infty 12} = \mathbf{S}_{12}^T\mathbf{r}_{\infty 21} - \mathbf{e}_{12}^T/\|\mathbf{e}_{12}\|^2$. The substitution of this value yields

$$\mathbf{S}_{13} - \mathbf{S}_{23}\mathbf{S}_{12} = \mathbf{e}_{32}\left((\mathbf{r}_{\infty 23}^T - \mathbf{r}_{\infty 21}^T)\mathbf{S}_{12} + \frac{\mathbf{e}_{12}^T}{\|\mathbf{e}_{12}\|^2}\right) + \mathbf{e}_{31}\Bigg(\underbrace{\mathbf{S}_{12}^T\mathbf{r}_{\infty 12} - \frac{\mathbf{e}_{12}}{\|\mathbf{e}_{12}\|^2}}_{\mathbf{r}_{\infty 12}} - \mathbf{r}_{\infty 13}\Bigg)^T.$$

APPENDIX C

Numerical Values

We provide numerical values corresponding to the examples of Section 8.

C.1. *Example* 1

$$\mathbf{F}_{12} = \begin{bmatrix} -0.1395\times10^{-5} & 0.1853\times10^{-5} & 0.001981 \\ -0.8986\times 10-5 & 0.1374\times10^{-5} & 0.01201 \\ 0.001424 & -0.01269 & 1 \end{bmatrix},$$

$$\mathbf{F}_{32} = \begin{bmatrix} -0.8549\times10^{-5} & 0.7565\times10^{-5} & 0.01596 \\ -0.1561\times10^{-5} & 0.2783\times10^{-5} & 0.004520 \\ -0.01474 & -0.01013 & 1 \end{bmatrix}$$

$$\mathbf{F}_{13} = \begin{bmatrix} 0.5978\times10^{-5} & 0.4767\times10^{-5} & -0.02712 \\ 0.4380\times10^{-5} & 0.2787\times10^{-5} & 0.005137 \\ 0.02105 & -0.007449 & 1 \end{bmatrix}$$

$\beta_1 = 2.5651, \mathbf{q}_{N1} = [-0.00759, 0.008898, -0.03468]^T$

$$\mathcal{P}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\mathcal{P}_2 = \begin{bmatrix} 6911.48 & -1363.61 & 2373.28 & 7873.86 \\ -1200.83 & 343.822 & -8656.55 & -1381.41 \\ 0.803724 & -0.161854 & 99.3223 & 1.000000 \end{bmatrix}$$

$$\mathcal{P}_3 = \begin{bmatrix} -635.855 & 126.875 & -1835.90 & -694.050 \\ -3351.10 & 640.177 & -1440.08 & -3859.31 \\ 0.877744 & -0.158361 & -42.4041 & 1.00000 \end{bmatrix}$$

$$\mathbf{H}_{\infty 12} = \begin{bmatrix} 0.7091 & -0.004109 & 397.0 \\ -0.01186 & 1.004 & -148.7 \\ -0.0006201 & 0.0001081 & 1.000 \end{bmatrix},$$

$$\mathbf{H}_{\infty 32} = \begin{bmatrix} 1.034 & 0.3276 & 163.0 \\ -0.2321 & 1.177 & -796.9 \\ -0.0005753 & 0.0005679 & 1.000 \end{bmatrix}$$

$$\mathbf{H}_{\infty 13} = \begin{bmatrix} 0.8853 & -0.1793 & 140.9 \\ -0.007023 & 0.7471 & 549.2 \\ -0.0002499 & -0.0003943 & 1.0000 \end{bmatrix}$$

$$\mathcal{A}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\mathcal{A}_2 = \begin{bmatrix} -5668.34 & 32.85 & -0.31736\times10^7 & 7873.86 \\ 94.778 & -8023.47 & 0.11885\times10^7 & -1381.41 \\ 4.9569 & -0.864224 & -7994.12 & 1.000 \end{bmatrix}$$

$$\mathcal{A}_3 = \begin{bmatrix} 541.50 & -514.85 & 404{,}452. & -694.05 \\ -20.1608 & 2144.7 & 0.15767\times10^7 & -3859.3 \\ -0.717456 & -1.1320 & 2870.78 & 1.000 \end{bmatrix}$$

C.2. *Example* 2

$$\mathbf{F}_{12} = \begin{bmatrix} 0.00030194 & 0.00022904 & 0.20447 \\ -0.00018455 & -0.000039449 & 0.070503 \\ 0.34804 & -0.17286 & 1.000 \end{bmatrix}$$

$$\mathbf{F}_{23} = \begin{bmatrix} 0.00049813 & -0.000052961 & -0.022319 \\ 0.000059537 & 0.000017487 & -0.0043212 \\ -0.0026382 & 0.0021470 & 1.00 \end{bmatrix}$$

$$\mathbf{F}_{13} = \begin{bmatrix} 0.43369 \times 10^{-5} & -0.000014001 & 0.0024273 \\ 0.000012721 & 0.19193 \times 10^{-5} & -0.0068886 \\ -0.0053471 & 0.0050099 & 1.00 \end{bmatrix}.$$

ACKNOWLEDGMENTS

Q.T.L. is grateful to J. Malik for encouragements, and support provided through the PATH project under MOU 94. We thank L. Robert and our colleagues at UC Berkeley, particularly T. Lindsey, for proofreading the paper, as well as the anonymous reviewers, whose suggestions have lead to a much improved presentation.

REFERENCES

1. S. T. Barnard, Interpreting perspective images, *Artif. Intell.* **21**(4), 1983, 435–462.
2. E. B. Barrett, M. H. Brill, N. N. Haag, and P. M. Payton, Invariant linear methods in photogrammetry and model-matching, in *Geometric Invariance in Computer Vision* (J. L. Mundy and A. Zisserman, Eds.), Chap. 14, pp. 277–292, MIT Press, Cambridge, MA, 1992. MIT Press, Cambridge, MA,
3. P. A. Beardsley, Applications of projective geometry to computer vision, Ph.D. thesis, University of Oxford, 1992.
4. P. A. Beardsley, A. Zisserman, and D. W. Murray, Navigation using affine structure from motion, in *Proceedings, European Conference on Computer Vision, Stockholm, Sweden, 1994,* pp. 85–96.
5. B. Boufama, R. Mohr, and F. Veillon, Euclidean constraints for uncalibrated reconstruction, in *Proceedings, International Conference on Computer Vision, Berlin, Germany, 1993,* pp. 466–470.
6. B. Brillault-O'Mahony, New method for vanishing point detection, *Comput. Vision Graphics Image Process. Image Understanding* **54**(2), 1991, 289–300.
7. B. Caprile and V. Torre, Using vanishing points for camera calibration, *Int. J. Comput. Vision* **4,** 1990, 127–140.
8. N. Daucher, M. Dhome, and J. T. Lapreste, Camera calibration from sphere images, in *Proceedings, European Conference on Computer Vision, Stockholm, Sweden, 1994,* pp. 449–454.
9. S. Demey, A. Zisserman, and P. A. Beardsley, Affine and projective structure from motion, in *Proceedings, British Machine Vision Conference, Leeds, UK, 1992,* pp. 49–58.
10. R. Deriche, Z. Zhang, Q.-T. Luong, and O. D. Faugeras, Robust recovery of the epipolar geometry for an uncalibrated stereo rig. in *Proceedings, European Conference on Computer Vision, Stockholm, Sweden, 1994,* pp. 567–576.
11. L. Dron, Dyanmic camera self-calibration from controlled motion sequences, in *Proceedings, of the Conference on Computer Vision and Pattern Recognition, New York, 1993,* pp. 501–506.
12. F. Du and M. Brady, Self-calibration of the intrinsic parameters of cameras for active vision systems, in *Proceedings, of the Conference on Computer Vision and Pattern Recognition, New York, 1993,* pp. 477–482.
13. O. D. Faugeras and F. Lustman, Let us suppose that the world is piecewise planar, in *Robotics Research, The Third International Symposium* (O. D. Faugeras and G. Giralt, Eds.), pp. 33–40, MIT Press, Cambridge, MA, 1986.
14. O. D. Faugeras, What can be seen in three dimensions with an uncalibrated stereo rig, in *Proceedings, European Conference on Computer Vision,* 1992, pp. 563–578.
15. O. D. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint,* MIT Press, Cambridge, MA, 1993.
16. O. D. Faugeras, Q.-T. Luong, and S. J. Maybank, Camera self-calibration: Theory and experiments, in *Proceedings, European Conference on Computer Vision, Santa Margerita, Italy, 1992,* pp. 321–334.
17. O. D. Faugeras and S. J. Maybank, Motion from point matches: Multiplicity of solutions, *Int. J. Comput. Vision* **4**(3), 1990, 225–246. Also INRIA Tech. Report 1157.
18. O. D. Faugeras and L. Robert, What can two images tell us about a third one? Technical Report RR-2018, INRIA, 1993. To appear in *Int. J. Comput. Vision.*
19. D. A. Forsyth, J. L. Mundy, A. Zisserman, and C. M. Brown, Projectively invariant representations using implicit algebraic curves, in *Proceedings, European Conference on Computer Vision, Antibes, France, 1990,* pp. 427–436.
20. G. H. Golub and C. F. Van Loan, *Matrix Computations,* John Hopkins Univ. Press, Baltimore, 1989.
21. R. I. Hartley, Cheirality invariants, in *Proceedings, DARPA Image Understanding Workshop,* University of Maryland, 1993, pp. 745–753.
22. R. I. Hartley, An algorithm for self calibration from several views, in *Proceedings, Conference on Computer Vision and Pattern Recognition, Seattle, WA, 1994,* pp. 908–912.
23. R. I. Hartley, Projective reconstruction from line correspondences, in *Proceedings, Conference on Computer Vision and Pattern Recognition, Seattle, WA, 1994,* pp. 903–907.
24. R. I. Hartley, Self-calibration from multiple views with a rotating camera, in *Proceedings, European Conference on Computer Vision, Stockholm, Sweden, 1994,* pp. 471–478.
25. R. I. Hartley and R. Gupta, Computing matched-epipolar projections, in *Proceedings, Conference on Computer Vision and Pattern Recognition, New York, 1993,* pp. 549–555.
26. R. I. Hartley, R. Gupta, and T. Chang, Stereo from uncalibrated cameras, in *Proceedings, Conference on Computer Vision and Pattern Recognition, Urbana, 1992,* pp. 761–764.
27. T. S. Huang and O. D. Faugeras, Some properties of the E-matrix in two view motion estimation, *IEEE Trans. Pattern Anal. Mach. Intell.* **11,** 1989, 1310–1312.
28. J. Illingworth (Ed.), *Proceedings, British Machine Vision Conference, Guildford, UK,* Sept. 1993.
29. J. J. Koenderink and A. J. van Doorn, Affine structure from motion, *J. Opt. Soc. Am. A* **8**(2), 1991, 377–385.
30. S. Laveau, O. D. Faugeras, and M. Hebert, Estimation of the relative geometry of a set of n cameras, technical report, INRIA, 1994.
31. H. C. Longuet-Higgins, A computer algorithm for reconstructing a scene from two projections, *Nature* **293,** 1981, 133–135.
32. Q.-T. Luong, On the geometry of three views, University of California at Berkeley, 1994.
33. Q.-T. Luong, R. Deriche, O. D. Faugeras, and T. Papadopoulo, On determining the fundamental matrix: Analysis of different methods and experimental results, Technical Report RR-1894, INRIA, 1993.
34. Q.-T. Luong and O. D. Faugeras, Self calibration of a moving camera

from point correspondences and fundmental matrices *Int. J. Comput. Vision,* 1996.

35. Q.-T. Luong and O. D. Faugeras. Determining the fundamental matrix with planes: Unstability and new algorithms, in *Proceedings, Conference on Computer Vision and Pattern Recognition, New York, 1993,* pp. 489–494.
36. Q.-T. Luong and O. D. Faugeras, Self-calibration of a stereo rig from unknown camera motions and point correspondence, in *Calibration and Orientation of Cameras in Computer Vision* (A. Grun and T. S. Huang, Eds.), Springer-Verlag, Berlin/New York, 1993. Also presented at XVII ISPRS, Washington, and INRIA Technical Report RR-2014.
37. Q.-T. Luong and O. D. Faugeras, The fundamental matrix: Theory, algorithms, and stability analysis, *Int. J. Comput. Vision,* **17**(1), 1996, 43–76.
38. Q.-T. Luong and O. D. Faugeras, An optimizaton framework for efficient self-calibration and motion determination, in *Proceedings International Conference on Pattern Recognition, Jerusalem, Israel, 1994,* pp. A248–252.
39. M. J. Magee and J. K. Aggarwal, Determining vanishing points in perspective images, *Comput. Vision Graphics Image Process.* **26,** 1984, 256–267.
40. S. J. Maybank, Properties of essential matrices, *Int. J. Imag. Systems Technol.* **2,** 1990, 380–384.
41. S. J. Maybank and O. D. Faugeras, A theory of self-calibration of a moving camera, *Int. J. Comput. Vision* **8**(2), 1992, 123–151.
42. P. F. McLauchlan, I. D. Reid, and D. W. Murray, Recursive affine structure and motion from image sequences, in *Proceedings, European Conference on Computer Vision, Stockholm, Sweden, 1994,* pp. 217–224.
43. R. Mohr, L. Morin, and E. Grosso, Relative positioning with poorly calibrated cameras, in *Geometric Invariance in Computer Vision* Chap. 22, pp. 440–462, MIT Press, Cambridge, MA, 1992
44. R. Mohr, F. Veillon, and L. Quan, Relative 3D reconstruction using multiple uncalibrated images, in *Proceedings, Conference on Computer Vision and Pattern Recognition, New York, 1993,* pp. 543–548.
45. T. Moons, L. Van Gool, M. Van Diest, and E. Pauwels, Affine reconstruction from perspective image pairs, in *Proccedings of the 2nd European Workship on Invariants, Ponta Delagada,, Azores, 1993,* pp. 249–266.
46. J. L. Mundy and A. Zisserman (Eds.), *Geometric Invariance in Computer Vision,* MIT Press, Cambridge, MA, 1992.
47. N. Navab and A. Shashua, Algebraic description of relative affine structure: Connections to Euclidean, affine, and projective structure, Technical Report 270, Media Lab, MIT, 1994.
48. P. Palmer and A. Tai, An optimised vanishing point detector, in *Proceedings, British Machine Vision Conference, Guildford, UK, 1993,* pp. 529–538.
49. J. Ponce, D. H. Marimont, and T. A. Cass, Analytical methods for uncalibrated stereo and motion reconstruction, in *Proceedings, European Conference on Computer Vision, Stockholm, Sweden, 1994,* pp. 463–470.
50. L. Quan, Affine stereo calibration for relative affine shape reconstruction, in *Proceedings, British Machine Vision Conference, 1993,* pp. 659–668.
51. L. Quan and R. Mohr, Determining perspective structures using hierarchical Hough transform, *Pattern Recognit. Lett.* **9**(4), 1989, 279–286.
52. L. Quan and R. Mohr, Affine shape representation from motion through reference points, *J. Math. Imaging Vision* **1,** 1992, 145–151.
53. L. Robert and O. D. Faugeras, Relative 3D positioning and 3D convex hull computation from a weakly calibrated stereo pair, in *Proceedings, International Conference on Computer Vison, Berlin, Germany, May 1993,* pp. 540–543.
54. L. G. Roberts, Machine perception of three-dimensional solids, in *Optical and Electro-Optical Information Processes* (J. T. Tippet *et al.*, Eds.), pp. 159–197, MIT-Press, Cambridge, MA, 1966.
55. J. G. Semple and G. T. Kneebone, *Algebraic Projective Geometry,* Clarendon Press, Oxford, 1952, reprinted 1979.
56. 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. A shorter version appeared at ECCV'94.
57. A. Shashua, On geometric and algebraic aspects of 3D affine and projective structures from perspective 2D views, in *Proceedings of the 2nd European Workshop on Invariants, Ponta Delagada, Azores, 1993.* Also MIT AI Memo 1405, July 1993.
58. A. Shashua, Projective depth: A geometric invariant for 3D reconstruction from two perspective/orthographic views and for visual recognition, in *Proceedings, International Conference on Computer Vision, Berlin, Germany, 1993,* pp. 583–590.
59. A. Shashua, Trilinearity in visual recognition by alignment, in *Proceedings, European Conference on Computer Vision, Stockholm, Sweden, 1994,* pp. 479–486.
60. A. Shashua and N. Navab, Relative affine structure: Theory and application to 3D reconstruction from perspective views, in *Proceedings, Conference on Computer Vision and Pattern Recognition, Seattle, WA 1994,* pp. 483–489.
61. P. H. S. Torr and D. W. Murray, Outlier detection and motion segmentation, Technical Report OUEL 1987/93, Oxford University, Sept. 1993.
62. G. Toscani, *Systèmes de Calibration et Perception du Mouvement en Vision Articielle,* Ph.D. thesis, Université de Paris-Sud Centre d'Orsay, Dec. 1987.
63. R. Y. Tsai, Synopsis of recent progress on camera calibration for 3D machine vision, in *The Robotics Review* (O. Khatib, J. J. Craig, and T. Lozano-Pérez, Eds.), pp. 147–159, MIT Press, Cambridge, MA, 1989.
64. R. Y. Tsai and T. S. Huang, Estimating three-dimensional motion parameters of a rigid planar patch, II: Singular value decomposition, *IEEE Trans. Acoust. Speech Signal Process.* **30,** 1982.
65. T. Viéville, Auto-calibration of visual sensor parameters on a robotic head, *Image Vision Comput.* **12,** 1994.
66. T. Viéville, P. Facao, and E. Clergue, Computation of ego-motion and structure from visual and inertial sensors using the vertical cue, in *Proceedings, International Conference on Computer Vision, Berlin, Germany, 1993,* pp. 591–598.
67. T. Viéville, O. D. Faugeras, and Q.-T. Luong, Motion of points and lines in the uncalibrated case, *Int. J. Comput. Vision,* **17**(1), 1996, 7–42.
68. T. Viéville, C. Zeller, and L. Robert, Using collineations to compute motion and structure in an uncalibrated image sequence, *Int. J. Comput. Vision,* 1996.
69. L. Wang and W. Tsai, Computing cameras parameters using vanishing line information from a rectangular parallepiped, *Mach. Vision Appl.* **3,** 1990, 129–141.
70. D. Weishall, Model-based invariants for 3-D vision, *Int. J. Comput. Vision* **10**(1), 1993, 27–42.
71. I. Weiss, Geometric invariants and object recognition, *Int. J. Comput. Vision* **10**(3), 1993, 207–231.
72. Z. Zhang and O. D. Faugeras, *3D Dynamic Scene Analysis,* Springer-Verlag, New York/Berlin, 1992.