A Nonlinear Method for Estimating the Projective Geometry of Three Views

Olivier Faugeras, Théodore Papadopoulo

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A nonlinear method for estimating the projective geometry of three views

O. Faugeras,T. Papadopoulo

N° 3221

Juillet 1997

THÈME 3

ISSN 0249-6399

A nonlinear method for estimating the projective geometry of three views

O. Faugeras,T. Papadopoulo*

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Robotvis

Rapport de recherche n˚3221 — Juillet 1997 — 27 pages

Abstract: Given three partially overlapping views of a scene from which a set of point correspondences have been extracted, recover the three trifocal tensors between the three views. We give a new way of deriving the trifocal tensor based on Grassmann-Cayley algebra that sheds some new light on its structure. We show that our derivation leads to a complete characterization of its geometric and algebraic properties which is fairly intuitive, i.e. geometric. We give a set of algebraic constraints which are satisfied by the 27 coefficients of the trifocal tensor and allow to parameterize it minimally with 18 coefficients. We then describe a robust method for estimating the trifocal tensor from point and line correspondences that uses this minimal parameterization. Our experimental results show that this method is superior to the linear methods which had been previously published.

Key-words: Multiple-view geometry, Trifocal tensor, Fundamental matrix, Grassmann-Cayley algebra, Calibration

(Résumé : tsvp)

The research described in this report has been partially supported by the European Reactive LTR Project 21914–CUMULI

* {Olivier.Faugeras,Theodore.Papadopoulo}@sophia.inria.fr

Unité de recherche INRIA Sophia Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA ANTIPOLIS Cedex (France)
Téléphone : 04 93 65 77 77 – Télécopie : 04 93 65 77 65

Estimation non-linéaire de la géométrie projective associée à trois vues

Résumé : Étant données trois vues d'une même scène pour lesquelles on dispose d'un ensemble de de triplets de points-image en correspondance, il s'agit de déterminer les trois tenseurs trifocaux qui lient ces trois vues. Une nouvelle présentation du tenseur trifocal basée sur l'algèbre de Grassmann-Cayley permet de mettre en évidence certaines propriétés structurelles de ces tenseurs. Nous avons ainsi obtenu une caractérisation complète et assez intuitive de leurs propriétés algébriques et géométriques. Nous décrivons également un ensemble de contraintes algébriques qui doivent êtres satisfaites par les 27 coefficients de ces tenseurs. Celles-ci nous permettent de définir une paramétrisation minimale des tenseurs à l'aide de 18 paramétres. Cette paramétrisation nous a permis de développer une méthode non-linéaire robuste pour l'estimation du tensor trifocal à partir de correspondances de points et de droites. Nos résultats expérimentaux montrent que cette méthode donne de bien meilleurs résultats que les méthodes linéaires qui ont été publiées jusque ici.

Mots-clé : Géométrie multi-vues, Tenseur trifocal, Matrice fondamentale, Algèbre de Grassmann-Cayley, Calibration

1 Introduction

This article deals with the following problem:
Given three partially overlapping views of a scene from which a set of point correspondences have been extracted, recover the three trifocal tensors between the three views.

This problem is important because once the trifocal tensors are known, they can be used for a variety of useful tasks, such as transfer [BBHP92, BGP93] computation of the perspective projection matrices of the three views, and 3D reconstruction of the scene up to a projective transformation. They could also be used, but this has not yet been much explored, to perform Euclidean self-calibration [AZH96].

Given three views, it has been shown originally by Shashua [Sha94b] that the coordinates of three corresponding points satisfied a set of algebraic relations of degree 3 called the trilinear relations. It was later on pointed out by Hartley [Har94a] that those trilinear relations were in fact arising from a tensor that governed the correspondences of lines between three views which he called the trifocal tensor. Hartley also correctly pointed out that this tensor had been used, if not formally identified as such, by researchers working on the problem of the estimation of motion and structure from line correspondences [SA90b]. Given three views, there are of course three such tensors, depending upon which view is selected as the one one wants to predict to.

The trinocular tensors play the same role in the analysis of scenes from three views as the fundamental matrix play in the two-view case, therefore the question of their estimation from feature correspondences arise naturally. The question of estimating the fundamental matrix between two views has received considerable attention in the last few years and robust algorithms have been proposed by a number of researchers [DZLF94, ZDFL95, TZ97, Har95]. The main difficulty of the estimation arises from the fact that the fundamental matrix must satisfy one nonlinear constraint, i.e. that its determinant is equal to 0, which prevents the straightforward application of quadratic least-squares methods.

The related question for the trinocular tensor has received much less attention except for the obvious application of quadratic least-squares methods [Har94a, Sha95]. What makes the use of these methods even more questionable in the case of the trinocular tensor is the fact that it is much more constrained than the fundamental matrix: even though it superficially seems to depend upon 26 parameters (27 up to scale), these 26 parameters are not independent since the number of degrees of freedom of three views has been shown to be equal to 18 in the projective framework (33 parameters for the 3 perspective projection matrices minus 15 for an unknown projective transformation) [LV94]. Therefore the trifocal tensor can depend upon at most 18 independent parameters and therefore its 27 components must satisfy a number of algebraic constraints, some of them have been elucidated

[SW95, AS96]. Unfortunately, since those constraints are a) partially unknown and b) complicated, no estimation algorithm has been published yet that takes them into account.

The contributions of this paper are four-folds. First we give a new way of deriving the trifocal tensor based on Grassmann-Cayley algebra, second we show that our derivation leads to a complete characterization of its geometric and algebraic properties, third we give a set of algebraic constraints which are satisfied by the 27 coefficients of the trifocal tensor and allow to parameterize it minimally with 18 coefficients, and fourth we describe a robust method for estimating the trifocal tensor from point and line correspondences that uses this minimal parameterization. Our experimental results show that this method is superior to the linear methods which had been previously published. For the sake of brevity, we have in general omitted the proofs of our theoretical assertions.

We assume that the reader is familiar with elementary Grassmann-Cayley algebra since the necessary ingredients have already been presented to the Computer Vision community in a number of publications such as, for example, [Car94, FM95a].

2 Monocular and binocular geometry

We take in this article the viewpoint of projective geometry. We consider that a camera can be modeled accurately as a pinhole and performs a perspective projection. If we consider two arbitrary systems of projective coordinates, for the image and the object space, the relationship between 2-D pixels and 3-D points can be represented as a linear projective operation which maps points of $\mathbb{P}^3$ to points of $\mathbb{P}^2$. This operation can be described by a 3×4 matrix $\boldsymbol{\mathcal{P}}$, called the *perspective projection* matrix of the camera:

$$\mathbf{m}^T = [x \;\; y \;\; z]^T \simeq [\mathcal{X} \; \mathcal{Y} \; \mathcal{Z} \; \mathcal{T}] \, \boldsymbol{\mathcal{P}}^T = \mathbf{M}^T \boldsymbol{\mathcal{P}}^T \tag{1}$$

This matrix is of rank 3. Its nullspace is therefore of dimension 1, corresponding to a unique point of $\mathbb{P}^3$, the optical center C of the camera.

We give a geometric interpretation of the rows of the projection matrix. We use the notation:

$$\boldsymbol{\mathcal{P}}^T = \left[\boldsymbol{\Gamma}^T \; \boldsymbol{\Lambda}^T \; \boldsymbol{\Theta}^T \right] \tag{2}$$

where $\boldsymbol{\Gamma}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\Theta}$ are the row vectors of $\boldsymbol{\mathcal{P}}$. Each of these vectors represent a plane in 3D. These three planes are called the *projection planes* of the camera. The projection equation (1) can be rewritten as:

$$x : y : z = \langle \boldsymbol{\Gamma}, \mathbf{M} \rangle : \langle \boldsymbol{\Lambda}, \mathbf{M} \rangle : \langle \boldsymbol{\Theta}, \mathbf{M} \rangle$$

where, for example, $\langle \mathbf{\Gamma}, \mathbf{M} \rangle$ is the dot product of the plane represented by $\mathbf{\Gamma}$ with the point represented by $\mathbf{M}$. This relation is equivalent to the three scalar equations, of which two are independent:

$$x\langle \mathbf{\Lambda}, \mathbf{M} \rangle - y\langle \mathbf{\Gamma}, \mathbf{M} \rangle = 0 \quad y\langle \mathbf{\Theta}, \mathbf{M} \rangle - z\langle \mathbf{\Lambda}, \mathbf{M} \rangle = 0 \quad z\langle \mathbf{\Lambda}, \mathbf{M} \rangle - x\langle \mathbf{\Theta}, \mathbf{M} \rangle = 0 \quad (3)$$

The planes of equation $\langle \mathbf{\Gamma}, \mathbf{M} \rangle = 0$, $\langle \mathbf{\Lambda}, \mathbf{M} \rangle = 0$ and $\langle \mathbf{\Theta}, \mathbf{M} \rangle = 0$ are mapped to the image lines of equations $x = 0$, $y = 0$, and $z = 0$, respectively. We have the proposition:

Proposition 1 *The three projection planes of a perspective camera intersect the retinal plane along the three lines going through the first three points of the standard projective basis.*

The optical center is the unique point C which satisfies $\mathcal{P}\mathbf{C} = \mathbf{0}$. Therefore this point is the intersection of the three planes represented by $\mathbf{\Gamma}$, $\mathbf{\Lambda}$, $\mathbf{\Theta}$. In the Grassmann-Cayley formalism, it is represented by the meet of those three planes $\mathbf{\Gamma} \vartriangle \mathbf{\Lambda} \vartriangle \mathbf{\Theta}$. This is illustrated in Fig. 1. Because of the definition of the meet operator, the projective coordinates of C are the four 3×3 minors of matrix $\mathcal{P}$:

Proposition 2 *The optical center C of the camera is the meet $\mathbf{\Gamma} \vartriangle \mathbf{\Lambda} \vartriangle \mathbf{\Theta}$ of the three projection planes.*

The three projection planes intersect along the three lines $\mathbf{\Gamma} \vartriangle \mathbf{\Lambda}$, $\mathbf{\Lambda} \vartriangle \mathbf{\Theta}$ and $\mathbf{\Theta} \vartriangle \mathbf{\Gamma}$ called the *projection rays*. These three lines meet at the optical center C and intersect the retinal plane at the first three points e_1, e_2 and e_3 of the standard projective basis.

In the case of two cameras, it is well-known that the geometry of correspondences between the two views can be described compactly by the fundamental matrix, noted $\mathbf{F}_{12}$ which associates to each pixel m_1 of the first view its epipolar line noted l_{m_1} in the second image:

$$\mathbf{l}_{m_1} \simeq \mathbf{F}_{12}\mathbf{m}_1$$

similarly $\mathbf{F}_{21} = \mathbf{F}_{12}^T$ associates to a pixel m_2 of the second view its epipolar line l_{m_2} in the first one.

The matrix $\mathbf{F}_{12}$ (resp. $\mathbf{F}_{21}$) is of rank 2, the point in its null-space is the epipole $e_{1,2}$ (resp. the epipole $e_{2,1}$):

$$\mathbf{F}_{12}\mathbf{e}_{1,2} = \mathbf{F}_{21}\mathbf{e}_{2,1} = \mathbf{0}$$

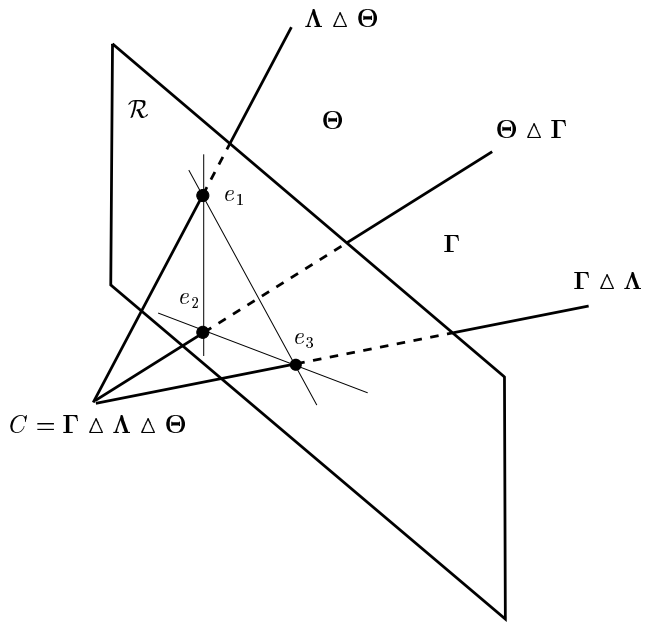

Figure 1: Geometrical interpretation of the three rows of the projection matrix as planes. The three projection planes $\mathbf{\Gamma}$, $\mathbf{\Lambda}$ and $\mathbf{\Theta}$ are projected into the axes of the retinal coordinate system. The three projection rays intersect the retinal plane at the first three points of the retinal projective basis. The three projection planes meet at the optical center.

3 Trinocular Geometry

3.1 Trifocal geometry from binocular geometry

When we add one more view, the geometry becomes more intricate, see figure 2. When the three optical centers $C_1,\ C_2,\ C_3$ are not aligned they define a plane, called the *trifocal plane*, which intersects the three image planes along the *trifocal lines* $t_1,\ t_2,\ t_3$ which contain the epipoles $e_{i,j},\ i \neq j\ i = 1, \ldots, 3,\ j = 1, \ldots, 3$. The three fundamental matrices $\mathbf{F}_{12},\ \mathbf{F}_{23}$ and $\mathbf{F}_{31}$ are not independent since they must satisfy the three constraints:

$$\mathbf{e}_{2,3}^T \mathbf{F}_{12} \mathbf{e}_{1,3} = \mathbf{e}_{3,1}^T \mathbf{F}_{23} \mathbf{e}_{2,1} = \mathbf{e}_{1,2}^T \mathbf{F}_{31} \mathbf{e}_{3,2} = 0 \tag{4}$$

which arise naturally from the trifocal plane.

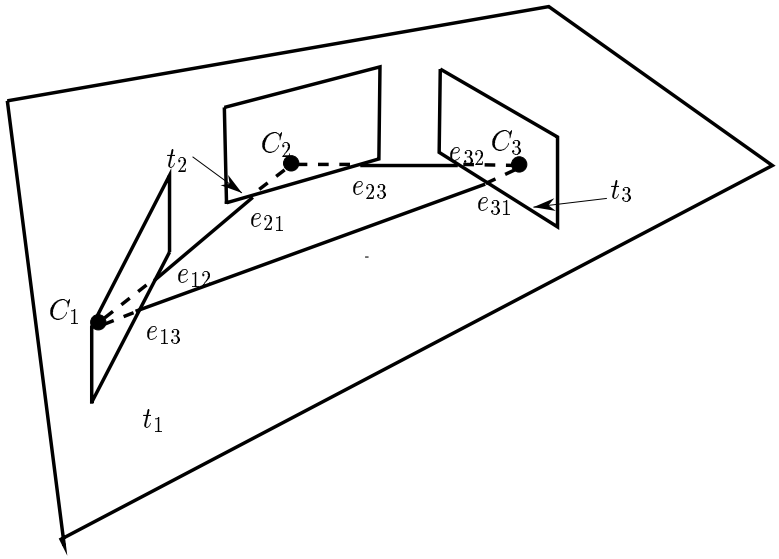

Figure 2: The trifocal geometry.

This has an important impact on the way we have to estimate the fundamental matrices when three views are available: very efficient and robust algorithms are now available to estimate the fundamental matrix between two views from point correspondences [ZDFL95, TZ97, Har95]. The constraints (4) mean that these algorithms cannot be used blindly to estimate the three fundamental matrices independently because the resulting matrices will not satisfy the constraints causing errors in further processes such as prediction.

Indeed, one of the important uses of the fundamental matrices in trifocal geometry is the fact that they in general allow to *predict* from two correspondences, say (m_1, m_2) where the point m_3 should be in the third image: it is simply at the intersection of the two epipolar lines represented by $\mathbf{F}_{13}\mathbf{m}_1$ and $\mathbf{F}_{23}\mathbf{m}_2$, when this intersection is well-defined.

It is not well-defined in two cases:

1. In the general case where the three optical centers are not aligned, when the 3D points lie in the trifocal plane (the plane defined by the three optical centers), the prediction with the fundamental matrices fails because, in the previous example both epipolar lines are equal to the trifocal line t_3.

2. In the special case where the three optical centers are aligned, the prediction with the fundamental matrices fails always since, for example, $\mathbf{F}_{13} \simeq \mathbf{F}_{23}$.

For those two reasons, as well as for the estimation problem mentioned previously, it is interesting to characterize the geometry of three views by another entity, the *trifocal tensor*.

The trifocal tensor is really meant at describing line correspondences and, as such, has been well-known under disguise in the part of the computer vision community dealing with the problem of structure from motion [SA90b, SA90a, WHA92] before it was formally identified by Hartley and Shashua [Har94a, Sha95].

3.2 The trifocal tensors

Let us consider three views, with projection matrices $\boldsymbol{\mathcal{P}}_n, n = 1, 2, 3$, a 3D line L with images l_n. Given two images l_j and l_k of L, L can be defined as the intersection (the meet) of the two planes $\boldsymbol{\mathcal{P}}_j^T \mathrm{l}_j$ and $\boldsymbol{\mathcal{P}}_k^T \mathrm{l}_k$:

$$\mathbf{L} \simeq \boldsymbol{\mathcal{P}}_j^T \mathrm{l}_j \vartriangle \boldsymbol{\mathcal{P}}_k^T \mathrm{l}_k$$

The vector $\mathbf{L}$ is the 6×1 vector of Plücker coordinates of the line L.

Let us write the right-hand side of this equation explicitly in terms of the row vectors of the matrices $\boldsymbol{\mathcal{P}}_j$ and $\boldsymbol{\mathcal{P}}_k$ and the coordinates of l_j and l_k:

$$\mathbf{L} \simeq (l_j^1 \boldsymbol{\Gamma}_j + l_j^2 \boldsymbol{\Lambda}_j + l_j^3 \boldsymbol{\Theta}_j) \vartriangle (l_k^1 \boldsymbol{\Gamma}_k + l_k^2 \boldsymbol{\Lambda}_k + l_k^3 \boldsymbol{\Theta}_k)$$

By expanding the meet operator in the previous equation, it can be rewritten in the following less compact form with the advantage of making the dependency on the projection planes of the matrices $\boldsymbol{\mathcal{P}}_j$ and $\boldsymbol{\mathcal{P}}_k$ explicit:

$$\mathbf{L} \simeq \mathrm{l}_j^T \begin{bmatrix} \boldsymbol{\Gamma}_j \vartriangle \boldsymbol{\Gamma}_k & \boldsymbol{\Gamma}_j \vartriangle \boldsymbol{\Lambda}_k & \boldsymbol{\Gamma}_j \vartriangle \boldsymbol{\Theta}_k \\ \boldsymbol{\Lambda}_j \vartriangle \boldsymbol{\Gamma}_k & \boldsymbol{\Lambda}_j \vartriangle \boldsymbol{\Lambda}_k & \boldsymbol{\Lambda}_j \vartriangle \boldsymbol{\Theta}_k \\ \boldsymbol{\Theta}_j \vartriangle \boldsymbol{\Lambda}_k & \boldsymbol{\Theta}_j \vartriangle \boldsymbol{\Lambda}_k & \boldsymbol{\Theta}_j \vartriangle \boldsymbol{\Theta}_k \end{bmatrix} \mathrm{l}_k \tag{5}$$

This equation should be interpreted as giving the Plücker coordinates of L as a linear combination of the lines defined by the meets of the projection planes of the perspective matrices $\boldsymbol{\mathcal{P}}_j$ and $\boldsymbol{\mathcal{P}}_k$, the coefficients being the products of the projective coordinates of the lines l_j and l_k.

The image l_i of L is therefore obtained by applying the matrix $\tilde{\boldsymbol{\mathcal{P}}}_i$ (defined as the 3x6 matrix $\tilde{\boldsymbol{\mathcal{P}}}_i^T = \left[\boldsymbol{\Lambda}_i^T \triangle \boldsymbol{\Theta}_i^T \; \boldsymbol{\Theta}_i^T \triangle \boldsymbol{\Gamma}_i^T \; \boldsymbol{\Gamma}_i^T \triangle \boldsymbol{\Lambda}_i^T\right]$, see [FM95a]) to the Plücker coordinates of L, hence the equation:

$$\mathrm{l}_i \simeq \tilde{\boldsymbol{\mathcal{P}}}_i(\boldsymbol{\mathcal{P}}_j^T \mathrm{l}_j \triangle \boldsymbol{\mathcal{P}}_k^T \mathrm{l}_k) \tag{6}$$

which is valid for $i \neq j \neq k$. Note that if we exchange view j and view k, we just change the sign of l_i and therefore we do not change l_i. A geometric interpretation of this is shown in figure 3. For convenience, we rewrite equation (6) in a more compact form:

$$\mathrm{l}_i \simeq \boldsymbol{\mathcal{T}}_i(\mathrm{l}_j, \mathrm{l}_k) \tag{7}$$

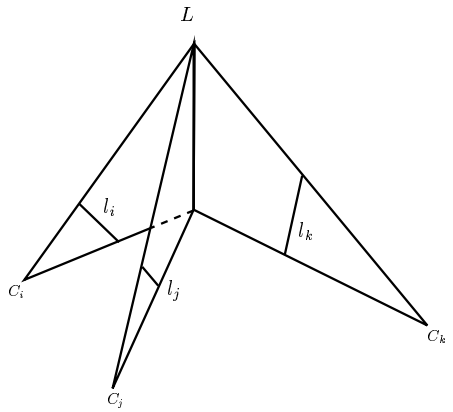

Figure 3: The line l_i is the image by camera i of the 3D line L intersection of the planes defined by the optical centers of the cameras j and k and the lines l_j and l_k, respectively.

This expression can be also put in a slightly less compact form with the advantage of making the dependency on the projection planes of the matrices $\boldsymbol{\mathcal{P}}_n, n = 1, 2, 3$ explicit:

$$\mathrm{l}_i \simeq \left[\mathrm{l}_j^T \mathbf{G}_i^1 \mathrm{l}_k \;\; \mathrm{l}_j^T \mathbf{G}_i^2 \mathrm{l}_k \;\; \mathrm{l}_j^T \mathbf{G}_i^3 \mathrm{l}_k\right]^T \tag{8}$$

This is, in the projective framework, the exact analog of the equation used in the work of Spetsakis and Aloimonos [SA90b] to study the structure from motion problem from line correspondences.

The three 3×3 matrices $\mathbf{G}_i^n, n = 1, 2, 3$ are obtained from equations (5) and (6):

$$\mathbf{G}_i^1 = \begin{bmatrix} [\mathbf{\Lambda}_i, \mathbf{\Theta}_i, \mathbf{\Gamma}_j, \mathbf{\Gamma}_k] & [\mathbf{\Lambda}_i, \mathbf{\Theta}_i, \mathbf{\Gamma}_j, \mathbf{\Lambda}_k] & [\mathbf{\Lambda}_i, \mathbf{\Theta}_i, \mathbf{\Gamma}_j, \mathbf{\Theta}_k] \\ [\mathbf{\Lambda}_i, \mathbf{\Theta}_i, \mathbf{\Lambda}_j, \mathbf{\Gamma}_k] & [\mathbf{\Lambda}_i, \mathbf{\Theta}_i, \mathbf{\Lambda}_j, \mathbf{\Lambda}_k] & [\mathbf{\Lambda}_i, \mathbf{\Theta}_i, \mathbf{\Lambda}_j, \mathbf{\Theta}_k] \\ [\mathbf{\Lambda}_i, \mathbf{\Theta}_i, \mathbf{\Theta}_j, \mathbf{\Gamma}_k] & [\mathbf{\Lambda}_i, \mathbf{\Theta}_i, \mathbf{\Theta}_j, \mathbf{\Lambda}_k] & [\mathbf{\Lambda}_i, \mathbf{\Theta}_i, \mathbf{\Theta}_j, \mathbf{\Theta}_k] \end{bmatrix}$$

Note that equation (6) allows us to predict the coordinates of a line l_i in image i given two images l_j and l_k of an unknown 3D line in images j and k, except in two cases:

1. When the two planes determined by l_j and l_k are identical i.e. when l_j and l_k are corresponding epipolar lines between views j and k. This is equivalent to saying that the 3D line L is in an epipolar plane of the camera pair (j, k). The meet that appears in equation (6) is then 0 and the line l_i is undefined, see figure 4. If L is not in an epipolar plane of the camera pair (i, j) then we can use the equation:

   $$\mathrm{l}_k \simeq \tilde{\boldsymbol{\mathcal{P}}}_k(\boldsymbol{\mathcal{P}}_i^T \mathrm{l}_i \vartriangle \boldsymbol{\mathcal{P}}_j^T \mathrm{l}_j)$$

   to predict l_k from the images l_i and l_j of L. If L is also in an epipolar plane of the camera pair (i, j) it is in the trifocal plane of the three cameras and prediction is not possible by any of the formulas such as (6). We will see in the next sections how the problem can be solved in the case of point correspondences.

2. When l_j and l_k are epipolar lines between views i and j and i and k, respectively. This is equivalent to saying that they are the images of the same optical ray in view i and that l_i is reduced to a point (see figure 5).

Except in those two cases, we have defined an application $\mathcal{T}_i$ from $\mathbb{P}^{*2} \times \mathbb{P}^{*2}$, the Cartesian product of two duals of the projective plane, into $\mathbb{P}^{*2}$. This application is represented by an application $\boldsymbol{\mathcal{T}}_i$ from $\mathbb{R}^3 \times \mathbb{R}^3$ into $\mathbb{R}^3$. This application is bilinear and antisymmetric and is represented by the three matrices $\mathbf{G}_i^n, n = 1, 2, 3$. It is called the *trifocal tensor* for view i. The properties of this application can be summarized in the following theorem:

Theorem 1 *The application $\mathcal{T}_i : \mathbb{P}^{*2} \times \mathbb{P}^{*2} \longrightarrow \mathbb{P}^{*2}$ is represented by the bilinear application $\boldsymbol{\mathcal{T}}_i$ such that $\boldsymbol{\mathcal{T}}_i(\mathrm{l}_j, \mathrm{l}_k) \simeq \tilde{\boldsymbol{\mathcal{P}}}_i(\boldsymbol{\mathcal{P}}_j^T \mathrm{l}_j \vartriangle \boldsymbol{\mathcal{P}}_k^T \mathrm{l}_k)$. $\boldsymbol{\mathcal{T}}_i$ has the following properties:*

1. *It is equal to* $\mathbf{0}$ *iff*
   - *(a) l_j and l_k are epipolar lines with respect to the ith view, or*
   - *(b) l_j and l_k are corresponding epipolar lines with respect to the pair (j, k) of cameras.*

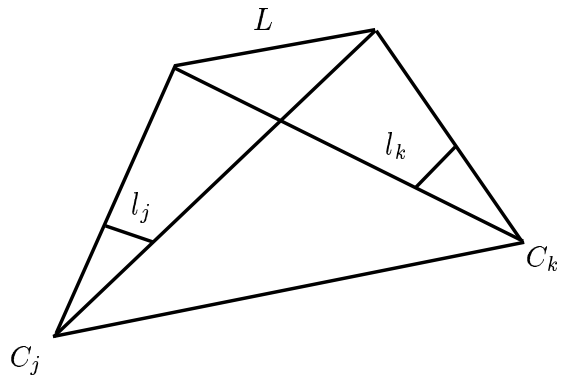

Figure 4: When l_j and l_k are corresponding epipolar lines, the two planes $\boldsymbol{\mathcal{P}}_j^T \mathrm{l}_j$ and $\boldsymbol{\mathcal{P}}_k^T \mathrm{l}_k$ are identical and therefore $\boldsymbol{\mathcal{T}}_i(\mathrm{l}_j, \mathrm{l}_k) = \mathbf{0}$.

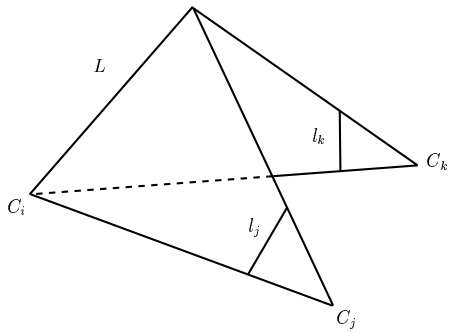

Figure 5: When l_j and l_k are epipolar lines with respect to view i, the line l_i is reduced to a point, hence $\boldsymbol{\mathcal{T}}_i(\mathrm{l}_j, \mathrm{l}_k) = \mathbf{0}$.

2. *Let l_{ki} be an epipolar line with respect to view i and l_{ik} (resp. l_{ji}) the corresponding epipolar line in view i (resp. in view j), then for all lines l_j in view j not equal to l_{ji}:* $\boldsymbol{\mathcal{T}}_i(\mathrm{l}_j, \mathrm{l}_{ki}) \simeq \mathrm{l}_{ik}$.

3. *Similarly, let l_{ji} be an epipolar line with respect to view i and l_{ij} (resp. l_{ki}) the corresponding epipolar line in view i (resp. in view k), then for all lines l_k in view k not equal to l_{ki}:* $\boldsymbol{\mathcal{T}}_i(\mathrm{l}_{ji}, \mathrm{l}_k) \simeq \mathrm{l}_{ij}$

A more pictorial view is shown in figure 6: the tensor is represented as a 3×3 cube, the three horizontal planes representing the matrices $\mathbf{G}_i^n, n = 1, 2, 3$. It can be thought of as a black box which takes as its input two lines, l_j and l_k and outputs a third one, l_i.

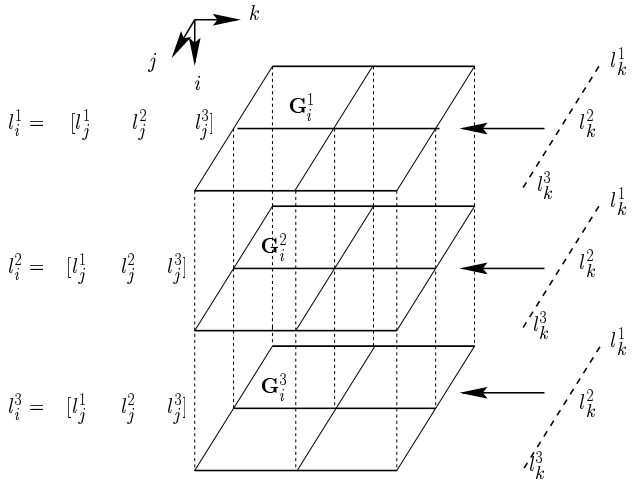

Figure 6: A three-dimensional representation of the trifocal tensor.

3.3 Algebraic and geometric properties of the trifocal tensors

The matrices $\mathbf{G}_i^n, n = 1, 2, 3$ have interesting properties which are closely related to the epipolar geometry of the views j and k. The nullspace of $\mathbf{G}_i^n$ is the set of lines l_k^n such that $\boldsymbol{\mathcal{T}}_i(\mathrm{l}_j, \mathrm{l}_k^n)$ has a zero in the n-th coordinate for all lines l_j. The corresponding lines l_i such that $\mathrm{l}_i = \boldsymbol{\mathcal{T}}_i(\mathrm{l}_j, \mathrm{l}_k^n)$ all go through the point represented by $\mathbf{e}_n, n = 1, 2, 3$ in the i-th retinal plane. This is true if and only if l_k^n is the image in the k-th retinal plane of the projection ray $\boldsymbol{\Lambda}_i \triangle \boldsymbol{\Theta}_i \, (n = 1)$, $\boldsymbol{\Theta}_i \triangle \boldsymbol{\Gamma}_i \, (n = 2)$ and $\boldsymbol{\Gamma}_i \triangle \boldsymbol{\Lambda}_i \, (n = 3)$. Theorem 1 shows that l_i is independent of l_j and represented by $\mathbf{e}_n \times \mathbf{e}_{i,k}$. We have proved the following proposition:

Proposition 3 *The matrices $\mathbf{G}_i^n$ are of rank 2 and their nullspaces are the three epipolar lines in the k-th retinal plane of the three projection rays of camera i. These three lines intersect at the epipole $e_{k,i}$.*

A similar reasoning applies to the matrices $\mathbf{G}_i^{nT}$:

Proposition 4 *The nullspaces of the matrices $\mathbf{G}_i^{nT}$ are the three epipolar lines, noted $l_j^n, n = 1, 2, 3$, in the j-th retinal plane of the three projection rays of camera i. These three lines intersect at the epipole $e_{j,i}$, see figure 7.*

The three corresponding epipolar lines for the pair (i, k) are obtained as $\boldsymbol{\mathcal{T}}_i(\mathbf{l}_j, \mathbf{l}_k^n), n = 1, 2, 3$ for any l_j not equal to l_j^n. They intersect at the epipole e_{ik}. Similarly, the three corresponding epipolar lines for the pair (i, j) are obtained as $\boldsymbol{\mathcal{T}}_i(\mathbf{l}_j^n, \mathbf{l}_k), n = 1, 2, 3$ for any l_k not equal to l_k^n. They intersect at the epipole $e_{i,j}$.

Algebraically, this implies that the three determinants $det(\mathbf{G}_i^n),\ n = 1, 2, 3$ are equal to 0. Another constraint implied by propositions 3 and 4 is that the 3×3 determinants formed with the three vectors in the nullspaces of the $\mathbf{G}_i^n, n = 1, 2, 3$ (resp. of the $\mathbf{G}_i^{nT}, n = 1, 2, 3$) are equal to 0. It turns out that the applications $\boldsymbol{\mathcal{T}}_i,\ i = 1, 2, 3$ satisfy other algebraic constraints which are also important in practice.

Let us assume for simplicity that $i = 1$ and show that the application $\mathcal{T}_1$, noted $\mathcal{T}$ satisfies 9 algebraic constraints of degree 6 which are defined as follows. Let $\mathbf{e}_n, n = 1, 2, 3$ be the canonical basis of $\mathbb{R}^3$ and let us consider the four lines $\boldsymbol{\mathcal{T}}(\mathbf{e}_{k_2}, \mathbf{e}_{k_3})$, $\boldsymbol{\mathcal{T}}(\mathbf{e}_{l_2}, \mathbf{e}_{k_3})$, $\boldsymbol{\mathcal{T}}(\mathbf{e}_{k_2}, \mathbf{e}_{l_3})$ and $\boldsymbol{\mathcal{T}}(\mathbf{e}_{l_2}, \mathbf{e}_{l_3})$ where the indexes k_2 and l_3 (resp. k_3 and l_3) are different. For example, if $k_2 = k_3 = 1$ and $l_2 = l_3 = 2$, the four lines are the images in camera 1 of the four 3D lines $\boldsymbol{\Gamma}_2 \triangle \boldsymbol{\Gamma}_3$, $\boldsymbol{\Lambda}_2 \triangle \boldsymbol{\Gamma}_3$, $\boldsymbol{\Gamma}_2 \triangle \boldsymbol{\Lambda}_3$ and $\boldsymbol{\Lambda}_2 \triangle \boldsymbol{\Lambda}_3$.

These four lines can be chosen in nine different ways satisfy an algebraic constraint which is detailed in the following theorem (proved in [FM95b]):

Theorem 2 *The bilinear mapping $\boldsymbol{\mathcal{T}}$ satisfies the 9 algebraic constraints of degree 6:*

$$\begin{aligned} &\mid \boldsymbol{\mathcal{T}}(\mathbf{e}_{k_2}, \mathbf{e}_{k_3})\boldsymbol{\mathcal{T}}(\mathbf{e}_{k_2}, \mathbf{e}_{l_3})\boldsymbol{\mathcal{T}}(\mathbf{e}_{l_2}, \mathbf{e}_{l_3}) \mid\mid \boldsymbol{\mathcal{T}}(\mathbf{e}_{k_2}, \mathbf{e}_{k_3})\boldsymbol{\mathcal{T}}(\mathbf{e}_{l_2}, \mathbf{e}_{k_3})\boldsymbol{\mathcal{T}}(\mathbf{e}_{l_2}, \mathbf{e}_{l_3}) \mid - \\ &\quad \mid \boldsymbol{\mathcal{T}}(\mathbf{e}_{l_2}, \mathbf{e}_{k_3})\boldsymbol{\mathcal{T}}(\mathbf{e}_{k_2}, \mathbf{e}_{l_3})\boldsymbol{\mathcal{T}}(\mathbf{e}_{l_2}, \mathbf{e}_{l_3}) \mid\mid \boldsymbol{\mathcal{T}}(\mathbf{e}_{k_2}, \mathbf{e}_{k_3})\boldsymbol{\mathcal{T}}(\mathbf{e}_{l_2}, \mathbf{e}_{k_3})\boldsymbol{\mathcal{T}}(\mathbf{e}_{k_2}, \mathbf{e}_{l_3}) \mid = 0 \end{aligned} \quad (9)$$

Referring to figure 6, what this theorem says is that if we take four vertical columns of the trifocal cube (shown as dashed lines in the figure) arranged in such a way that they form a prism with a square basis, then the expression (9) is equal to 0. Representing each line as $\boldsymbol{\mathcal{T}}_{.k_2k_3}$, etc . . . , we can rewrite equation (9) as:

$$\mid \boldsymbol{\mathcal{T}}_{.k_2k_3}\boldsymbol{\mathcal{T}}_{.k_2l_3}\boldsymbol{\mathcal{T}}_{.l_2l_3} \mid\mid \boldsymbol{\mathcal{T}}_{.k_2k_3}\boldsymbol{\mathcal{T}}_{.l_2k_3}\boldsymbol{\mathcal{T}}_{.l_2l_3} \mid - \mid \boldsymbol{\mathcal{T}}_{.l_2k_3}\boldsymbol{\mathcal{T}}_{.k_2l_3}\boldsymbol{\mathcal{T}}_{.l_2l_3} \mid\mid \boldsymbol{\mathcal{T}}_{.k_2k_3}\boldsymbol{\mathcal{T}}_{.l_2k_3}\boldsymbol{\mathcal{T}}_{.k_2l_3} \mid = 0 \quad (10)$$

The same kinds of relations hold for the other two principal directions of the cube (shown as solid lines of two different widths in the same figure):

Theorem 3 *The bilinear mapping $\mathcal{T}$ satisfies also the 18 algebraic constraints of degree 6:*

$$\mid \mathcal{T}_{k_2.k_3}\mathcal{T}_{k_2.l_3}\mathcal{T}_{l_2.l_3} \mid\mid \mathcal{T}_{k_2.k_3}\mathcal{T}_{l_2.k_3}\mathcal{T}_{l_2.l_3} \mid - \mid \mathcal{T}_{l_2.k_3}\mathcal{T}_{k_2.l_3}\mathcal{T}_{l_2.l_3} \mid\mid \mathcal{T}_{k_2.k_3}\mathcal{T}_{l_2.k_3}\mathcal{T}_{k_2.l_3} \mid = 0 \tag{11}$$

and:

$$\mid \mathcal{T}_{k_2k_3.}\mathcal{T}_{k_2l_3.}\mathcal{T}_{l_2l_3.} \mid\mid \mathcal{T}_{k_2k_3.}\mathcal{T}_{l_2k_3.}\mathcal{T}_{l_2l_3.} \mid - \mid \mathcal{T}_{l_2k_3.}\mathcal{T}_{k_2l_3.}\mathcal{T}_{l_2l_3.} \mid\mid \mathcal{T}_{k_2k_3.}\mathcal{T}_{l_2k_3.}\mathcal{T}_{k_2l_3.} \mid = 0 \tag{12}$$

Note that these 27 constraints, and the 5 constraints implied by propositions 3 and 4 are not algebraically independent but they must be taken into account when estimating the trifocal tensor as will be demonstrated in the experimental part of this paper.

3.4 Recovering a coherent set of fundamental matrices from the trifocal tensors

We can then show that the epipolar geometry between the three views, i.e. the fundamental matrices, can be recovered from the $\mathcal{T}_i, i = 1, 2, 3$ and that these fundamental matrices are coherent, i.e. they satisfy equations (4).

Proposition 5 *The two fundamental matrices $\mathbf{F}_{ij}$ and $\mathbf{F}_{ik}$ can be recovered uniquely from $\mathcal{T}_i$.*

Proof : We can recover the two fundamental matrices $\mathbf{F}_{ij}$ and $\mathbf{F}_{ik}$ from $\mathcal{T}_i$ since for each of them we know the two epipoles and three corresponding epipolar lines, see propositions 3 and 4. □

The missing link is $\mathbf{F}_{jk}$. In order to recover it, the simplest is to compute $\mathcal{T}_j$ as explained in theorem 4 and apply the same technique as above.

The question of computing $\mathcal{T}_j$ from $\mathcal{T}_i$, $j \neq i$ has already been addressed partially in [SW95, AS96]. We give a somewhat simpler result here through the following theorem:

Theorem 4 *The relation $\mathcal{T}_j$ is obtained from $\mathcal{T}_i, i \neq j$, from the following relations:*

$$\left[\mathbf{G}_j^{1(m)}\ \mathbf{G}_j^{2(m)}\ \mathbf{G}_j^{3(m)}\right] = \left[\lambda_1\mathbf{G}_i^{2(m)} \times \mathbf{G}_i^{3(m)}\ \lambda_2\mathbf{G}_i^{3(m)} \times \mathbf{G}_i^{1(m)}\ \lambda_3\mathbf{G}_i^{1(m)} \times \mathbf{G}_i^{2(m)}\right]$$

where $\mathbf{G}_i^{k(m)}, k, m = 1, 2, 3$ *(resp.* $\mathbf{G}_j^{k(m)}$*) is the* m*th column vector of matrix* $\mathbf{G}_i^k$ *(resp.* $\mathbf{G}_j^k$*) and the* $\lambda_n,\ n = 1, 2, 3$ *are given, for example, by* $\lambda_1 = \mid \boldsymbol{\mathcal{T}}_{.31}\boldsymbol{\mathcal{T}}_{.22}\boldsymbol{\mathcal{T}}_{.32} \mid\mid \boldsymbol{\mathcal{T}}_{.31}\boldsymbol{\mathcal{T}}_{.23}\boldsymbol{\mathcal{T}}_{.33} \mid$, $\lambda_2 = \mid \boldsymbol{\mathcal{T}}_{.31}\boldsymbol{\mathcal{T}}_{.21}\boldsymbol{\mathcal{T}}_{.32} \mid\mid \boldsymbol{\mathcal{T}}_{.31}\boldsymbol{\mathcal{T}}_{.23}\boldsymbol{\mathcal{T}}_{.33} \mid$ *and* $\lambda_3 = \mid \boldsymbol{\mathcal{T}}_{.31}\boldsymbol{\mathcal{T}}_{.22}\boldsymbol{\mathcal{T}}_{.32} \mid\mid \boldsymbol{\mathcal{T}}_{.31}\boldsymbol{\mathcal{T}}_{.21}\boldsymbol{\mathcal{T}}_{.33} \mid$

A similar result can be obtained for changing from $\mathcal{T}_i$ to $\mathcal{T}_k$.

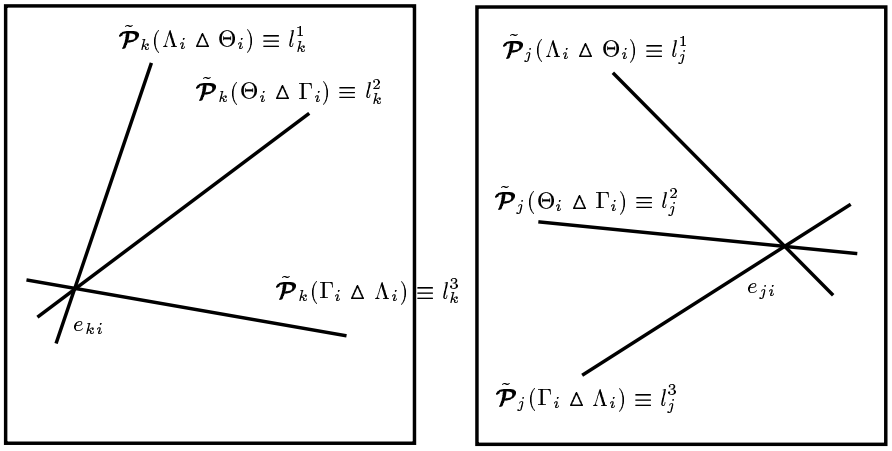

Figure 7: The lines l_j^n (resp. l_k^n), $n = 1, 2, 3$ in the nullspaces of the matrices $\mathbf{G}_i^{nT}$ (resp. $\mathbf{G}_i^n$) are the images of the three projection rays of camera i. Hence, they intersect at the epipole $e_{j,i}$ (resp. $e_{k,i}$). The corresponding epipolar lines in camera i are obtained as $\mathcal{T}_i(l_j^n, l_k)$ (resp. $\mathcal{T}_i(l_j, l_k^n)$) for $l_k \neq l_k^n$ (resp. $l_j \neq l_j^n$).

The important point to note here is that the resulting set of fundamental matrices is coherent, i.e. they satisfy equations (4).

The perspective projection matrices of the three views can then be recovered uniquely, for example through the results of Luong and Viéville, [LV94].

3.5 The trilinearities

So far, we have been working only with lines. It is easy to derive algebraic relations between corresponding pixels which are *necessary* conditions for triples of points to be in correspondence. By considering the three points $e_k,\ k = 1, 2, 3$, we can in each image i consider the three lines represented by $\mathbf{m}_i \times \mathbf{e}_k$, thereby generating three groups of nine trilinear relations in the coordinates of the three pixels m_i, e.g. $\mathbf{m}_i^T \boldsymbol{\mathcal{T}}_i(\mathbf{m}_j \times \mathbf{e}_m, \mathbf{m}_k \times \mathbf{e}_n)$ For a

fixed view i, it is easy to convince oneself that only four of these nine relations are linearly independent, since the three vectors $\mathbf{m} \times \mathbf{e}_m,\ m = 1, 2, 3$ are linearly dependent. Thus we can talk of the set of the 12 trilinear relations and of the three groups (one per view) of four trilinear relations. Note that since each group corresponds to a view and we have seen in proposition 4 that the three trifocal tensors are not algebraically independent, the trilinear relations between groups are algebraically dependent. The study of this dependence was done for example in [FM95b].

We also showed there the following proposition:

Proposition 6 *A triple of image points (m_1, m_2, m_3) is in correspondence iff the coordinates satisfy the four trilinear constraints of any of the three groups.*

As shown by Hartley and Shashua, the trilinear constraints can be used to transfer points from to views to the third. More precisely, if m_j and m_k are two corresponding pixels in views j and k, i.e. satisfying the epipolar constraint, then the twelve lines defined in image i by the twelve trilinear constraints intersect at a single point m_k.

We mentioned above that this process failed with the fundamental matrixes in two cases. It is easy to show that this does not happen in the case of the trilinear constraints. In detail: We saw previously that if the point M is in the trifocal plane, its image in one of the cameras, for example m_3, cannot be predicted from its images m_1 and m_2 in the other two cameras. The traces of this plane in the three retinal planes are the three trifocal lines $t_i,\ i = 1, 2, 3$. But the trilinear relations can be used to compute the position of m_3 on this line. To show this, let us assume that $\mathbf{m}_1 = a_1\,\mathbf{e}_{1,2} + b_1\,\mathbf{e}_{1,3}$, $\mathbf{m}_2 = a_2\,\mathbf{e}_{2,3} + b_2\,\mathbf{e}_{2,1}$, $\mathbf{m}_3 = a_3\,\mathbf{e}_{3,1} + b_3\,\mathbf{e}_{3,2}$ are on the trifocal lines. We now prove the following proposition:

Proposition 7 *The restriction of any of the 12 trilinearities to the trifocal plane is either identically 0 or proportional to $a_1a_2a_3 + b_1b_2b_3$.*

We saw in section 3.1 that if the optical centers of the three cameras were aligned, it was not possible to transfer a point, say from views 2 and 3 to view 1, because the two epipolar lines were identical. This problem does not occur if we use the trifocal tensor $\mathcal{T}_1$.

Indeed, let m_2 and m_3 be two pixels in images 2 and 3 satisfying the epipolar constraint. Let l_2 and l_3 be two lines going through the points m_2 and m_3, respectively. Then, under the condition of the next proposition, the line $\mathcal{T}(\mathrm{l}_2, \mathrm{l}_3)$ intersects the epipolar line of m_2 in image 1, which is identical to the epipolar line of m_3, at the point m_1 corresponding to (m_2, m_3).

Proposition 8 *Let m_2 and m_3 be two points in images 2 and 3 such that m_2 and m_3 satisfy the epipolar constraint. Let l_2, l_3 be two lines going through the points m_2 and m_3,*

respectively, and such that $l_k,\ k = 2, 3$ *does not go through the epipole* $e_{k,1}$*. Then, the point of intersection* m_1 *of the line* $\mathcal{T}(\mathrm{l}_2, \mathrm{l}_3)$ *with the epipolar line in image 1 of* m_2 *(which is identical to the epipolar line of* m_3*) is well defined and is the point in image 1 corresponding to the pair* m_2, m_3*).*

These two propositions are the reasons why it is better to use the trifocal tensor than the fundamental matrices when three views are available.

In the next section, we turn to the problem of estimating the trilinear from point and line correspondences in such a way that the result is guaranteed to satisfy all the algebraic constraints described in propositions 3 and 4 and in theorems 2 and 3. We will call these constraints the *trilinear constraints*.

3.6 Minimal parameterization of the trifocal tensors

Let us see how the constraints described by the 9 equations (10) can be used to parameterize the trifocal tensors with 18 parameters.

To do so, we choose two of the constraints of (10), for example $k_2 = l_2 = 1,\ k_3 = l_3 = 2$ and $k_2 = l_2 = 2,\ k_3 = l_3 = 3$. Notice that $\mathcal{T}_{\cdot 22}$ is the only vector that appears in both constraints. In general, the three vectors $\mathcal{T}_{\cdot 23}$, $\mathcal{T}_{\cdot 32}$ and $\mathcal{T}_{\cdot 33}$ are linearly independent, so that $\mathcal{T}_{\cdot 22}$ can be expressed in the basis they form: $\mathcal{T}_{\cdot 22} = \alpha_1 \mathcal{T}_{\cdot 23} + \alpha_2 \mathcal{T}_{\cdot 32} + \alpha_3 \mathcal{T}_{\cdot 33}$ Equation (10) then implies that $\alpha_1 + \alpha_2 \alpha_3 = 0$ It is then easy to parameterize the set of the vectors $\mathcal{T}_{\cdot 22}$ that verify the constraint given by $k_2 = l_2 = 2,\ k_3 = l_3 = 3$: $\mathcal{T}_{\cdot 22} = \mathbf{K}\,[-pq\ p\ q]^T$, where $\mathbf{K}$ is the matrix $[\mathcal{T}_{\cdot 23}\ \mathcal{T}_{\cdot 32}\ \mathcal{T}_{\cdot 33}]$. Reporting this value of $\mathcal{T}_{\cdot 22}$ in the constraint given by $k_2 = l_2 = 1,\ k_3 = l_3 = 2$, we obtain a polynomial equation P of total degree 4, and of degree 2 in p and q.

The trifocal tensor can thus be parameterized by 18 parameters. Since there is a global scale factor, we can fix for example the first coordinate of $\mathcal{T}_{\cdot 11}$ to be 1. All the vectors $\mathcal{T}_{\cdot 11}$, $\mathcal{T}_{\cdot 12}$, $\mathcal{T}_{\cdot 21}$, $\mathcal{T}_{\cdot 23}$, $\mathcal{T}_{\cdot 32}$ and $\mathcal{T}_{\cdot 33}$ are thus described by 17 parameters. Adding p as the 18th parameter gives a parameterization of the trifocal tensor. Indeed, then $\mathcal{T}_{\cdot 22}$ can recovered using the polynomial P. Then, as shown in [FM95a], $\mathcal{T}_{\cdot 13}$ and $\mathcal{T}_{\cdot 31}$ can be recovered up to a scale factor by the formula $(\mathcal{T}_{\cdot 12} \times (\mathcal{T}_{\cdot 22} \times (\mathcal{T}_{\cdot 23}) \times (\mathcal{T}_{\cdot 33} \times \mathcal{T}_{\cdot 32}))) \times (\mathcal{T}_{\cdot 23} \times (\mathcal{T}_{\cdot 22} \times (\mathcal{T}_{\cdot 12}) \times (\mathcal{T}_{\cdot 21} \times \mathcal{T}_{\cdot 11})))$ and $(\mathcal{T}_{\cdot 21} \times (\mathcal{T}_{\cdot 22} \times (\mathcal{T}_{\cdot 32}) \times (\mathcal{T}_{\cdot 33} \times \mathcal{T}_{\cdot 23}))) \times (\mathcal{T}_{\cdot 32} \times (\mathcal{T}_{\cdot 22} \times (\mathcal{T}_{\cdot 21}) \times (\mathcal{T}_{\cdot 12} \times \mathcal{T}_{\cdot 11})))$, respectively. Each of the remaining scale factors can be recovered using the constraints given by $k_2 = 2,\ k_3 = 1,\ l_2 = 3,\ l_3 = 2$ and $k_2 = 1,\ k_3 = 2,\ l_2 = 2,\ l_3 = 3$, respectively.

It is important to note that, since we have to solve the polynomial P to recover the trifocal tensor, each vector of 18 parameters gives in fact two trifocal tensors. Both these two tensors are valid: i.e. they verify all the constraints of trifocal tensors. Thus, there is,

in general, no way to distinguish between those and both have to be considered: only some extra information can be of some help.

4 Methods of estimation

We now assume that we have a set of three corresponding images from which triplets of corresponding points (one point per image) have been extracted. In this section, we study how the estimation of the trifocal tensor can be done in such a way that the trilinear constraints can be taken into account.

As for the fundamental matrix [ZDFL95], estimating the trifocal tensor is done in two steps. First, we obtain an initial estimate $\tilde{\mathcal{T}_{i0}}$ that does satisfy the trilinear constraints and second, starting from this estimate, we find a new estimate $\tilde{\boldsymbol{\mathcal{T}}}_i$ that is close to $\tilde{\mathcal{T}_{i0}}$ and satisfies all the trilinear constraints. Once this is done, the resulting tensor can be parameterized as shown in section 3.6 and non-linear minimization can be used to refine it with respect to the image data.

4.1 Initial estimate

It is well known [Har94a, Sha95] that an approximation to the trilinear tensor can be estimated linearly. To do so, we neglect the trilinear constraints and take all the 27 coefficients of the trifocal tensor as unknowns.

We consider pairs of points in each image. Every such pair defines a line and, of course, pairs of corresponding points give corresponding lines in the three images. We then use eq. (7) to get the following equation:

$$\boldsymbol{\mathcal{T}}_i(\mathrm{l}_j, \mathrm{l}_k) \times \mathrm{l}_i \ , \tag{13}$$

which shows that each triplet of corresponding lines provides two linear equations in the coefficients of $\boldsymbol{\mathcal{T}}_i$. Consequently, as soon as 14 line matches are available (if they are in general position), the trifocal tensor can be estimated using equations (13). In practice, we have much more line matches and we use singular value decomposition to get the least squares solution of the system generated by equations (13). Of course, the coordinates of points and/or lines have to be scaled properly [Har95] in order to get good numerical results. The result of this step is a tensor $\tilde{\mathcal{T}_{i0}}$ that gives good estimates for the projection of a line in image i given its projections in images j and k.

4.2 Enforcing the constraints

As we will see in section 5, due to the uncertainty on the positions of the points in the images, the tensor $\tilde{\mathcal{T}}_{i0}$ does not in general satisfy the trilinear constraints. Thus, in order to obtain a trifocal tensor that can be parameterized by 18 parameters, we have first to enforce the trilinear constraints.

To do so, we have developed a minimization scheme based on those constraints. To describe it, we define two different sets of constraints. The first set, S_1, consists of the 3 algebraic constraints that state that the ranks of the matrices $\mathbf{G}_i^n,\ n = 1 \ldots 3$ are zero. The second one, S_2, consists of the 9 algebraic constraints described by formula (11).

As we will see in section 5, imposing these constraints only is enough to impose all the other constraints described in section 3.3. Let us call $\mathbf{p}$, the vector of the 27 coefficients of the trifocal tensor $\boldsymbol{\mathcal{T}}_i$. As for $\boldsymbol{\mathcal{T}}_i$, $\mathbf{p}$ is defined up to a global scale factor. To avoid this problem, we add the constraint that $\|\mathbf{p}\| = 1$. The core of the minimization is a least-squares minimization of the algebraic equations of set S_2 with respect to $\mathbf{p}$. This can be done easily by linearizing those equations and applying a Newton step. If we call $C_2(\mathbf{p})$, the vector of the values of the S_2-constraints at $\mathbf{p}$ and $J_2(\mathbf{p})$ the Jacobian matrix of those constraints with respect to $\mathbf{p}$, this step can be written as:

$$J_2(\mathbf{p})d\mathbf{p} = -C_2(\mathbf{p}), \tag{14}$$

where $d\mathbf{p}$ is the unknown and represents the increment to apply to $\mathbf{p}$ in order to obtain a refined estimate $\mathbf{p} + d\mathbf{p}$. Since J_2 is a 9x27 matrix, we know that this linear system is at most of rank 9 (actually it can be shown that it is in general of rank 8 and that there is a linear relation relating all those constraints), so that the system (14) has infinitely many solutions: of those, we select the one that is of minimal norm. To avoid steps that will increase the values of the constraints and to make the minimization slightly more efficient, the new point is not taken as just $\mathbf{p} + d\mathbf{p}$ but is merely the point on the segment $[\mathbf{p}, \mathbf{p} + d\mathbf{p}]$ that minimizes the sum of the squares of the values of the C_2 constraints.

Notice that because each of the constraints C_2 is globally homogeneous (of degree 6), we know (using the Euler relation) that the rows of J_2 should be orthogonal to $\mathbf{p}$ when the constraints are satisfied. Because of this, the step $d\mathbf{p}$ remains in the tangent plane at $\mathbf{p}$ to the sphere defined by $\|\mathbf{p}\| = 1$, which means that this iteration step is compatible with that constraint (i.e. $\|\mathbf{p} + d\mathbf{p}\| \approx 1$ since $\|d\mathbf{p}\|$ is usually small, which means that the minimization keeps $\mathbf{p}$ away from the singularity $\mathbf{p} = 0$). Of course, $\|\mathbf{p}\| = 1$ is imposed at the beginning of every new iteration.

Unfortunately, this step alone is not sufficient to impose the S_1 constraints and practically, we have noticed that after an initial improvement the scheme stays usually trapped into a local minimum from which it does not escape. Thus, before it is considered any

new configuration $\mathbf{p}$ is first modified to enforce the S_1-constraints. As for fundamental matrices, this is done by taking the singular value decomposition of each of the $\mathbf{G}_i^n$, and reconstructing it from the decomposition after having set its smallest singular value to be zero.

4.3 Non-linear refinement of the trifocal tensor

Let us call L^i_{jk} the line joining the points j and k in image i. Furthermore, we note L^i the set of all the lines of image i and $d(m, L)$ the Euclidean distance of the point m to the line L. Using these notations, it is easy to describe a new criterion C that is reminiscent of the one that is commonly used for estimating the fundamental matrices:

$$C = \sum_{i\in\{1,2,3\}} \sum_{L^i_{jk}\in L^i} d(j, \boldsymbol{\mathcal{T}}_i(L^{i'}_{jk}, L^{i''}_{jk})) + d(k, \boldsymbol{\mathcal{T}}_i(L^{i'}_{jk}, L^{i''}_{jk})), \tag{15}$$

where i' and i'' are such that $\{i, i', i''\} = \{1, 2, 3\}$. Unlike the criterion optimized with the linear method, this criterion does not privilege any of the views and gives the same weight over each feature of the views.

Since we now have a tensor that satisfies the constraints of section 3.3, it can be parameterized as described in section 3.6. This vector of parameters can be used to optimize C using a Levenberg-Marquardt method. But in order to do so, we still have to solve one small problem: remember that each vector of parameters is not giving one but two trifocal tensors that cannot be distinguished in general. But we have now access to some new information, the points and lines in the three images, and those can be used to select the proper trifocal tensor. To do so, we just select the tensor that gives the smallest value for the criterion C!

5 Experimental results

In order to compare the inherent merits of each method, we have conducted a set of experiments with three triplets of real images referred to as **Triplet 1**, **Triplet 2**, and **Triplet 3**. One example of each image is shown in figure 8. Each triplet contains about 30 point matches that are used to estimate the trifocal tensor. Those points were obtained using an interactive tool and are thus of good quality (reasonable accuracy, of the order of .5 pixels, and no false matches). In order to test the robustness of the various algorithms to pixel noise, we have also rounded to the closest integer value the pixels coordinates for the second and third triplets which are referred to as **Triplet 2'** and **Triplet 3'**.

With each of these data sets, we have used several different methods to estimate the three trifocal tensors $\boldsymbol{\mathcal{T}}_1$, $\boldsymbol{\mathcal{T}}_2$ and $\boldsymbol{\mathcal{T}}_3$ that are associated to the triplet of views:

Figure 8: One image excerpted of each triplet used for the experiments. In each case, the point matches used for the experiments are shown. Notice that the first of these triplets is actually a triplet of mosaics made from different pictures. Starting from top to bottom and from left to right these experimental sets are called **Triplet 1**, **Triplet 2** and **Triplet 3** respectively.

- One of the trifocal tensors $\mathcal{T}_i,\ i = 1..3$ is computed using the linear algorithm [Sha94a, Har94b] and the formula of theorem 4 is used to obtain the two other trifocal tensors. This set of experiments is called $\mathbf{Lin}_i$.
- The second case is similar to the previous one but the constraints were enforced as described in section 4 before applying the change of view. This set of experiments is denoted by $\mathbf{Cst}_i$ hereafter.
- In the third case, we estimate the three trifocal tensors independently using the linear algorithm. This set of experiments is called $\mathbf{Data}$.
- The last case uses the full strategy described in section 4: One of the trifocal tensors $\mathcal{T}_i,\ i = 1..3$ is computed using the linear algorithm, then the constraints on its coefficients are enforced and the resulting tensor is used as the starting point in a non-linear optimization procedure. This set of experiments is called $\mathbf{NonLin}_i$ hereafter.

For each of these cases, we have used various different criteria to assess the quality of the resulting tensors from two different points of view: the coherence of the epipolar geometry in the three views and the fidelity to the data. The quintessence of these results is presented in the next three sections.

5.1 Constraints

The first measure of the quality of the trifocal tensors is to see how well the constraints of section 3.3 are verified. To do so we display in table 1, the sum of the squares of the normalized value of the 27 criteria for each of the experiments. By normalized, we mean that the squared value of the criterion is divided by the sum of the squares of the two components that are involved in the constraint: each constraint being of the form $x + y = 0$, we display the value of $\frac{(x+y)^2}{x^2+y^2}$. The table clearly shows that our reprojection scheme that uses only a subset of 9 of the trilinear constraints is in fact enforcing all of them (row $\mathbf{Cst}_1$).

5.2 Coherence of the trifocal geometry

To test the coherence of the epipolar geometry, we have used the constraints described in Eq. (4). To do so, the three fundamental matrices that are needed are computed from the trifocal tensor using the method described in [SW95]. The six epipoles are then computed from those matrices. For each of the four estimation techniques and for each triplet of views, we show two types of results: first the angle, in degrees, of the two vectors representing the epipole and the corresponding trifocal line (in fact 90 degrees minus the angle).

	Triplet 1	**Triplet 2**	**Triplet 2'**	**Triplet 3**	**Triplet 3'**
$\mathbf{Lin}_1$	14.6	4.9	13.4	10.9	15.5
$\mathbf{Lin}_i$	14.6	0.69	5.6	4.4	9.8
$\mathbf{Cst}_1$	$1.2e^{-27}$	$5.1e^{-27}$	$5.0e^{-28}$	$1.8e^{-28}$	$4.2e^{-27}$

Table 1: This table shows a measure of how well the constraints described in section 3.3 are verified. As shown by the third row, the values after reprojection can be safely regarded as being zero. The first and the second rows show the results for linear methods. For the results of the second row, the linear estimate that gives the best error has been selected. These two rows clearly show that linear methods are leading to values that are far from being null. Results for the experiment $\mathbf{NonLin}_i$ are not shown here as the constraints are trivially satisfied since we use a parameterization of the trifocal tensor.

	Triplet 1		**Triplet 2**		**Triplet 2'**		**Triplet 3**		**Triplet 3'**	
	Angle	Dist.	Angle	Dist.	Angle	Dist.	Angle	Dist.	Angle	Dist.
Data	0.2	1.5	0.03	0.6	0.01	0.3	21.5	15.2	2.3	4.5
	1.7	4.7	0.01	0.06	0.06	1.2	1.7	11.2	0.9	1.2
	0.1	8.1	0.4	0.4	0.05	0.5	1.6	14.1	1.1	4.7
$\mathbf{Lin}_1$	0.1	45.	0.01	0.2	0.04	1.6	0.6	0.7	3.4	16.3
	0.4	1.1	0.01	0.1	0.02	0.6	0.6	4.3	0.8	1.0
	0.3	18.	0.0	0.02	0.01	0.3	0.9	34.4	0.5	4.7
$\mathbf{Cst}_1$	0.0	$2.1e^{-11}$	0.0	$3.7e^{-13}$	0.0	$2.6e^{-12}$	0.0	$2.1e^{-11}$	0.0	$1.2e^{-11}$
	0.0	$8.9e^{-12}$	0.0	$3.1e^{-13}$	0.0	$8.3e^{-14}$	0.0	$1.0e^{-11}$	0.0	$2.3e^{-12}$
	0.0	$2.1e^{-13}$	0.0	$3.6e^{-15}$	0.0	$1.9e^{-12}$	0.0	$1.6e^{-11}$	0.0	$6.3e^{-12}$
$\mathbf{NLin}_1$	0.0	$4.0e^{-12}$	0.0	$3.7e^{-11}$	0.0	$3.8e^{-10}$	0.0	$2.1e^{-13}$	0.0	$5.1e^{-12}$
	0.0	$1.6e^{-11}$	0.0	$3.3e^{-11}$	0.0	$3.1e^{-10}$	0.0	$5.3e^{-13}$	0.0	$1.1e^{-12}$
	0.0	$2.4e^{-12}$	0.0	$8.1e^{-12}$	0.0	$9.9e^{-11}$	0.0	$1.7e^{-13}$	0.0	$2.0e^{-12}$

Table 2: This table shows various measurements of how well the trifocal geometry is satisfied. For each experiment, there are three rows, each of which represents one of the three constraints of Eq. (4).

For example, in the first constraint of equations (4), the angle is between the 3-D vectors $\mathbf{e}_{2,3}$ and $\mathbf{F}_{12}\mathbf{e}_{1,3}$. Second, we show the distance, in pixels, between the epipole and the corresponding trifocal line. For the same example as before, it is the distance between the point $e_{2,3}$ and the line t_2. The results shown in rows $\mathbf{Cst}_1$ and $\mathbf{NLin}_1$ clearly show the importance of enforcing the constraints on the coefficients of the tensors.

	Triplet 1		**Triplet 2**		**Triplet 2'**		**Triplet 3**		**Triplet 3'**	
	Average	Max.	Average	Max.	Average	Max.	Average	Max.	Average	Max.
Data	$1.4e^{-3}$	0.7	$1.6e^{-3}$	0.5	$7.9e^{-3}$	1.8	$4.3e^{-4}$	0.2	$2.2e^{-3}$	2.2
	$4.3e^{-4}$	0.1	$4.3e^{-3}$	1.2	$1.8e^{-2}$	5.4	$2.3e^{-4}$	0.1	$7.0e^{-3}$	8.3
	$3.8e^{-3}$	2.4	$4.7e^{-4}$	0.2	$2.3e^{-3}$	4.4	$1.4e^{-3}$	1.6	$1.5e^{-3}$	1.2
$\mathbf{Lin}_1$	$1.4e^{-3}$	0.7	$1.6e^{-3}$	0.5	$7.9e^{-3}$	1.8	$4.3e^{-4}$	0.2	$2.2e^{-3}$	2.2
	0.30	127.	$9.7e^{-3}$	2.6	$4.9e^{-2}$	12.	$1.3e^{-1}$	99	$4.7e^{-2}$	27
	1.54	620.	$4.3e^{-3}$	1.1	$2.6e^{-2}$	5.5	$3.5e^{-1}$	250	0.3	170
$\mathbf{Lin}_i$	$5.7e^{-2}$	21.	$1.6e^{-3}$	0.5	$7.9e^{-3}$	1.8	$9.9e^{-2}$	38	0.16	105
	$4.3e^{-4}$	0.1	$9.7e^{-3}$	2.6	$4.9e^{-2}$	12.	$1.1e^{-2}$	5.5	$6.8e^{-3}$	4.0
	1.07	555.	$4.3e^{-3}$	1.1	$2.6e^{-2}$	5.5	$1.4e^{-3}$	1.6	$1.5e^{-3}$	1.2
$\mathbf{Cst}_1$	$1.1e^{-2}$	4.7	$1.5e^{-3}$	0.4	$3.3e^{-2}$	7.8	$9.6e^{-4}$	0.5	$7.9e^{-3}$	7.8
	$1.9e^{-3}$	1.0	$5.9e^{-3}$	1.8	$1.2e^{-2}$	3.2	$1.7e^{-3}$	1.4	$1.6e^{-3}$	1.3
	$4.1e^{-2}$	26.	$5.6e^{-4}$	0.2	$4.1e^{-3}$	0.9	$1.5e^{-3}$	1.4	$1.0e^{-2}$	9.3
$\mathbf{Cst}_i$	$1.2e^{-2}$	7.3	$1.8e^{-3}$	0.6	$1.6e^{-2}$	4.5	$1.2e^{-3}$	0.6	$4.2e^{-3}$	2.7
	$3.3e^{-3}$	2.8	$3.1e^{-3}$	0.8	$2.2e^{-2}$	6.9	$7.1e^{-4}$	0.3	$2.4e^{-3}$	1.7
	$1.8e^{-2}$	11.	$3.8e^{-4}$	0.1	$2.3e^{-3}$	0.6	$1.2e^{-3}$	0.8	$3.3e^{-3}$	2.7
$\mathbf{NLin}_1$	$2.8e^{-3}$	0.9	$5.5e^{-4}$	0.1	$7.1e^{-3}$	1.6	$3.3e^{-4}$	0.1	$6.8e^{-4}$	0.2
	$3.1e^{-3}$	1.4	$6.9e^{-4}$	0.2	$8.8e^{-3}$	2.0	$1.7e^{-4}$	0.1	$5.6e^{-4}$	0.4
	$6.7e^{-3}$	3.1	$6.6e^{-4}$	0.1	$2.5e^{-3}$	0.4	$2.4e^{-4}$	0.1	$8.1e^{-4}$	0.3
$\mathbf{NLin}_i$	$2.2e^{-3}$	1.0	$5.5e^{-4}$	0.1	$3.4e^{-3}$	0.8	$3.2e^{-4}$	0.1	$6.8e^{-4}$	0.2
	$2.3e^{-3}$	1.8	$6.9e^{-4}$	0.2	$7.2e^{-3}$	1.5	$1.6e^{-4}$	0.1	$5.6e^{-4}$	0.4
	$2.2e^{-3}$	0.7	$6.6e^{-4}$	0.1	$2.5e^{-3}$	0.5	$2.3e^{-4}$	0.1	$8.1e^{-4}$	0.3

Table 3: Prediction errors in pixels for all experiments. The average error is small for all methods even though the linear methods sometimes fail miserably and more importantly, unpredictably, for a few matches, as shown in the column showing the maximum error. The rows labelled $\mathbf{Lin}_i$, $\mathbf{Cst}_i$ and $\mathbf{NLin}_i$ display the best results for each of the three classes of algorithms.

5.3 Fidelity to the data

Table 3 shows all the error measurements, in pixels, that were obtained for each different method and for each triplet of views. For each experiment and for each view of the triplet, for each image i two different errors are shown: those are respectively the average and maximal distance $d(j, \boldsymbol{\mathcal{T}}_i(L^{i'}_{jk}, L^{i''}_{jk}))$. As shown in the table, the nonlinear methods are much more robust and reliable, bringing the errors down for *all* correspondences.

6 Conclusion

We have proposed a new of describing the trifocal tensors between three views based on the elegant Grassmann-Cayley formalism and shown that it shed new light on their structure. In particular, we have shown that the tensors could be parameterized minimally and simply with 18 parameters thanks to a set of algebraic constraints on the coefficients of the tensors. We have then proposed a nonlinear method for estimating the tensors from points and lines correspondences that makes full use of their algebraic structure and shown on some real triplets of images that the results obtained using the new method are superior to those obtained by previously published linear methods according to three different and complementary criteria.

References

[AS96] S. Avidan and A. Shashua. Tensorial transfer: Representation of $n > 3$ views of 3d scenes. In *Proceedings of the ARPA Image Understanding Workshop*. darpa, morgan-kaufmann, February 1996.

[AZH96] M. Armstrong, A. Zisserman, and R. Hartley. Self-calibration from image triplets. In Bernard Buxton, editor, *Proceedings of the 4th European Conference on Computer Vision*, Cambridge, UK, April 1996.

[BBHP92] Eamon B. Barrett, Michael H. Brill, Nils N. Haag, and Paul M. Payton. Invariant linear methods in photogrammetry and model-matching. In Joseph L. Mundy and Andrew Zimmerman, editors, *Geometric Invariance in Computer Vision*, chapter 14. MIT Press, 1992.

[BGP93] Eamon Barrett, Gregory Gheen, and Paul Payton. Representation of three-dimensional object structure as cross-ratios of determinants of stereo image points. In Joseph L. Mundy and Andrew Zisserman, editors, *Applications of Invariance in Computer Vision*. ESPRIT, ARPA/NSF, 1993. Ponta Delgada, Azores.

[Car94] Stefan Carlsson. Multiple image invariance using the double algebra. In Joseph L. Mundy, Andrew Zissermann, and David Forsyth, editors, *Applications of Invariance in Computer Vision*, volume 825 of *Lecture Notes in Computer Science*, pages 145–164. Springer-Verlag, 1994.

[DZLF94] R. Deriche, Z. Zhang, Q.-T. Luong, and O. Faugeras. Robust recovery of the epipolar geometry for an uncalibrated stereo rig. In Eklundh [Ekl94], pages 567–576, Vol. 1.

[Ekl94] J-O. Eklundh, editor. volume 800-801 of *Lecture Notes in Computer Science*, Stockholm, Sweden, May 1994. Springer-Verlag.

[FM95a] Olivier Faugeras and Bernard Mourrain. On the geometry and algebra of the point and line correspondences between n images. In *Proceedings of the 5th International Conference on Computer Vision* [icc95], pages 951–956.

[FM95b] Olivier Faugeras and Bernard Mourrain. On the geometry and algebra of the point and line correspondences between n images. Technical Report 2665, INRIA, October 1995.

[Har94a] Richard Hartley. Lines and points in three views-an integrated approach. In *Proceedings of the ARPA Image Understanding Workshop*. Defense Advanced Research Projects Agency, Morgan Kaufmann Publishers, Inc., 1994.

[Har94b] Richard Hartley. Projective reconstruction and invariants from multiple images. *PAMI*, 16(10):1036–1040, 1994.

[Har95] R.I. Hartley. In defence of the 8-point algorithm. In *Proceedings of the 5th International Conference on Computer Vision* [icc95], pages 1064–1070.

[icc95] Boston, MA, June 1995. IEEE Computer Society Press.

[LV94] Q.-T. Luong and T. Viéville. Canonic representations for the geometries of multiple projective views. In Eklundh [Ekl94], pages 589–599.

[SA90a] Minas E. Spetsakis and John Aloimonos. Structure from Motion Using Line Correspondences. *The International Journal of Computer Vision*, 4:171–183, 1990.

[SA90b] Minas E. Spetsakis and Y. Aloimonos. A unified theory of structure from motion. In *Proc. DARPA IU Workshop*, pages 271–283, 1990.

[Sha94a] Amnon Shashua. Projective structure from uncalibrated images: structure from motion and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):778–790, 1994.

[Sha94b] Amnon Shashua. Trilinearity in visual recognition by alignment. In Eklundh [Ekl94], pages 479–484.

[Sha95] Amnon Shashua. Algebraic functions for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):779–789, 1995.

[SW95] A. Shashua and M. Werman. On the trilinear tensor of three perspective views and its underlying geometry. In *Proceedings of the 5th International Conference on Computer Vision* [icc95].

[TZ97] P.H.S. Torr and A. Zissermann. Performance characterization of fundamental matrix estimation under image degradation. *Machine Vision and Applications*, 9:321–333, 1997.

[WHA92] J. Weng, T.S. Huang, and N. Ahuja. Motion and structure from line correspondences: Closed-form solution, uniqueness and optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(3), March 1992.

[ZDFL95] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence Journal*, 78:87–119, October 1995.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399