Tensor Embedding of the Fundamental Matrix

Shai Avidan and Amnon Shashua

Institute of Computer Science, The Hebrew University,
91904 Jerusalem, Israel.
email:{avidan,shashua}@cs.huji.ac.il

Abstract

We revisit the bilinear matching constraint between two perspective views of a 3D scene. Our objective is to represent the constraint in the same manner and form as the trilinear constraint among three views. The motivation is to establish a common terminology that bridges between the fundamental matrix F (associated with the bilinear constraint) and the trifocal tensor $\mathcal{T}_i^{jk}$ (associated with the trilinearities). By achieving this goal we can unify both the properties and the techniques introduced in the past for working with multiple views for geometric applications.

Doing that we introduce a $3 \times 3 \times 3$ tensor $\mathcal{F}_i^{jk}$, we call the bifocal tensor, that represents the bilinear constraint. The bifocal and trifocal tensors share the same form and share the same contraction properties. By close inspection of the contractions of the bifocal tensor into matrices we show that one can represent the family of rank-2 homography matrices by $[\delta]_\times F$ where δ is a free vector. We then discuss four applications of the new representation: (i) Quasi-metric viewing of projective data, (ii) triangulation, (iii) view synthesis, and (iv) recovery of camera ego-motion from a stream of views.

1 Introduction

The geometry of multiple views is governed by certain multi-linear constraints, bilinear for pairs of views and trilinear for triplets of views — all other multi-linear constraints (four views and beyond) are spanned by the bilinear and trilinear constraints.

The traditional representation of the coefficients of the bilinear constraint is by a 3×3 matrix, F, that satisfies $p'^{\top} F p = 0$ for all matching image points p, p' (represented in the 2D projective space) across two views. On the other hand, the three-view relations are represented by a set of 4 trilinear constraints, each of the form $p^i s_j r_k \mathcal{T}_i^{jk} = 0$ where s and r are lines coincident with the matching points p' and p'', respectively. In other words, the bilinear constraint represents a "point+point" relation, whereas each of the trilinear constraints represents a "point+line+line" relation.

Because of the difference in form between the fundamental matrix and the trifocal tensor, the analysis tools are different and the properties discovered for one do not easily carry over to the other. For example, the trifocal tensor contracts (reduces) to matrix forms that carry geometric information: one type of contraction produces subgroups of 2D homography matrices and another type of contraction produces a subgroup of 2D correlation matrices. There is no such equivalence known for the fundamental matrix, for instance.

In this paper we revisit the bilinear constraint and represent it using a $3\times3\times3$ tensor $p^i s_j r_k \mathcal{F}_i^{jk} = 0$ where s, r are two coincident lines with the matching point p'. We call the tensor $\mathcal{F}_i^{jk}$ the "bifocal" tensor and show that not only it shares the same form as the trifocal tensor but it also shares the same properties. We can therefore consider contractions of the bifocal tensor just was is done with the trifocal counterpart.

Through the inspection of tensor contractions we derive the representation of the subgroup of rank-2 homography matrices in the simple form of $[\delta]_\times F$ where δ is a free vector. We introduce the group of "primitive homographies" and discuss 4 applications of the new representation: (i) Quasi-metric viewing of projective data, (ii) triangulation, (iii) view synthesis, and (iv) recovery of camera ego-motion from a stream of views. This work in its initial form was presented at the meeting found in [1].

2 Notations

A point $\boldsymbol{x}$ in the 3D projective space $\mathcal{P}^3$ is projected onto the point p in the 2D projective space $\mathcal{P}^2$ by a 3×4 camera projection matrix $\mathbf{A} = [A, v']$ that satisfies $p \cong \mathbf{A}\boldsymbol{x}$, where $\cong$ represents equality up to scale. The left 3×3 minor of $\boldsymbol{A}$, denoted by A, stands for a 2D projective transformation of some arbitrary plane (the reference plane) and the fourth column of $\boldsymbol{A}$, denoted by v', stands for the epipole (the projection of the center of camera 1 on the image plane of camera 2). In a calibrated setting the 2D projective transformation is the rotational component of camera motion (the reference plane is at infinity) and the epipole is the translational component of camera motion. Since only relative camera positioning can be recovered from image measurements, the camera matrix of the first camera position in a sequence of positions can be represented by $[I; 0]$.

We will occasionally use tensorial notations, which are briefly described next. We use the covariant-contravariant summation convention: a point is an object whose coordinates are specified with superscripts, i.e., $p^i = (p^1, p^2, ...)$. These are called contravariant vectors. An element in the dual space (representing hyper-planes — e.g., lines in $\mathcal{P}^2$), is called a covariant vector and is represented by subscripts, i.e., $s_j = (s_1, s_2, ...)$. Indices repeated in covariant and contravariant forms are summed over, i.e., $p^i s_i = p^1 s_1 + p^2 s_2 + ... + p^n s_n$. This is known as a contraction. For example, if p is a point incident to a line s in $\mathcal{P}^2$, then $p^i s_i = 0$. Vectors are also called 1-valence tensors. 2-valence tensors (matrices) have two indices and the transformation they represent depends on the covariant-contravariant positioning of the indices. For example, a_i^j is a mapping from points to points, and hyper-planes to hyper-planes, because $a_i^j p^i = q^j$ and $a_i^j s_j = r_i$ (in matrix form: $Ap = q$ and $A^\top s = r$); a_{ij} maps points to hyper-planes; and a^{ij} maps hyper-planes to points. When viewed as a matrix the row and column positions are determined accordingly: in a_i^j and a_{ji} the index i runs over the columns and j runs over the rows, thus $b_j^k a_i^j = c_i^k$ is $BA = C$ in matrix form. An outer-product of two 1-valence tensors (vectors), $a_i b^j$, is a 2-valence tensor c_i^j whose i, j entries are $a_i b^j$ — note that in matrix form $C = ba^\top$. An n-valence tensor described as an outer-product of n vectors is a rank-1 tensor. Any n-valence tensor can be described as a sum of rank-1 n-valence tensors. The rank of an n-valence tensor is the *smallest* number of rank-1 n-valence tensors with sum equal to the tensor. For example, a rank-1 trivalent tensor is $a_i b_j c_k$ where a_i,b_j and c_k are three vectors. The rank of a trivalent tensor α_{ijk} is the smallest r such that,

$$\alpha_{ijk} = \sum_{s=1}^{r} a_{is} b_{js} c_{ks}. \tag{1}$$

We will make extensive use of the "cross-product tensor" ϵ defined next. The cross product (vector product) operation $c = a \times \mathrm{b}$ is defined for vectors in $\mathcal{P}^2$. The vector c is the line joining the points a, b, or the point of intersection of the lines a, b. The product operation can also be represented as the product $c = [a]_\times b$ where $[a]_x$ is called the "skew-symmetric matrix of a" and has the form:

$$[a]_\times = \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}$$

In tensor form we have $\epsilon_{ijk} a^i b^j = c_k$ representing the cross products of two points (contravariant vectors) resulting in the line (covariant vector) c_k. Similarly, $\epsilon^{ijk} a_i b_j = c^k$ represents the point intersection of the to lines a_i and b_j. The tensor ϵ is defined such that $\epsilon_{ijk} a^i$ produces the matrix $[a]_\times$ (i.e., ϵ contains $0, -1, 1$ in its entries such that its operation on a single vector produces the skew-symmetric matrix of that vector).

3 Tensor Embedding of the Fundamental Matrix

Our goal is to derive a trivalent tensor representation (i.e., a $3 \times 3 \times 3$ tensor) of the 3×3 fundamental matrix and to illuminate the advantages of doing so. In particular, once we have the trivalent tensor representation in our hand we wish to investigate its contraction properties (as was done for the trifocal tensor in [22]) and recast them back in matrix form.

We start with deriving the fundamental matrix from basic principles. Let A be a 2D homography (collineation) from image 1 to image 2 due to some plane π, i.e., if p is a point in image 1, then Ap is a point coincident with the epipolar line $p' \times v'$ in image 2, where the exact location of Ap on the epipolar line is determined by the position of the plane π. Thus, $(v' \times p')^\top Ap = 0$, or in tensor notation,

$$\begin{aligned} 0 &= \epsilon_{lj\rho} p'^j v'^\rho p^i a_i^l \\ &= p'^j \underbrace{\left(\epsilon_{lj\rho} v'^\rho a_i^l\right)}_{F_{ji}} p^i \end{aligned}$$

where $\epsilon_{lj\rho}p'^j v'^\rho$ is the cross-product $p' \times v'$. The matrix $F_{ji} = \epsilon_{lj\rho}v'^\rho a_i^l$ is the fundamental matrix that satisfies the bilinear constraint $p^i p'^j F_{ji} = 0$ (cf. [14, 6]). In matrix form, since $\epsilon_{lj\rho}v'^\rho$ is the skew-symmetric matrix $[v']_\times$, then $F = [v']_\times A$.

Next, we begin with the bilinear constraint $p^i p'^j F_{ji} = 0$ and consider replacing the point p' with a cross product of any two incident lines s, r, i.e., $p'^l = \epsilon^{ljk}s_j r_k$. The reason for doing so will be apparent later on. We have therefore a "point+line+line" relationship $p^i s_j r_k \mathcal{F}_i^{jk} = 0$ as follows:

$$\begin{aligned} p^i p'^l F_{li} &= p^i \underbrace{(\epsilon^{ljk}s_j r_k)}_{p'^l} F_{li} \\ &= p^i s_j r_k \underbrace{(\epsilon^{ljk}F_{li})}_{\mathcal{F}_i^{jk}} \\ &= 0 \end{aligned}$$

and the tensor $\mathcal{F}_i^{jk} = \epsilon^{ljk}F_{li}$ is a trivalent form of the fundamental matrix. This form is equivalent to considering the trifocal tensor of views 1,2,3 where views 2,3 are identical. Thus we obtain a relationship between three views, but only two of the views are distinct. We can represent the "bifocal" tensor $\mathcal{F}_i^{jk}$ directly as a function of v' and A as follows:

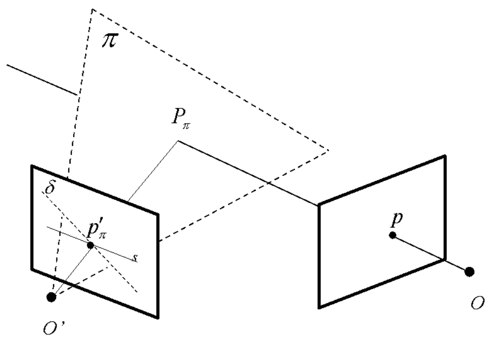
$$\mathcal{F}_i^{jk} = v'^j a_i^k - v'^k a_i^j.$$

The importance of the trivalent tensor embedding of the fundamental matrix (which we will denote by bifocal tensor from now on) is that we have arrived to an equivalent representation with 3-view geometry: both the trifocal and bifocal tensors are $3 \times 3 \times 3$ and operate on a configuration of a point+line+line. In the case of three views, the lines are in two distinct views (the line s coincides with p' and the line r coincides with p'') and there are 4 such relationships (due to the fact that there are two choices for each line). In the case of two views the two lines are in the same view and therefore there is only one configuration of point+line+line.

The advantage of this equivalence in form between the trifocal and bifocal tensors appears when one considers contractions into bivalent forms (matrices). The properties of contractions of the trifocal tensor are well understood (see [22, 18] and in the appendix here) and provide the building blocks for making use of the trifocal tensor in applications. We can apply now an identical analysis on the bifocal tensor which we will do next.

3.1 Bifocal Tensor Contractions

Given an arbitrary vector δ, the trifocal tensor reduces to a matrix of three types: $\delta^i \mathcal{T}_i^{jk}, \delta_j \mathcal{T}_i^{jk}$ and $\delta_k \mathcal{T}_i^{jk}$. Note that when $\delta = (1,0,0), (0,1,0)$ or $(0,0,1)$ we obtain "slices" of the tensor. The first type produces a rank-2 correlation matrix, i.e., a mapping from all 2D lines to collinear points (where the orientation of collinearity is determined by δ) — by slicing the tensor in that way we



(a)

Fig. 1. The matrix $[\delta]_\times F$ is a homography matrix due to a plane π coincident with the center of projection O' and the line δ in view 2. The line s is the epipolar line and the point P_π is at the intersection of the optic ray from the first view and the plane π. The point p'_π is the projection of P_π onto view 2. Therefore the point+line+line configuration of p, δ, s satisfies the bifocal tensor relation $p^i s_j \delta_k \mathcal{F}_i^{jk} = 0$, where $\delta_k \mathcal{F}_i^{jk}$ is the matrix $[\delta]_\times F$.

obtain the three matrices of "line geometry" introduced in the calibrated context by [23, 24, 28]. The second and third types produce homography matrices (collineations). The second type is a homography matrix from view 1 to 3 due to a plane determined by the line δ_j in view 2 and the center of projection of camera 2. Likewise, the third type is a homography from view 1 to 2 via a plane determined by the line δ_k in view 3 and the center of projection of camera 3. These homography matrices were introduced in [22] and are described in more detail in the appendix here.

We wish to consider the same types of contractions on the bifocal tensor $\mathcal{F}_i^{jk}$ — by equivalence of form, we should obtain collineations and correlations as well. Consider the contraction

$$\delta_k \mathcal{F}_i^{jk}$$

for some arbitrary vector δ. By substitution in the definition of $\mathcal{F}_i^{jk}$ we obtain

$$\delta_k \mathcal{F}_i^{jk} = \underbrace{(\epsilon^{ljk}\delta_k)}_{[\delta]_\times} F_{li}$$

which in matrix form becomes $[\delta]_\times F$. Our question therefore is about the geometric interpretation of this matrix (for an arbitrary δ). Given the form-equivalence of the two tensors the answer is immediate: $[\delta]_\times F$ is a homography matrix from view 1 to view 2 via a plane coincident with the center of projection O' of camera 2 and the line δ in view 2. The family of such matrices over all choices of δ corresponds to the family of homography matrices whose planes are coincident with O'. The family is spanned by three matrices (since δ is spanned by three vectors), and for example, the three slices using $\delta = (1,0,0),(0,1,0)$ or $(0,0,1)$ will provide the basis for this subgroup of homography matrices.

More formally, consider the plane π defined by the point O' and the line δ in view 2. Consider a point p in view 1 and the ray from the center of projection O of the first camera and the point p. The ray intersects π at P_π which projects to a point p'_π in view 2 which is coincident with the line δ (by construction). Let s_j be the epipolar line of p in view 2, thus $p^i s_j \delta_k \mathcal{F}_i^{jk} = 0$ because they provide a point+line+line configuration, and this holds for all points p (see Fig. 1). Thus, the matrix $\delta_k \mathcal{F}_i^{jk}$ maps view 1 onto points along the corresponding epipolar lines and is therefore a homography matrix, and since the projected points are collinear the rank of the matrix is 2. We have the following result:

Theorem 1. *The matrix $[\delta]_\times F$ is a homography matrix of rank 2 from view 1 to view 2 due to the plane coincident with the center of projection of camera 2 and the line δ in view 2.*

Note that the theorem generalizes the observation due to [16] that $[v']_x F$ is a homography matrix. We see that this is true for any choice of skew-symmetric matrix $[\delta]_\times$.

The same result applies for the contraction $\delta_j \mathcal{F}_i^{jk}$ (with a change of sign). Note that with the trifocal tensor there is a difference between the contractions $\delta_j \mathcal{T}_i^{jk}$ and $\delta_k \mathcal{T}_i^{jk}$ in which the former produces a homography matrix from view 1 to 3 and the latter produces a homography matrix from view 1 to 2. In the case of the bifocal tensor views 2 and 3 coincide thus the two types of contractions are equivalent.

The remaining contraction type is $\delta^i \mathcal{F}_i^{jk}$. In the case of the trifocal tensor the contraction $\delta^i \mathcal{T}_i^{jk}$ produces a *correlation* matrix which maps the space of lines from view 2 to a set of collinear points (on the epipolar line of the point δ in view 1) in view 3. The transpose of that matrix is the same type of mapping, but from view 3 to view 2 (see Appendix). We should obtain something similar for the bifocal tensor and since view 2 and 3 coincide the matrix $\delta^i \mathcal{F}_i^{jk}$ should map the space of lines in view 2 onto collinear points in view 2 that define the epipolar line, $F\delta$, of the point δ. Indeed, by substitution we obtain:

$$\begin{aligned} \delta^i \mathcal{F}_i^{jk} &= \epsilon^{ljk} \delta^i F_{li} \\ &= [F\delta]_\times \end{aligned} \tag{2}$$

Thus, $[F\delta]_\times s$ for all lines s in view 2 is the point of intersection of the epipolar line $F\delta$ and the line s. In other words, the matrix $[F\delta]_\times$ is the correlation matrix we described above. Note that the reason we have obtained a trivial mapping is due to fact that this type of contraction is associated with reconstruction for lines. The three matrices $\delta^i \mathcal{T}_i^{jk}$ for $\delta = (1,0,0),(0,1,0)$ and $(0,0,1)$ are known to arise from considerations of matching lines across three views (cf. [23, 28, 10]). However, the relative camera positions cannot be recovered from matching lines across two views only (only from matching points), which is why the corresponding correlation matrices $\delta^i \mathcal{F}_i^{jk}$ of the bifocal tensor become trivial.

To summarize, the embedding of the fundamental matrix in trivalent tensor format (the bifocal tensor) provides a unified terminology of a "point+line+line" that applies for both the bifocal and trifocal relationships across multiple views. In particular, as is the case with the trifocal tensor, contractions of the bifocal tensor into reduced forms (matrices) have a geometric significance. The contractions properties of the bifocal tensor are listed in Table 1. We see a clear analogy to the type of resulting matrices (homography and correlations) one obtains from the same contractions applied to the trifocal tensor. Furthermore, the homography contraction provides the basis for all rank-2 homography matrices whose planes are coincident with the center of projection of camera 2. All linear combinations of the rank-2 homography matrices are of the form $[\delta]_\times F$ for some vector δ.

Table 1. *The three types of contractions of the bifocal tensor (embedding of the fundamental matrix F as a trivalent tensor $\mathcal{F}_i^{jk}$), their matrix form, and the property they produce. Note that the first two contractions produce a homography matrix of a plane whose orientation is determined by the vector of contraction δ.*

Contraction	Matrix Form	Result
$\delta_k \mathcal{F}_i^{jk}$	$[\delta]_\times F$	Homography Matrix.
$\delta_j \mathcal{F}_i^{jk}$	$[\delta]_\times F$	Same as above.
$\delta^i \mathcal{F}_i^{jk}$	$[F\delta]_\times$	Trivial Correlation Mapping.

4 The Primitive Homography Matrices

We have seen that the family of matrices $[\delta]_\times F$ parameterized by the choice of the vector δ spans the family of homography matrices from view 1 to view 2 due to the planes coincident with the center of projection O' of camera 2. The vector δ determines the orientation of the plane and is the line of intersection of the plane and view 2. Since δ is spanned by three vectors, say $(1, 0, 0), (0, 1, 0)$ and $(0, 0, 1)$, the bifocal tensor contractions provide three distinct homography matrices that span the subgroup of homography matrices (those whose planes are coincident with O'). Since the entire group of all homography matrices lies in a 4 dimensional subspace [20], i.e., spanned by 4 homography matrices whose planes do not all coincide with a single point, we must produce an additional homography matrix in order to complete the basis of the subgroup defined by $[\delta]_\times F$ to a full basis for the entire group. The elements (matrices) of the full basis will be called "primitive homographies". The additional homography matrix we seek must therefore be associated with a plane coincident with the center of projection O of camera 1 (and is therefore of rank 1). We have the following Lemma which is adapted from [17]:

Lemma 1. *Given the fundamental matrix F and the epipole v' defined by $F^\top v' = 0$, then the family of matrices $v'\delta^\top$ are homography matrices from view 1 to view 2 due to planes coincident with the center of projection O of camera 1 and the vector δ is the intersection line of the plane and view 1.*

Proof: Let A_1, A_2 be any two homography matrices. Thus, A_1p, A_2p and v' are collinear for all points p in view 1. Let $q \neq v$, where v is the epipole in view 1 ($Fv = 0$), be some point in view 1 and let λ be a scalar defined such that $A_1q - \lambda A_2 q \cong v'$. Let $H = A_1 - \lambda A_2$ be a homography matrix (because all homography matrices are closed under linear combinations). Clearly, since $Hq \cong v'$, then $Hp \cong v'$ for all p (because $Hv \cong v'$ as well). Thus $H = v'\delta^\top$ for some vector δ. $\square$

Therefore, as long as $\delta^\top v \neq 0$, where v is the epipole in view 1 (i.e., $Fv = 0$), then the homography matrix $v'\delta^\top$ does not coincide with O' (only with O) and thus can be used to complete the full basis for the group of homography matrices. Without loss of generality assume that $(1,0,0)$ is not coincident with v, thus we have a basis of 4 homography matrices $H_1, ..., H_4$, denoted as "primitive homographies", defined below:

$$H_i = [e_i]_\times F, \qquad i = 1,2,3 \tag{3}$$
$$H_4 = v' e_1^\top \tag{4}$$

where e_i are the identity vectors: $e_1 = (1,0,0), e_2 = (0,1,0)$ and $e_3 = (0,0,1)$.

5 Applications Using Primitive Homography Matrices

The primitive homography matrices are a useful tool for representing geometric data. We will consider two examples here, the first on obtaining a "quasi-metric" representation of 3D space from a pair of uncalibrated cameras, and the second on "triangulation" from 3 views.

5.1 Quasi-Metric Reference Plane

Let p_i, p'_i, $i = 1, ..., N$, be matching points in view 1 and 2 respectively. Given the fundamental matrix F and the epipole v' in view 2, then the 3D projective representation of the object space points P_i can be described relative to a reference plane π:

$$p'_i \cong A_\pi p_i + \rho_i v' = [A_\pi, v'] P_i$$

where A_π is the homography matrix mapping view 1 onto view 2 due to the plane π. The scalar ρ_i represents the relative deviation of the point P_i from the plane π and is called the "relative affine structure"[21]. The choice of the plane π determines the projective representation of object space. For purposes of visualization, it is useful to choose π such that it is situated "in-between" the space points making it possible to treat ρ_i as simple depth variable. In other words, let $A_\pi = \sum_j \alpha_j H_j$, we seek to solve for the scalar α_j, $j = 1, ..., 4$, that minimize:

$$\sum_{j=1}^{4} (\alpha_j H_j) p_i \cong p'_i \quad i = 1, ..., N$$

which provides an over-determined linear set of equations. We will refer to π as the "quasi-metric" plane. The choice of the quasi-metric plane provides a better chance that the projective viewing of the object (treating the coordinates x_i, y_i, ρ_i as Euclidean coordinates by the viewing program) will have less projective distortions than other choices.

5.2 Triangulation from 3 Views

Hartley and Sturm [11] considered the problem, they called "triangulation", of modifying the locations of input matching points $\hat{p}, \hat{p}'$ that are given with noise to new locations p, p' that satisfy $p'^\top F p = 0$ such that $(p - \hat{p})^2 + (p' - \hat{p}')^2$ is minimized. The triangulation problem in 3 views can be stated in a similar manner: given p in view 1, the matching process produces an error in the matches in view 2 and 3. The input matches are $\hat{p}'$ and $\hat{p}''$ and we wish to find new matches p', p'' with p such that the triplet p, p', p'' satisfy the trilinear equations while $(p' - \hat{p}')^2 + (p'' - \hat{p}'')^2$ is minimized. Note that we do not add an error term to p and rather take p as a reference. The reason for that is twofold: first due to the asymmetry of the trifocal tensor with respect to view ordering as it is defined with respect to a reference view (unlike the fundamental matrix which remains fixed under view ordering). Secondly, in most matching approaches that use a correlation principle, like the popular Lucas-Kanade [15] method with the coarse-to-fine implementation by Sarnoff Corp. [4], there is also an intrinsic asymmetry that assumes one of the views as a reference. Taken together, we can without loss of generality assume that the effect of error in the matching process is represented in the displacement of $\hat{p}'$ and $\hat{p}''$ from their true locations p', p''.

The triangulation process using the trifocal tensor can proceed as follows. We first note that the following relationship exists:

$$p' \cong Ap + \rho v' \tag{5}$$

$$p'' \cong Bp + \rho v'' \tag{6}$$

where A, B are two homography matrices from view 1 to 2 and from view 1 to 3 via some reference plane π (any plane). Given the trifocal tensor $\mathcal{T}_i^{jk}$ one can recover the epipoles v', v'' (and fundamental matrices) [10, 22] and proceed to recover a pair of homography matrices A, B as described below.

One can solve for A be either choosing some linear combination of the primitive homographies or solving for the quasi-metric plane as described in the previous section. Thus we can assume that A is known. The corresponding homography B cannot be chosen arbitrarily because it must be associated with the same plane π that was associated with the homography A.

Let $\bar{H}_l$, $l = 1, ..., 4$, be the primitive homographies from view 1 to 3. Let the sought after matrix B be represented by $B = \sum_l \beta_l \bar{H}_l$. We seek a solution of the scalars β_l. We have the following relationship:

$$\begin{aligned} \mathcal{T}_i^{jk} &= v'^j b_i^k - v''^k a_i^j \\ &= v'^j (\Sigma_{l=1}^4 \beta_l \bar{H}_l)_i^k - \lambda v''^k a_i^j \end{aligned} \tag{7}$$

where the left-hand side is known (the trifocal tensor) and the right-hand side contains 5 unknowns which together form an over-determined linear system. The scalar λ fixes the scale because v', v'', A are all determined up to scale. Taken together, from the trifocal tensor and with the use of the primitive homographies we can extract a set of compatible homographies (associated with the same plane) and the epipoles v', v''.

We are now left with minimizing the following expression:

$$\min_{\rho}\{(\hat{x}' - \frac{a_1^\top p + \rho v_1'}{a_3^\top p + \rho v_3'})^2 + (\hat{y}' - \frac{a_2^\top p + \rho v_2'}{a_3^\top p + \rho v_3'})^2 + (\hat{x}'' - \frac{b_1^\top p + \rho v_1''}{b_3^\top p + \rho v_3''})^2 + (\hat{y}'' - \frac{b_2^\top p + \rho v_2''}{b_3^\top p + \rho v_3''})^2\}$$

which is minimized with respect to ρ. This yields a 4th order polynomial in ρ which thus has a closed-form solution. The geometric interpretation of this minimization process is that the solution ρ determines the points p', p'' on their corresponding epipolar lines such that the distance $(p' - \hat{p}')^2 + (p'' - \hat{p}'')^2$ is minimized. Note that unlike the case of two views, one cannot place p' and p'' anywhere on their epipolar lines because they are coupled together by a 1-parameter degree of freedom. In particular, the projections of $\hat{p}'$ and $\hat{p}''$ on their epipolar lines may not be an admissible solution.

6 Other Applications of the Bifocal Tensor Representation

In the previous sections we presented applications of the primitive homographies which in turn are due to the discovery of $[\delta]_\times F$ representing the family of rank-2 homographies which in turn are due to the bifocal tensor representation. However, one could possibly re-derive the result $[\delta]_\times F$ from purely matrix considerations without relying on the bifocal tensor. Nevertheless, there are applications that critically rely on the tensor embedding of the fundamental matrix in the form of the bifocal tensor — and in this section we briefly discuss two of them.

6.1 View-Synthesis

The notion of image-based rendering is gaining momentum both in the computer graphics and computer vision communities. Using the trifocal tensor for image-based rendering was proposed by [2]. In a nutshell, the method links together two real views of a 3D scene with a third virtual view of the scene. The tensor is

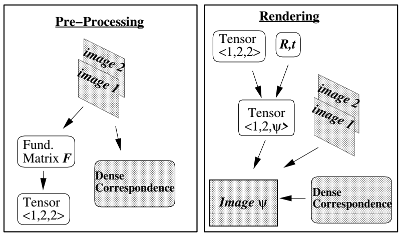


Fig. 2. View synthesis is divided into two parts. In the pre-processing stage, done only once, we compute the dense correspondence and the bifocal tensor. The rendering stage, done for every novel image, transforms the "seed" bifocal tensor to a general three-view trifocal tensor, using user-specified parameters R, t and renders the novel view using the transformed tensor, the model images and the dense correspondence.

then used to reproject a point appearing in the first two views directly onto the virtual view, without ever recovering 3D structure. Moving the virtual camera in space is done by modifying the tensor to reflect the change in the relative position of the virtual view. To bootstrap the seed tensor one would need three real views of the object, but only two of them will be later used for the generation of the virtual view. However, using the tensor-embedded fundamental matrix, one can use only two real images to generate the bifocal "seed tensor".

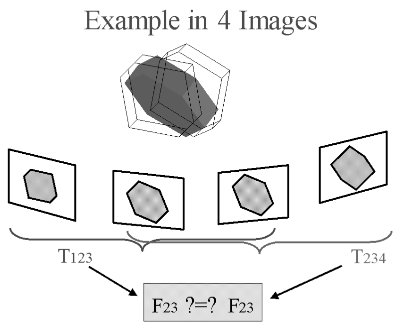
One starts with the bifocal tensor which is then transformed using the user specified motion of the virtual camera to the appropriate trifocal tensor (of the original two model views and the virtual view to be synthesized). From there on the trifocal tensors transform as the virtual camera changes positions (see Fig. 2). Thus, for this application to work it is necessary to have a uniform terminology for handling 2 and 3 views.

6.2 Ego-motion Recovery

When considering the problem of recovering the camera ego-motion (projection matrices) from a stream of views, one faces the problem of maintaining a consistency of pairwise fundamental matrices. The consistency requirement arises from the simple fact that from an algebraic standpoint a camera trajectory must be *concatenated* from pairs or triplets of images. Therefore, a sequence of independently computed fundamental matrices or trifocal tensors, maybe optimally consistent with the image data, but not necessarily consistent with a unique camera trajectory (see Figure 3).

The consistency problem can be approached by introducing the following equation which relates the trifocal tensor between views 1,2,3 and the bifocal tensor between views 1,2 and the elements of the fundamental matrix between views 2,3:

$$\mathcal{T}_i^{jk} = c_l^k \mathcal{F}_i^{jl} - v'''^k a_i^j \tag{8}$$



(a)

Fig. 3. One can compute two tensors T_{123}, T_{234} from the four images of the 3D scene. However, each tensor can give rise to a different reconstruction of the 3D structure due to noise or errors in measurements, and therefor the camera trajectory between images 2 and 3, as captured by the fundamental matrix F_{23}, is inconsistent between the two tensors. Figure taken from [3]

where $\mathcal{T}_i^{jk}$ is the tensor of views 1,2,3, the matrix A, whose elements are a_i^j, is a homography from views 1 to 2 via some arbitrary plane π, $\mathcal{F}_i^{jl}$ is the bifocal tensor of views 1,2, and $\mathbf{C} = [C; v''']$ is the camera motion from view 2 to 3 where c_l^k is a homography matrix from view 2 to 3 via the (same) plane π.

As a result, given the fundamental matrix between views 1,2 and (at least) 6 matching points between views 1,2,3 one can solve for the fundamental matrix between views 2 and 3 (i.e., $[v''']_\times C$) which is consistent with the trifocal relationship among views 1,2,3. Also, as a byproduct, the projection matrix $[C, v''']$ is consistent with the same projective representation due to the fact that the homographies A, C are of the same reference plane. The details and demonstration of this idea can be found in [3].

7 Summary

We have introduced a new representation of the bilinear matching constraint between a pair of views in terms of a $3 \times 3 \times 3$ tensor which we termed the "bifocal" tensor. The motivation for the new representation is to establish a unified terminology between the elements of 2-view and 3-view constraints. The unified terminology is achieved by representing the 2-view constraint in a way analogously (and identical in form) to the trifocal tensor relationship. As a result, we were able to transfer the properties known today about the trifocal tensor (especially the contraction into homography matrices) to the realm of the 2-view case.

The byproduct of the new representation is twofold. First, we have derived the family of rank-2 homography matrices represented by $[\delta]_\times F$ and introduced the "primitive homographies" and their applications. Second, we mentioned two other applications for which the unified terminology is necessary.

Taken together, it is useful to have a common language for analyzing the geometric constraints arising from multiple-view geometry — both at the theoretical level for purposes of obtaining a clean representation and for applications where the common language is sometimes necessary (as was shown in Section 6).

Acknowledgments

We thank Richard Hartley for comments on the previous version of this work (in [1]) which led to a simpler representation of $\mathcal{F}_i^{jk}$. A.S. also thanks Nir Avrahami for noticing the need for the scale factor λ in eqn. 7 and Jacob Goldberg for suggesting the use of the relative-affine invariant for "triangulation" in Section 5.2.

References

1. S. Avidan and A. Shashua. Unifying two-view and three-view geometry. In *ARPA, Image Understanding Workshop*, 1997.
2. S. Avidan and A. Shashua. Novel view synthesis by cascading trilinear tensors. *IEEE Transactions on Visualization and Computer Graphics*, 4(3), 1998. Short version can be found in CVPR'97.
3. S. Avidan and A. Shashua. Threading fundamental matrices. In *Proceedings of the European Conference on Computer Vision*, Frieburg, Germany, June 1998. Springer, LNCS 1406.
4. J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proceedings of the European Conference on Computer Vision*, Santa Margherita Ligure, Italy, June 1992.
5. A. Criminisi, I. Reid, and A. Zisserman. Duality, rigidity and planar parallax. In *Proceedings of the European Conference on Computer Vision*, Frieburg, Germany, 1998. Springer, LNCS 1407.
6. O.D. Faugeras. Stratification of three-dimensional vision: projective, affine and metric representations. *Journal of the Optical Society of America*, 12(3):465–484, 1995.
7. O.D. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondences between N images. In *Proceedings of the International Conference on Computer Vision*, Cambridge, MA, June 1995.
8. O.D. Faugeras and T. Papadopoulo. A nonlinear method for estimating the projective geometry of three views. In *Proceedings of the International Conference on Computer Vision*, Bombay, India, January 1998.
9. R. Hartley. Lines and points in three views — a unified approach. In *Proceedings of the DARPA Image Understanding Workshop*, Monterey, CA, November 1994.
10. R.I. Hartley. Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision*, 22(2):125–140, 1997.
11. R.I. Hartley and P. Sturm. Triangulation. In *Proceedings of the DARPA Image Understanding Workshop*, pages 972–966, Monterey, CA, Nov. 1994.
12. A. Heyden. Reconstruction from image sequences by means of relative depths. In *Proceedings of the International Conference on Computer Vision*, pages 1058–1063, Cambridge, MA, June 1995.

13. M. Irani, P. Anandan, and D. Weinshall. From reference frames to reference planes: Multiview parallax geometry and applications. In *Proceedings of the European Conference on Computer Vision*, Frieburg, Germany, 1998. Springer, LNCS 1407.
14. H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.
15. B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings IJCAI*, pages 674–679, Vancouver, Canada, 1981.
16. Q.T. Luong and T. Vieville. Canonic representations for the geometries of multiple projective views. In *Proceedings of the European Conference on Computer Vision*, pages 589–599, Stockholm, Sweden, May 1994. Springer Verlag, LNCS 800.
17. A. Shashua. Algebraic functions for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):779–789, 1995.
18. A. Shashua. Trilinear tensor: The fundamental construct of multiple-view geometry and its applications. In G. Sommer and J.J. Koenderink, editors, *Algebraic Frames For The Perception Action Cycle*, number 1315 in Lecture Notes in Computer Science. Springer, 1997. Proceedings of the workshop held in Kiel, Germany, Sep. 1997.
19. A. Shashua and P. Anandan. The generalized trilinear constraints and the uncertainty tensor. In *Proceedings of the DARPA Image Understanding Workshop*, Palm Springs, CA, February 1996.
20. A. Shashua and S. Avidan. The rank4 constraint in multiple view geometry. In *Proceedings of the European Conference on Computer Vision*, Cambridge, UK, April 1996.
21. A. Shashua and N. Navab. Relative affine structure: Canonical model for 3D from 2D geometry and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):873–883, 1996.
22. A. Shashua and M. Werman. Trilinearity of three perspective views and its associated tensor. In *Proceedings of the International Conference on Computer Vision*, June 1995.
23. M.E. Spetsakis and J. Aloimonos. Structure from motion using line correspondences. *International Journal of Computer Vision*, 4(3):171–183, 1990.
24. M.E. Spetsakis and J. Aloimonos. A unified theory of structure from motion. In *Proceedings of the DARPA Image Understanding Workshop*, 1990.
25. G. Stein and A. Shashua. Model based brightness constraints: On direct estimation of structure and motion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Puerto Rico, June 1997.
26. G. Stein and A. Shashua. On degeneracy of linear reconstruction from three views: Linear line complex and applications. In *Proceedings of the European Conference on Computer Vision*, Frieburg, Germany, 1998. Springer, LNCS 1407.
27. B. Triggs. Matching constraints and the joint image. In *Proceedings of the International Conference on Computer Vision*, pages 338–343, Cambridge, MA, June 1995.
28. 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), 1992.

A Appendix

A.1 Trilinearities and the Trifocal Tensor

Three views, $p = [I;0]\boldsymbol{x}, p' \cong \boldsymbol{A}\boldsymbol{x}$ and $p'' \cong \boldsymbol{B}\boldsymbol{x}$, are known to produce four trilinear forms whose coefficients are arranged in a tensor representing a bilinear function of the camera matrices $\boldsymbol{A}, \boldsymbol{B}$:

$$\mathcal{T}_i^{jk} = v'^j b_i^k - v''^k a_i^j \tag{9}$$

where $A = [a_i^j, v'^j]$ (a_i^j is the 3×3 left minor and v' is the fourth column of A) and $B = [b_i^k, v''^k]$. The tensor acts on a triplet of matching points in the following way:

$$p^i s_j^\mu r_k^\rho \mathcal{T}_i^{jk} = 0 \tag{10}$$

where s_j^μ are any two lines (s_j^1 and s_j^2) intersecting at p', and r_k^ρ are any two lines intersecting p''. Since the free indices are μ, ρ each in the range 1,2, we have 4 trilinear equations (unique up to linear combinations). If we choose the *standard* form where s^μ (and r^ρ) represent vertical and horizontal scan lines, i.e.,

$$s_j^\mu = \begin{bmatrix} -1 & 0 & x' \\ 0 & -1 & y' \end{bmatrix}$$

then the four trilinear forms, referred to as *trilinearities* [17], have the following explicit form:

$$\begin{aligned}
x''\mathcal{T}_i^{13}p^i - x''x'\mathcal{T}_i^{33}p^i + x'\mathcal{T}_i^{31}p^i - \mathcal{T}_i^{11}p^i &= 0, \\
y''\mathcal{T}_i^{13}p^i - y''x'\mathcal{T}_i^{33}p^i + x'\mathcal{T}_i^{32}p^i - \mathcal{T}_i^{12}p^i &= 0, \\
x''\mathcal{T}_i^{23}p^i - x''y'\mathcal{T}_i^{33}p^i + y'\mathcal{T}_i^{31}p^i - \mathcal{T}_i^{21}p^i &= 0, \\
y''\mathcal{T}_i^{23}p^i - y''y'\mathcal{T}_i^{33}p^i + y'\mathcal{T}_i^{32}p^i - \mathcal{T}_i^{22}p^i &= 0.
\end{aligned}$$

These constraints were first derived in [17]; the tensorial derivation leading to eqns. 9 and 10 was first derived in [19]. The tensor is often referred to as "trilinear" or "trifocal", and we adopt here the term trifocal tensor. The trifocal tensor has been well known in disguise in the context of Euclidean line correspondences and was not identified at the time as a tensor but as a collection of three matrices (a particular contraction of the tensor, correlation contractions, as explained next) [23, 24, 28]. The link between the trilinearities and the matrices of line geometry was identified later by Hartley [9, 10]. Additional work in this area can be found in [22, 7, 27, 12, 20, 3, 2, 25, 8, 13, 5, 26].

The tensor has certain contraction properties and can be sliced in three principled ways into matrices with distinct geometric properties. These properties is what makes the tensor distinct from simply being a collection of three matrices and will be briefly discussed next — further details can be found in [22, 18].

A.2 Contraction Properties and Tensor Slices

Consider the matrix arising from the contraction,

$$\delta_k \mathcal{T}_i^{jk} \tag{11}$$

which is a 3×3 matrix, we denote by E, obtained by the linear combination $E = \delta_1 \mathcal{T}_i^{j1} + \delta_2 \mathcal{T}_i^{j2} + \delta_3 \mathcal{T}_i^{j3}$ (which is what is meant by a contraction), and δ_k is an *arbitrary* covariant vector. The matrix E has a general meaning introduced in [22]:

Proposition 1 (Homography Contractions). *The contraction $\delta_k \mathcal{T}_i^{jk}$ for some arbitrary δ_k is a homography matrix from image one onto image two determined by the plane containing the third camera center C'' and the line δ_k in the third image plane. Generally, the rank of E is 3. Likewise, the contraction $\delta_j \mathcal{T}_i^{jk}$ is a homography matrix from image one onto image three.*

For proof see [22]. Clearly, since δ is spanned by three vectors, we can generate up to at most three distinct homography matrices by contractions of the tensor. We define the *Standard Homography Slicing* as the homography contractions associated by selecting δ be $(1,0,0)$ or $(0,1,0)$ or $(0,0,1)$, thus the three standard homography slices between image one and two are $\mathcal{T}_i^{j1}, \mathcal{T}_i^{j2}$ and $\mathcal{T}_i^{j3}$, and we denote them by E_1, E_2, E_3 respectively, and likewise the three standard homography slices between image one and three are $\mathcal{T}_i^{1k}, \mathcal{T}_i^{2k}$ and $\mathcal{T}_i^{3k}$, and we denote them by W_1, W_2, W_3 respectively.

Similarly, consider the contraction

$$\delta^i \mathcal{T}_i^{jk} \tag{12}$$

which is a 3×3 matrix, we denote by T, and where δ^i is an *arbitrary* contravariant vector. The matrix T has a general meaning is well, as detailed below [18]:

Proposition 2. *The contraction $\delta^i \mathcal{T}_i^{jk}$ for some arbitrary δ^i is a rank 2 correlation matrix from image two onto image three, that maps the dual image plane (the space of lines in image two) onto a set of collinear points in image three that form the epipolar line corresponding to the point δ^i in image one. The null space of the correlation matrix is the epipolar line of δ^i in image two. Similarly, the transpose of T is a correlation from image three onto image two with the null space being the epipolar line in image three corresponding to the point δ^i in image one.*

For proof see [18]. We define the *Standard Correlation Slicing* as the correlation contractions associated with selecting δ be $(1,0,0)$ or $(0,1,0)$ or $(0,0,1)$, thus the three standard correlation slices are $\mathcal{T}_1^{jk}, \mathcal{T}_2^{jk}$ and $\mathcal{T}_3^{jk}$, and we denote them by T_1, T_2, T_3, respectively. The three standard correlations date back to the work on structure from motion of lines across three views [23, 28] where these matrices were first introduced.