On the Structure and Properties of the Quadrifocal Tensor

Amnon Shashua   and   Lior Wolf

School of Computer Science and Engineering,
The Hebrew University,
Jerusalem 91904, Israel
e-mail: {shashua,lwolf}@cs.huji.ac.il

Abstract

The quadrifocal tensor which connects image measurements along 4 views is not yet well understood as its counterparts the fundamental matrix and the trifocal tensor. This paper establishes the structure of the tensor as an "epipole-homography" pairing

$$Q^{ijkl} = v'^{j} H^{ikl} - v''^{k} H^{ijl} + v'''^{l} H^{ijk}$$

where v', v'', v''' are the epipoles in views 2,3,4, H is the "homography tensor" the 3-view analogue of the homography matrix, and the indices i,j,k,l are attached to views 1,2,3,4 respectively — i.e., H^{ikl} is the homography tensor of views 1,3,4.

In the course of deriving the structure Q^{ijkl} we show that Linear Line Complex (LLC) mappings are the basic building block in the process. We also introduce a complete break-down of the tensor slices: $3 \times 3 \times 3$ slices are homography tensors, and 3×3 slices are LLC mappings. Furthermore, we present a closed-form formula of the quadrifocal tensor described by the trifocal tensor and fundamental matrix, and also show how to recover projection matrices from the quadrifocal tensor. We also describe the form of the 51 non-linear constraints a quadrifocal tensor must adhere to.

1   Introduction

The study of the geometry of multiple views has revealed the existence of certain multi-linear forms that connect image measurements of points and lines across 2,3,4 views. The coefficients of these multi-linear forms make up the fundamental matrix [8, 1] in the case of two views, the trifocal tensor [9, 4, 11, 14] of three views, and the quadrifocal tensor in case of four views [2, 16, 7, 3].

Among the three, the quadrifocal tensor is the least understood. What is known so far is somewhat fragmented and includes (i) the existence of 16 quad-linear forms per quadruple of matching points across 4 views [2, 16, 17], (ii) quadlinear forms between a pair of matching quadruples form a rank 31 system [3], (iii) the fact that the coefficients of the tensor are Grassman coordinates [2, 16], (iv) that the quadlinear forms are spanned by trilinear and bilinear forms

[2] (using, however, symbolic algebra and random camera configurations), (v) 3×3 slices of the quadrifocal tensor are rank-2 matrices [7], and (vi) equations for recovering fundamental matrices and trifocal tensors from the quadrifocal tensor [7], albeit not as far as recovering the camera projection matrices in a general manner (only for "reduced" representations [6,5]), and (vii) the source of the 51 non-linear constraints and their form is still an open issue.

In this paper we derive the quadrifocal tensor "bottom up" similarly to the way the trifocal tensor was derived in [9, 4, 10] and establish the following results. First and foremost we obtain an explicit formula that describes the tensor as a sum of epipole-homography outer-products:

$$Q^{ijkl} = v'^{j} H^{ikl} - v''^{k} H^{ijl} + v'''^{l} H^{ijk}$$

where v', v'', v''' are the epipoles in views 2,3,4, H is the "homography tensor" the 3-view analogue of the homography matrix, and the indices i, j, k, l are attached to views 1,2,3,4 respectively — i.e., H^{ikl} is the homography tensor of views 1,3,4.

In the course of deriving the representation above we find that the Linear Line Complex (LLC) mapping (introduced in the past in the context of ambiguities in reconstruction [15]) forms a basic building block in the construction of the tensor. The explicit representation allows us to introduce a complete break-down of the tensor slices: $3 \times 3 \times 3$ slices are homography tensors, and 3×3 slices are LLC mappings. Furthermore, we present a closed-form formula of the quadrifocal tensor described by the trifocal tensor and fundamental matrix, and also show how to generally recover projection matrices from the quadrifocal tensor. Finally, we describe the form of the 51 non-linear constraints a quadrifocal tensor must adhere to.

2   Notations and Background

We will be working with the projective 3D space and the projective plane. In this section we will describe the basic elements we will be working with (i) collineations of the plane (ii) camera projection matrices, (iii) Linear Line Complex (LLC) mapping, and (iv) tensor notations.

A point in $\mathcal{P}^2$ is defined by three numbers, not all zero, that form a coordinate vector defined up to a scale factor. The dual projective plane represents the space of lines which are also defined by a triplet of numbers. A point p in the projective plane coincides with a line s if and only if $p^{\top} s = 0$, i.e., the scalar product vanishes. In the projective plane any four points in general position can be uniquely mapped to any other fours points in the projective plane. Such a mapping is called *collineation* and is defined by 3×3 invertible matrices, defined up to scale. These matrices are sometimes referred to as *homographies*. If H is a homography matrix, then H^{-T} (inverse transpose) is the dual homography that maps lines onto lines.

A point in $\mathcal{P}^3$ is defined by four numbers, not all zero, that form a coordinate vector defined up to a scale factor. The dual projective space represents the space of planes which are also defined by a quadruple of numbers. The projection from

3D space to 2D space is determined by a 3×4 matrix. A useful parameterization (which is the one we adopt in this paper) is to have the 3D coordinate frame and the 2D coordinate frame of view 1 aligned. Thus, in the case of two views we have $[I; 0], [A; v']$ be the two camera matrices from 3D space to views 1,2 respectively. The matrix A is a homography matrix from view 1 to 2 whose corresponding plane is the "reference" plane, v' is the intersection point between the two camera centers (the null spaces of the respective projection matrices) and view 2 (known as the epipole). Additional views, $[B; v''], [C; v''']$, etc., must all agree on the same reference plane, In other words, the homographies $A, B, C, ..$ all form a group, i.e., the homography matrix from view 2 to 3 is BA^{-1}, for example. Note that the choice of the reference plane is free — a fact that provides 3 free parameters (the "gauge" of the system) when setting up a set of projection matrices that agree with the image measurements. If $A_\pi, A_{\bar{\pi}}$ are two homography matrices between views 1,2 associated with planes $\pi, \bar{\pi}$, then $A_\pi \cong A_{\bar{\pi}} + v'w^\top$ where w is the projection onto view 1 of the line intersection of π and $\bar{\pi}$. If λ is a line in view 2, for example, then $\lambda^\top [B; v'']$ is the plane passing through the line λ and the second camera center.

Another useful transformation between views of a 3D scene is the one resulting from a Linear Line Complex (LLC) configuration. An LLC configuration consists of a set of lines in 3D that have a common line intersection (referred to as the kernel of the set). Let L be the kernel of the set, and let its projections onto views 1,2 be l, l' respectively (see Fig. 1). Let A be a homography matrix of any plane containing L, then $G = A[l]_x$ (where $[]_x$ is the skew-symmetric matrix of cross-products, see later) is a unique transformation (does not depend on the choice of the plane) that satisfies $s'^\top A[l]_x s = 0$ for all matching lines s, s' in views 1,2 respectively arising from lines of the LLC configuration. The left and right null spaces of G are the projections of L on views 1,2.

It will be most convenient to use tensor notations from now on because the material we will be using in this paper involves coupling together pairs of collineations and epipoles into a "joint" object. When working with tensor objects the distinction of when coordinate vectors stand for points or lines matters. A point is an object whose coordinates are specified with superscripts, i.e., $p^i = (p^1, p^2, p^3)$. These are called contravariant vectors. A line in $\mathcal{P}^2$ is called a covariant vector and is represented by subscripts, i.e., $s_j = (s_1, s_2, s_3)$. Indices repeated in covariant and contravariant forms are summed over, i.e., $p^i s_i = p^1 s_1 + p^2 s_2 + p^3 s_3$. 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 (a collineation, for example), and hyperplanes (lines in $\mathcal{P}^2$) to hyperplanes, 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 hyperplanes; and a^{ij} maps hyperplanes 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
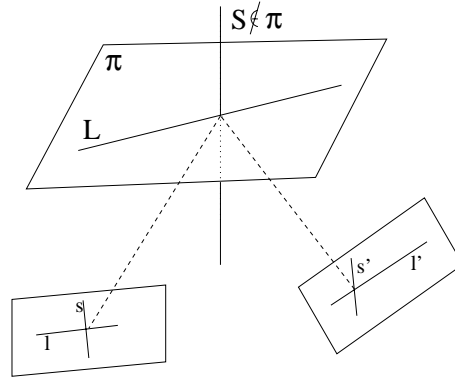
Fig. 1. A linear line complex (LLC) is a configuration of lines in 3D that have a common line intersection, the kernel, L. Let l be the projection of L in view 1 and let A be a homography of some plane containing L. Then $G = A[l]_x$ is the LLC mapping that $s'^{\top}Gs = 0$ for all matching lines s, s' in views 1,2 respectively arising from lines of the LLC configuration.

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 $C = ba^{\top}$. A 3-valence tensor has three indices, say H_i^{jk}. The positioning of the indices reveals the geometric nature of the mapping: for example, $p^i s_j H_i^{jk}$ must be a point because the i,j indices drop out in the contraction process and we are left with a contravariant vector (the index k is a superscript). Thus, H_i^{jk} maps a point in the first coordinate frame and a line in the second coordinate frame into a point in the third coordinate frame. The trifocal tensor in multiple-view geometry is an example of such a tensor. A single contraction, say $p^i H_i^{jk}$, of a 3-valence tensor leaves us with a matrix. Note that when p is $(1, 0, 0)$ or $(0, 1, 0)$, or $(0, 0, 1)$ the result is a "slice" of the tensor.

We will make extensive use of the "cross-product tensor" ϵ defined next. The cross product (vector product) operation $c = a \times b$ is defined for vectors in $\mathcal{P}^2$. The product operation can also be represented as the product $c = [a]_\times b$ where $[a]_\times$ is called the "skew-symmetric matrix of a". In tensor form we have $\epsilon_{ijk} a^i b^j = c_k$ representing the cross product 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_{ijk} is the anti-symmetric tensor defined such that $\epsilon_{ijk} a^i b^j c^k$ is the determinant of the 3×3 matrix whose columns are the vectors a, b, c. As such, ϵ_{ijk} contains $0, +1, -1$ where the vanishing entries correspond to arrangement of indecis with repetitions (21 such entries), whereas the odd permutations of ijk correspond to -1 entries and the even permutations to $+1$ entries.

3 Quadrifocal Tensor Bottom-Up

Consider 4 views with the following 3×4 projection matrices: $[I; 0], [A; v'], [B; v''], [C; v''']$ associated with views 1,2,3,4 respectively. By definition, the matrices A, B, C are homography matrices from view 1 onto 2,3,4 respectively through some reference plane π. Let P be some point in 3D projective space projecting onto p, p', p'', p''' in the four images, i.e., $p \cong [I; 0]P, p' \cong [A; v']P, p'' \cong [B; v'']P$ and $p''' \cong [C; v''']P$. Let L be some 3D line passing through P and let the projections of L onto views 3,4 be denoted by r, t, thus L is the intersection of the two planes $r^\top [B; v'']$ and $t^\top [C; v''']$. See Fig. 2 as a reference from now on.
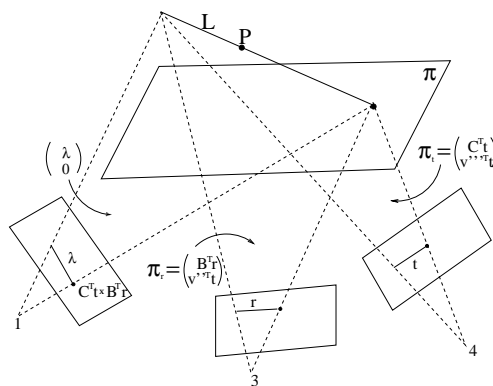
Fig. 2. The construction of a LLC mapping between views 1,2 where the kernel line L is determined by the lines r, t in views 3,4. In the construction one needs to express the projection of L in view 1 and consider the projection of the intersection point of L and the reference plane π on view 1 — the point $B^\top r \times C^\top t$.

We wish to construct an LLC mapping, a 3×3 matrix $Q(r, t)$, between views 1,2 whose kernel is the line L. Before we do so , it is worthwhile noting what we would gain from it. From the definition of an LLC mapping, we have:

$$s^\top Q(r, t) q = 0$$

for *all* lines s passing through p' and *all* lines q passing through p. In other words, $Q(r, t) = r_k t_l Q^{ijkl}$ is a 3×3 double contraction of the quadrifocal tensor (linear combination of 3×3 slices). Thus, our mission would be almost completed if we derive $Q(r, t)$ — almost, because one must separate the contribution of q, s, r, t from the contribution of the camera matrices in order to get a form $q_i s_j r_k t_l Q^{ijkl} = 0$. To conclude, we have two steps left, one is to derive $Q(r, t)$ as the LLC between views 1,2 whose kernel is L, and the second is to separate the image measurements and the camera matrices from the equation $s^\top Q(r, t) q = 0$ (this is where the power of tensor notations becomes critical).

Let $\alpha r^\top[B;v''] + \beta t^\top[C;v''']$ be the pencil of planes whose axis is L, parameterized by the scalars α, β. Let λ be the projection of L onto view 1, thus the plane $\lambda^\top[I;0]$ belongs to the pencil:

$$\begin{pmatrix} \lambda \\ 0 \end{pmatrix} \cong \alpha \begin{pmatrix} B^\top r \\ v''^\top r \end{pmatrix} + \beta \begin{pmatrix} C^\top t \\ v'''^\top t \end{pmatrix}$$

Therefore, $\lambda \cong (v'''^\top t) B^\top r - (v''^\top r) C^\top t$. Let σ be any plane through L, then the homography matrix from view 1 to 2 through σ is, $A_\sigma \cong A + v'n^\top$, where n is the projection of the intersection line between planes π, σ onto view 1. By definition of LLC mapping we have:

$$Q(r,t) = A_\sigma[\lambda]_x \cong A[\lambda]_x + v'(n \times \lambda)^\top.$$

Note that $n \times \lambda$ is the projection of the intersection point between L and π onto view 1, and furthermore,

$$n \times \lambda \cong C^\top t \times B^\top r$$

because $C^\top t$ is the projection of the intersection of the planes π and $t^\top[C;v''']$ in view 1, and $B^\top r$ is the projection of the intersection of the planes π and $r^\top[B;v'']$ in view 1. The intersection of the two lines in view 1 is the projection of the intersection point between L and π onto view 1. Taken together, we have an explicit equation for $Q(r,t)$:

$$Q(r,t) = v'(t^\top C \times r^\top B) - (v''^\top r)A[C^\top t]_x + (v'''^\top t)A[B^\top r]_x \qquad (1)$$

And the quadlinearity $s^\top Q(r,t)q = 0$ for all lines q, s, r, t in views 1,2,3,4 respectively that coincide with their respective image point p, p', p'', p''' is:

$$(s^\top v')(t^\top C \times r^\top B)q - (v''^\top r)s^\top A[C^\top t]_x q \qquad (2)$$
$$+ (v'''^\top t)s^\top A[B^\top r]_x q = 0.$$

We have so far observed that the LLC mapping is a basic building block in constructing the quadlinearity above. It is worthwhile noting that the quadlinearity above can be also derived "top-down" by a determinant expansion, as follows. Since the planes $q^\top[I;0]$, $r^\top[A;v']$, $s^\top[B;v'']$, $t^\top[C;v''']$ meet at the point P, the determinant below must vanish:

$$\begin{vmatrix} q^\top & 0 \\ s^\top A & s^\top v' \\ r^\top B & r^\top v'' \\ t^\top C & t^\top v''' \end{vmatrix} = 0$$

After expanding the determinant by its fourth column we obtain Eqn. 2 above. In order to continue, we introduce another basic block the "homography tensor" of three views. Referring to Fig. 3, consider the line of intersection $\bar{L}$ between the plane π and the plane $r^\top[B;v'']_x$. Consider some point $\bar{P}$ on $\bar{L}$ and its projections $\bar{p}, \bar{p}', p''$ onto views 1,2,3 respectively. Since the projection of $\bar{L}$
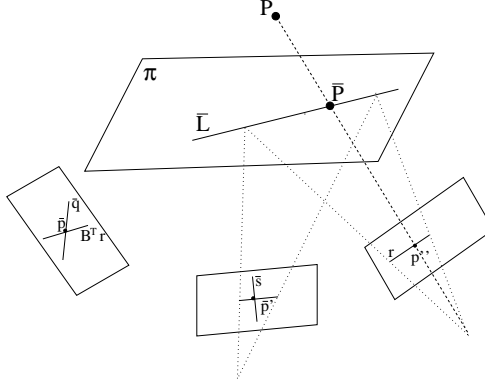
Fig. 3. The construction of the Homography Tensor. A line r in the third view determines by intersection of the viewing plane and the plane π a line $\bar{L}$ in 3D forming a kernel of an LLC mapping $A[B^\top r]_x$ between views 1,2. In tensor form we have $q_i s_j r_k H^{ijk} = 0$ where q, s, r are lines through a matching triplet of points corresponding to points on the plane π. Thus, H^{ijk} is a 3-view analogue to the homography matrix. See [12] for more details.

onto view 1 is $B^\top r$, then the LLC mapping between views 1,2 with $\bar{L}$ as the kernel is $A[B^\top r]_x$. By definition of LLC mapping, we have $\bar{s}^\top A[B^\top r]_x \bar{q} = 0$ for all lines $\bar{q}, \bar{s}$ that are coincident with $\bar{p}, \bar{p}'$ respectively.

From this point on we will move to tensor notations - necessary step in order to separate the image measurements q, s, r, t from the camera projection matrices A, B, C, v', v'', v'''. We adopt the notation that indices i, j, k, l are associated exclusively with views 1,2,3,4. For example, since v' is a point (epipole) in view 2, then when placed in a tensor equation it will always appear as v'^j; likewise s is a line in view 2, then in a tensor equation it will always appear as s_j.

Rewriting $\bar{s}^\top A[B^\top r]\bar{q} = 0$ in tensor form we have:

$$\bar{q}_i \bar{s}_j r_k \big(\epsilon^{iun} a_u^j b_n^k\big) = 0$$

and denote

$$H^{ijk} = \epsilon^{iun} a_u^j b_n^k. \tag{3}$$

The tensor H^{ijk} is a *homography* (collineation) mapping of the plane π associated with 3 views. For example,

$$\bar{q}_i \bar{s}_j H^{ijk} \cong p''^k$$
$$\bar{s}_j r_k H^{ijk} \cong \bar{p}^i$$
$$\bar{q}_i r_k H^{ijk} \cong \bar{p}'^j .$$

Just like a homography matrix it can map directly a point in any view onto its matching point in any other view (not described here). Its 3×3 slices are LLC maps: we saw that $r_k H^{ijk}$ is the LLC map $A[B^\top r]_x$ between views 1,2 of

the line $\bar{L}$; likewise, $s_j H^{ijk}$ is the LLC map $B[A^\top s]_x$ between views 1,3 whose kernel is the intersection of π and the plane $s^\top[A; v']$, and $q_i H^{ijk}$ is the LLC map $B[q]_x A^\top$ between views 2,3 whose kernel is the intersection of π and the plane $q^\top[I; 0]$. Note also that $r^\top B[q]_x A^\top s = (B^\top r)(q \times A^\top s) = q^\top(B^\top r \times A^\top s)$. The tensor H^{ijk} is therefore the extension of the 2-view homography matrix and is referred to as "Homography Tensor" (Htensor in short) — further details are beyond the scope of this paper and the reader is referred to [12].

Returning to the quadlinearity in eqn. 2, we notice that all three terms consist of homography tensors of the plane π of views $(1, 3, 4)$ and $(1, 2, 4)$ and $(1, 2, 3)$. Using our notation of indices, we have H^{ikl} for the Htensor of views $(1, 3, 4)$, H^{ijl} for the Htensor of views $(1, 2, 4)$ and H^{ijk} for the Htensor of views $(1, 2, 3)$. Therefore, the quadrifocal tensor is:

$$Q^{ijkl} = v'^j H^{ikl} - v''^k H^{ijl} + v'''^l H^{ijk} \tag{4}$$

and the quadlinearity in eqn. 2 is simply

$$q_i s_j r_k t_l Q^{ijkl} = 0.$$

Finally, the form of Q^{ijkl} does not depend on the position of the reference plane π. Changing the reference plane to $\bar{\pi}$ results in the new set of camera projection matrices $[I; 0], [A + v'w^\top], [B + v''w^\top], [C + v'''w^\top]$ where w is the projection onto view 1 of the intersection line between π and $\bar{\pi}$. By substitution in eqn 4 one notices that the terms with w drop out — details are in the full version of this paper [13].

4 Properties of the Quadrifocal Tensor

Since every quadruple of lines q, s, r, t coincident with the matching points p, p', p'', p''', respectively, contributes one equation $q_i s_j r_k t_l Q^{ijkl} = 0$ we have a total of 16 linearly independent equations per matching quadruple of points. Hartley [3] first noticed and proved that two quadruples contribute only 31 linearly independent equation, and every additional quadruple contributes one less equation, thus 6 matching quadruples contribute $16 + 15 + 14 + 13 + 12 + 11 = 81$ linearly independent equations for the 81 coefficients of the tensor. One can obtain a simpler geometric proof of why this is so:

Each quadlinearity is spanned by a set of 16 quadlinearities

$$q_i^\rho s_j^\mu r_k^\nu t_l^\delta Q^{ijkl} = 0 \quad \rho, \mu, \nu, \delta = 1, 2$$

where q_i^1, q_i^2 are two lines, say the horizontal and vertical scan lines, passing through p, etc. Given any lines q, s, r, t passing through the matching points, then each line is linearly spanned by the horizontal and vertical scan lines and this linear combination carries through to a linear combination of the 16 quadlinearities above. Given a second matching quadruple, $\bar{p}, \bar{p}', \bar{p}'', \bar{p}'''$, then the quadlinearity resulting from taking the lines $q = p \times \bar{p}, s = p' \times \bar{p}', r = p'' \times \bar{p}''$ and

$t = p''' \times \bar{p}'''$ is *spanned* by the 16 quadlinearities of the first set of matching quadruple p, p', p'', p''' (see Fig. 4). Thus, the two subspaces have a single non-trivial intersection and the total rank is 31 (instead of 32). Likewise, the n'th additional quadruple of matching points has $n - 1$ quadlineairites spanned by the previous $n - 1$ subspaces.
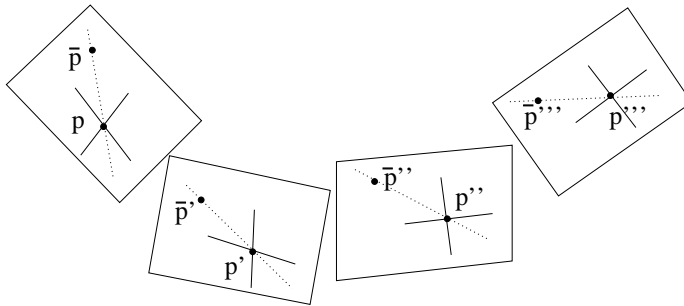
Fig. 4. A quadruple of matching points provides 16 constraints. A second quadruple provides only 15 additional constraints because the constraint defined by the lines connecting the the two sets of points is already covered by the 16 constraints of the first quadruple.

4.1 Slicing Breakdown

We move our attention to the breakdown of slices of the tensor. From the construction of the tensor and the symmetrical role that all views play (all indices are contravariant, unlike the trifocal tensor which has both covariant and contravariant indices) we conclude:

Theorem 1. *Let x, y, $x \neq y$, be a pair of indices from the set $\{i, j, k, l\}$, let M_x, M_y be the camera projection matrices associated with the choice of view that the indices x, y represent, and let δ, μ range over the standard basis $(1, 0, 0), (0, 1, 0), (0, 0, 1)$. Every 3×3 slice $\delta_x \mu_y Q^{ijkl}$ of the quadrifocal tensor corresponds to an LLC mapping between the remaining views not represented by x, y and whose kernel is the line intersection of the planes $\delta^\top M_x$ and $\mu^\top M_y$. For example, if $x = i, y = j$ then $\delta_i \mu_j Q^{ijkl}$ provide 9 slices, each slice is an LLC map between views 3,4 and whose kernel is the intersection of the planes $\delta^\top [I; 0]$ and $\mu^\top [A; v']$.*

Note that in particular a 3×3 slice is a rank-2 matrix (as observed by [7]), but not every rank-2 matrix is an LLC mapping. Note also that any linear combination of the 9 slices $\delta_x \mu_y Q^{ijkl}$, for a fixed choice of x, y, is also an LLC mapping. Thus the finding above is a stronger constraint on the structure of the 3×3 slices of the tensor then what was known so far.

Next we state (proof in [13]) that every $3 \times 3 \times 3$ slice of the quadrifocal tensor corresponds to a homography tensor:

Theorem 2. *Let x be an index from the set $\{i, j, k, l\}$, let M_x be the camera projection matrices associated with the choice of view that the index x represents, and let δ range over the standard basis $(1,0,0), (0,1,0), (0,0,1)$. Then every $3 \times 3 \times 3$ slice $\delta_x Q^{ijkl}$ is a homography tensor between the three remaining views of the plane $\delta^\top M_x$.*

4.2 Quadrifocal Constructed from Trifocal and Bifocal

We move our attention to the construction of the quadrifocal tensor from lower order tensors: the trifocal tensor of 3 views and the fundamental matrix of two views. Using symbolic algebra on random camera configurations, Faugeras & Mourrain [2] have concluded that the quadlinear forms are spanned by trilinear and bilinear ones. We will use the LLC building block to derive a closed form formula on representing the quadrifocal tensor as a function of trifocal tensors and fundamental matrices. Let $Y(\delta, \mu) = \delta_i \mu_j Q^{ijkl}$ be the LLC slice of Q between views 3,4. Then,

$$Y(\delta, \mu) = [\delta_i \mu_j \mathcal{T}_l^{ij}]_x F_{34} [\delta_i \mu_j \mathcal{T}_k^{ij}]_x$$

where $\mathcal{T}_l^{ij}$ is the trifocal tensor of views $(4, 1, 2)$ and $\mathcal{T}_k^{ij}$ is the trifocal tensor of views $(3, 1, 2)$, and F_{34} is the fundamental matrix between views 3,4.

To see why this expression holds, let L be the line intersection of the planes $\delta^\top [I; 0]$ and $\mu^\top [A; v']$. Note that $\delta_i \mu_j \mathcal{T}_l^{ij}$ is the projection of L onto view 4, and $\delta_i \mu_j \mathcal{T}_k^{ij}$ is the projection of L onto view 3. The fundamental matrix flanked by both sides by the skew-symmetric matrix of the projections of the kernel line is the LLC map between views 3,4.

By varying δ, μ to range over $(1,0,0), (0,1,0), (0,0,1)$ we obtain a closed form formula of the nine 3×3 slices

$$Q^{11kl}, Q^{12kl}, ..., Q^{33kl}$$

denoted by $Y(1, 1), ..., Y(3, 3)$, making up the quadrifocal tensor. However, each slice is up to scale. The 9 scale factors $\lambda_1, ..., \lambda_9$ can be recovered (up to a global scale) by setting δ, μ each to be $(1, 1, 1)$ in which case we have:

$$Y(\delta, \mu) = \lambda_1 Y(1, 1) + ... + \lambda_9 Y(3, 3)$$

which provides a linear system for recovering the scale factors. In conclusion we have:

Theorem 3. *The quadrifocal tensor can be constructed from two trifocal tensors and one fundamental matrix.*

4.3 Fundamental Matrix from Quadrifocal Tensor

We next move our attention to the construction of lower order tensors from the quadrifocal tensor. We will start with the fundamental matrix. Let $Y(\delta, \mu) = \delta_i \mu_j Q^{ijkl}$ be the LLC slice of Q between views 3,4 with kernel line L. Let l''', l'' be

the projections of L onto views 3,4 respectively, thus $Yl'' = 0$ and $Y^\top l''' = 0$. Let F_{34} be the fundamental matrix between views 3,4. Let r be some line in view 3, then Yr is a point coincident with $F_{34}[l'']_x r$ (see Fig. 5). Thus, $r^\top Y^\top F_{34}[l'']_x r = 0$ for all r, therefore $Y^\top F_{34}[l'']_x$ is a skew-symmetric matrix. The skew-symmetric constraint provides 2 linear equations for F_{34}. By varying δ, μ over the standard basis we obtain nine LLC slices, each provides 2 constraints on F_{34}.
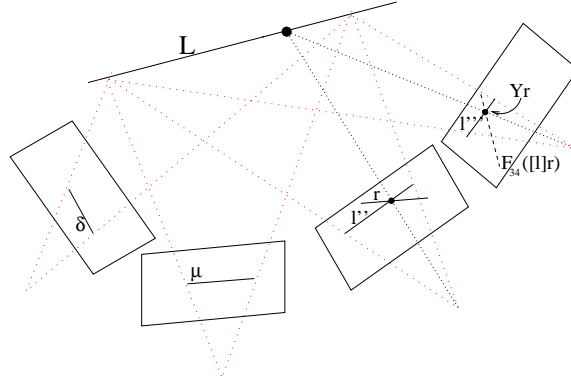
Fig. 5. Constructing the fundamental matrix F_{34} between views 3,4 from the Quadrifocal tensor. We use the 3×3 slices $\delta_i \mu_j Q^{ijkl}$ which form a LLC mapping Y between views 3,4 with a kernel line L (whose projection in views 1,2 are δ, μ). The left and right null spaces of Y are the projections of L on views 3,4. If r is some line in view 3, the Yr is a point coincident with $F_{34}[l'']_x r$, thus $Y^\top F_{34}[l'']$ is a skew-symmetric matrix providing 2 constraints on F_{34}.

4.4 Trifocal from Quadrifocal

To recover the trifocal tensor, say $\mathcal{T}_i^{jk}$ of views $(1,2,3)$ consider the following. Consider two lines δ_l, μ_l in view 4. Then $H^{ijk} = \delta_l Q^{ijkl}$ and $\bar{H}^{ijk} = \mu_l Q^{ijkl}$ are two Htensors of views $(1,2,3)$ associated with two distinct planes. Then, $s_j r_k H^{ijk}$ is a point in view 1, and $s_j r_k \bar{H}^{ijk}$ is another point in view 1, but these two points lie on the line $s_j r_k \mathcal{T}_i^{jk}$. Thus, we have the constraint:

$$s_j r_k \mathcal{T}_i^{jk} \cong \left(s_j r_k H^{ijk} \right) \times \left(s_j r_k \bar{H}^{ijk} \right)$$

By varying s, r to range over the standard basis, we obtain slices (each up to scale) of the trifocal tensor:

$$\mathcal{T}_i^{11}, ..., \mathcal{T}_i^{33}.$$

The nine scale factors are recovered in two stages. First, by setting $s = (1, 1, 1)$ and varying r over the standard basis we obtain a linear system for sets of three

scale factors. We are thus left with three 3×3 slices $\mathcal{T}_i^{j1}, \mathcal{T}_i^{j2}, \mathcal{T}_i^{j3}$ each up to scale. By setting both s, r to $(1, 1, 1)$ we obtain a linear system for the three remaining scale factors.

4.5 Projection Matrices from Quadrifocal Tensor

Finally, we move our attention to the construction of the camera projection matrices from the quadrifocal tensor. Consider again eqn. 4:

$$Q^{ijkl} = v'^j H^{ikl} - v''^k H^{ijl} + v'''^l H^{ijk}.$$

The epipolar points v', v'', v''' are the null spaces of the fundamental matrices which were recovered in Section 4.3. Thus, we have 81 constraints for solving for the three Htensors H^{ikl}, H^{ijl} and H^{ijk} that together form 81 unknowns. There are two interesting points to make here. First, all three Htensors correspond to the same reference plane in space — thus if we extract the constituent homography matrices out of the Htensors, then together with the epipoles we have an admissible set of camera projection matrices. Second, due to the gauge-invariance property of the multi-view tensors we have three degrees of freedom thus the rank of the estimation of the 81 variables of the three Htensors is 78. We are free to choose any solution spanned by the null space (the choice will determine the gauge, i.e., the location of the reference plane).

What is left is to show how to extract the homography matrices from view 1 to 2,3,4 from the Htensors. Consider for example H^{ijk} the Htensor of views $(1, 2, 3)$ with constituent homography matrices A, B. Because 3×3 slices correspond to LLC maps, it is possible to extract from them the homographies A, B. For example, $\delta_k H^{ijk}$ produces an LLC map $A[B^\top \delta]_x$ — by allowing δ to range over the standard basis $(1, 0, 0), (0, 1, 0), (0, 0, 1)$, we obtain three such matrices, denoted by E_1, E_2, E_3. We have that $AE_i^\top + E_i A^\top = 0$, $i = 1, 2, 3$, thus providing 18 linear equations for A. Similarly, one can find in this manner the other homographies B, C — each up to scale. The three scale factors can be determined by using eqn. 4 again, this time the Htensors are constructed from the homography matrices where the unknowns are the three scale factors.

4.6 The 51 Non-linear Constraints

The quadrifocal tensor is represented by 29 parameters $(44 - 15 = 29)$ thus we expect 51 non-linear constraints ("admissibility" constraints) on the 81 coefficients (up to scale) of the tensor. This issue has so far been unresolved. Heyden [7] conjectures that the source of these constraints comes from the rank-2 property of the 3×3 slices. But as we saw in Theorem 1 the matter is more complicated because the slices are LLC maps, and not every rank-2 matrix is an LLC map. We will use the $3 \times 3 \times 3$ slices, the Htensors, to buildup those 51 constraints, as described below.

Consider the three $3 \times 3 \times 3$ slices $\delta_l Q^{ijkl}$ by letting δ range over the standard basis $(1, 0, 0), (0, 1, 0), (0, 0, 1)$. From Theorem 2 we know that these slices

are Htensors denoted by $H^{ijk}_{(m)}$, $m = 1, 2, 3$. Note that all three Htensors are of views $(1, 2, 3)$ each corresponding to a different plane. Let $A_{(m)}, B_{(m)}$ be the constituent homography matrices of the m'th Htensor. We know that $A_{(1)}, A_{(2)}, A_{(3)}$ are homography matrices from view 1 to 2 corresponding to three different planes π_1, π_2, π_3, and $B_{(1)}, B_{(2)}, B_{(3)}$ are homography matrices from view 1 to 3 corresponding to the same planes π_1, π_2, π_3, respectively.

We will divide the 51 constraints into two sets. The first set consists of $9 \times 3 = 27$ constraints which describe the constituent homography matrices from their corresponding Htensors. The second set consists of 24 constraints which embody the relationship between $A_{(m)}, B_{(m)}$ as described above.

Recall that an Htensor produces 18 linear constraints for each of its homography matrices. Consider the three 3×3 slices $H^{ij1}_{(1)}, H^{ij2}_{(1)}, H^{ij3}_{(1)}$ of the Htensor $H^{ijk}_{(1)}$ and denote the resulting matrices by E_1, E_2, E_3. We know that $A_{(1)}E_i^\top + E_i A_{(1)}^\top = 0$, $i = 1, 2, 3$, which provide 18 constraints for $A_{(1)}$. Choose 8 constraints from these 18 constraints. Thus, each entry of the matrix $A_{(1)}$ can be represented by a determinant expansion of an 8×8 matrix whose components come from the tensor elements that participate in those 8 constraints. The remaining 10 constraints must be of rank 8, thus by substituting $A_{(1)}$ in the remaining 10 constraints we have 8 polynomials of degree 9 on the entries of E_1, E_2, E_3 that must be satisfied in order that $A_{(1)}E_i^\top + E_i A_{(1)}^\top = 0$ for $i = 1, 2, 3$. In this way we may solve for $B_{(1)}$, but this does not add new constraints because we are using the same information used to derive $A_{(1)}$. The scale of the Htensor is set because it is a slice of the quadrifocal tensor - yet the scales of $A_{(1)}, B_{(1)}$ are arbitrary. Therefore, there is another constraint that is captured as follows. Let α be some unknown scale of $A_{(1)}$, such that $H_{(1)} = \alpha A_{(1)} \otimes B_{(1)}$, where $\otimes$ is a short-cut denoting the Htensor equation 3. This provides 27 equations. Choose any two of them and eliminate α — the results is non-linear equaiton in the elements of $A_{(1)}$ and $B_{(1)}$. Taken togetner, we have 9 non-linear constraints from $H_{(1)}$ and since this is true for $m = 2, 3$ as well, we have 27 non-linear constraints.

We have 27 constraints, and have represented $A_{(m)}, B_{(m)}$ by determinant expansions of the entries of the quadrifocal tensor. Because $A_{(m)}$ are between view 1 to 2, and $B_{(m)}$ are between view 1 to 3, and each pair $A_{(m)}, B_{(m)}$ are associated with the same plane π_m, $m = 1, 2, 3$ respectively, we have:

$$A_{(2)} = \lambda_1 A_{(1)} + v' w^\top$$
$$\lambda_2 B_{(2)} = \lambda_3 B_{(1)} + v'' w^\top$$
$$A_{(3)} = \lambda_4 A_{(1)} + v' \bar{w}^\top$$
$$\lambda_5 B_{(3)} = \lambda_6 B_{(1)} + v'' \bar{w}^\top$$

where $w, \bar{w}$ is the projection onto view 1 of line intersection between π_1, π_2 and between π_1, π_3 respectively. We have 36 equations in $A_{(m)}, B_{(m)}$ (which are represented in terms of determinant expansions of the quadrifocal tensor elements), the epipoles v', v'' (which are also represented as non-linear functions of the quadrifocal elements — as we saw in Section 4.3) and 12 variables: 6 from

$w, \bar{w}$ and the scales $\lambda_1, ..., \lambda_6$. By elimination of the 12 variables we are left with $36 - 12 = 24$ polynomials on the elements of the quadrifocal tensor alone. Taken together we have $27 + 24 = 51$ constraints (plus the constraint arising from the global scale factor). The two sets of constraints are algebraically independent as they use different information: the first set arises from the rank-2 LLC map of the 3×3 slice of the Htensors which provides constraints on the homography matrices, and the second set arises from the *relationships* between the individual homography matrices.

5   Summary

We have derived an explicit form of the quadrifocal tensor (eqn. 4) analogous to the forms of the lower order tensors of multi-view geometry, the trifocal tensor and the fundamental matrix. The lower-order tensors have explicit forms as an epipole-homography outer-products. The fundamental matrix is formed by a single epipole-homography coupling $F = [v']_x A$, and the trifocal tensor is formed by two pairs of epipole-homography: $\mathcal{T}_i^{jk} = v'^j b_i^k - v''^k a_i^j$. We have shown that the quadrifocal tensor is formed by three pairs of epipole-Htensor couplings:

$$Q^{ijkl} = v'^j H^{ikl} - v''^k H^{ijl} + v''''^l H^{ijk},$$

where the Htensors are 3-view analogue to the 2-view homography matrices (they perform the operation of collineations between views of a 2D configuration).

Using the explicit form of the tensor and the tool of LLC mapping for analysis, the slicing breakdown is relatively simple: the 3×3 slices of the quadrifocal tensor form LLC mappings — in particular these are rank-2 matrices, but not every rank-2 matrix is an LLC mapping. The $3 \times 3 \times 3$ slices are homography tensors (Htensors), i.e., collineations of 3-view sets — which is analogous to the covariant 3×3 slices of the trifocal tensor which form homography matrices. Moreover, the construction of the quadrifocal tensor from the lower-order tensors and, vice-versa, the construction of the lower-order tensors from the quadrifocal tensor as well as the camera projection matrices were presented in detail — again, become relatively straightforward once the explicit form (eqn. 4) is available.

Finally, the explicit form and the discovery that the $3 \times 3 \times 3$ slices are homography tensors has provided a simple route to deriving the 51 non-linear constraints that all quadrifocal tensors must adhere to.

References

1. 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.
2. 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.

3. R.I. Hartley. Multilinear relationships between coordinates of corresponding image points and lines. In *Proceedings of the International Workshop on Computer Vision and Applied Geometry*, Nordfjordeid, Norway, August 1995. International Sophus Lie Center.

4. R.I. Hartley. Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision*, 22(2):125–140, 1997.

5. R.I. Hartley. Computation of the quadrifocal tensor. In *Proceedings of the European Conference on Computer Vision*, pages 20–35, Freiburg, Germany, June 1998.

6. 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.

7. A. Heyden. A common framework for multiple view tensors. In *Proceedings of the European Conference on Computer Vision*, pages 3–19, Freiburg, Germany, June 1998.

8. H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

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

10. 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.

11. 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.

12. A. Shashua and Lior Wolf. Homography tensors: On algebraic entities that represent three views of static or moving planar points. *In these proceedings.*

13. A. Shashua and Lior Wolf. On the Structure and Properties of the Quadrifocal Tensor. Computer Science and Engineering TR99-38, Hebrew University, Dec. 1999.

14. M.E. Spetsakis and J. Aloimonos. Structure from motion using line correspondences. *International Journal of Computer Vision*, 4(3):171–183, 1990.

15. G. Stein and A. Shashua. On degeneracy of linear reconstruction from three views: Linear line complex and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(3):244–251, 1999.

16. B. Triggs. Matching constraints and the joint image. In *Proceedings of the International Conference on Computer Vision*, pages 338–343, Cambridge, MA, June 1995.

17. M. Werman and A. Shashua. Elimination: An approach to the study of 3D-from-2D. In *Proceedings of the International Conference on Computer Vision*, June 1995.

18. Lior Wolf and A. Shashua. Richly occluded scenes and the problem of reprojection versus reconstruction. Computer Science and Engineering TR99-34, Hebrew University, 1999.