Fundamental Tensor: On the Geometry of Three Perspective Views

Amnon Shashua and Michael Werman

Hebrew University of Jerusalem
Institute of Computer Science
91904 Jerusalem
Israel

Abstract.

Reconstruction of 3D space from 2D images is generally a "2-view" problem. That is, the algebra and geometry of the problem is all contained in measurements across two views, the rest (i.e., when more than two views are available) is statistics.

This paper embarks on the investigation of the algebra and geometry of three views. Based on the recent result [18] showing the existence of certain trilinear functions of three views with a corresponding tensor of 27 intrinsic coefficients, we derive a set of relations between the tensor and geometric invariants of the 3D world, intrinsic structures of two views (fundamental matrix and epipoles), and show the existence of new intrinsic structures (beside the tensor itself) associated with three views.

On the practical side, the tensor offers a host of new algorithms for recovering 3D information from 2D views. These methods cut through the epipolar geometry (i.e., three views are used simultaneously, rather than in pairs, reconstruction proceeds independently of epipolar geometry), make room for statistics (additional views can be incorporated to obtain further redundant equations, just like in traditional 2-view-based schemes), and generally exploit the information available from measurements across views in a more efficient manner than any technique based on 2-view geometry.

1 Introduction

The geometric and algebraic relationship between three-dimensional objects in the world and their two-dimensional images, created by a pin-hole camera model, can be divided roughly into two categories. The first category is about establishing invariants of 3D world against camera transformations, i.e., measurements across two or more views that are invariant to the camera geometries. The second category is about establishing functions of views that are invariant to the 3D world, i.e., functions of image coordinates across two or more views that hold regardless of the 3D object whose views are at hand. For example, projective structure from two views [2, 7, 20], relative affine structure from two views [19, 21, 22, 17] and affine structure from two views (cf. [10]) are examples of the first category, which generally involve questions of recovery of 3D information from multiple 2D views. Examples of the second category include the epipolar geometry (bilinear function of two views) embodied in the fundamental matrix [11, 2], and the trilinear functions of three views [18] (both reduce to linear functions under orthographic projection originally discovered by [8, 26]).

The connection between the two categories — recovery and invariant functions — has been thoroughly investigated only in connection with epipolar geometry. That is, the connection between the fundamental matrix, the epipoles, translational and rotational components of calibrated camera motion, intrinsic structures (plane at infinity, intrinsic singular plane), affine structure, relative affine structure and projective structure is well understood (see following section for review). In fact, the state-of-the-art of 3D recovery from corresponding points across multiple 2D views relies solely on fundamental structures associated with two views — the additional views simply add a statistical component whose typical forms are least-squares, recursive estimation, non-linear optimization based on variants of Newton methods, and so forth.

Invariant functions of three views have been recently discovered in [18]. The functions have a trilinear form with 27 intrinsic coefficients that form a tensor. Given the fundamental role of the invariant function of two views (epipolar geometry) to 3D recovery of invariants, the natural goal taken here is to unfold the relations between this tensor and 3D invariants, intrinsic geometric structures, and 2-view geometry and algebra. This goal has also a significant practical motivation as well. Performance evaluation of the trilinear functions on real images for purposes of reprojection (synthesizing novel views from a pair of model views) has demonstrated an *order of magnitude* improvement compared to conventional techniques that rely on epipolar geometry [18]. Hence, questions of 3D recovery may be better solved using the tensor, rather than through conventional 2-view structures (see also the Discussion section later in this paper)

The current literature on 3-view geometry is focused on 3D recovery from line correspondences [5, 27, 23, 13]. It is known that 13 corresponding lines across three views are sufficient for 3D reconstruction. Recently, and independently of our investigation, Hartley has observed ([4] and private communication) that a certain arrangement of the coefficients associated with the trilinear functions yields the entire set of parameters (arranged in three matrices) associated with lines across three views. Hartley's observation is one way to unfold the role of the tensor in the geometry of three views. Our investigation takes a different approach unrelated to lines. We show that there are three ways to break down the tensor into sets of matrices, one of them yields the matrices associated with lines as Hartley observed, the other two decompositions yield structures and results that penetrate much deeper and with greater practical consequences on the design of new algorithms than the decomposition into line parameters of three views. We will review Hartley's observation in more detail later in this paper. Other connections to the recent literature, albeit less direct, include the work of Quan [15] on recovering 6-point invariants (up to a finite number of solutions) from three views, and an observation by Tomasi (private communication) on connections between the trilinear coefficients and the frame parameters of [24].

The major result (Theorem 1) we obtain in our investigation of the tensor of trilinear coefficients is the existence of new intrinsic structures (three intrinsic homography matrices) that completely remove the role of 2-view structures (fundamental matrix and epipoles) for 3D recovery. In other words, it is not just the possibility of recovering 2-view structures from the tensor that makes the tensor interesting, it is the possibility of cutting through any involvement of the fundamental matrix, epipoles and any other intrinsic structure associated with two views that gives the tensor a fundamental role (and for that reason we call it "fundamental tensor"). Corollaries of this result include: (i) showing that it is possible to obtain projective structure of the scene *directly* from the tensor without recovering any 2-view structure, (ii) 2-view structures (fundamental matrix, epipoles) are recoverable (linearly) from the tensor in a variety of ways, (iii) closed form solutions for Euclidean structure are derived from the tensor, and (iv) there is no need for fiducial points for 3D recovery — one of the views turns out to provide the projective reference frame in an intrinsic manner.

We start with a somewhat extensive review of the state-of-the-art of 2-view geometry as has been unfolded in the most recent literature, then proceed to review the trilinearity relation across three views and the associated 27 intrinsic coefficients, and end the review with Hartley's observation on the connection between the trilinear coefficients and the parameters of lines across three views. The review is necessary because progress in the area of 3D-from-2D geometry is moving relatively fast, hence a pause for regrouping and unifying various results is helpful to better understand the contributions of this paper.

The novel contributions of this paper start from Section 3.

2 Preliminaries

The connecting thread across "2-view geometry" (invariant 3D structure, and intrinsic structures), and "3-view geometry" (trilinearity result, fundamental tensor, intrinsic structures) is the mapping between the images of a fiducial (reference) *plane* in 3D projective space $\mathcal{P}^3$. The connection between fiducial planes and projective structure from two views was originally described in [20], and a wide spectrum of results connecting the projective transformation between two views of a fiducial plane, referred from now as the "homography matrix", and intrinsic structures of 2-view geometry can be found in [16, 6, 19, 14, 21, 12, 22, 17, 9]. We will start with reviewing the state-of-the-art of 2-view geometry from the homography matrix standpoint, then proceed to the "trilinearity" result [18] which forms the basis for this paper, and end this section with the connection to lines across three views.

As for notations, we consider object space to be the three dimensional projective space $\mathcal{P}^3$, and image space to be the two dimensional projective space $\mathcal{P}^2$. An object is modeled as a set of points in $\mathcal{P}^3$ and its views (images) are denoted by $\psi_i \subset \mathcal{P}^2$. The projection centers will be denoted by O, O', O'' for views ψ_1, ψ_2, ψ_3, respectively (we will be considering at most three views in this paper). Image coordinates are noted by $p = (x, y, 1)^\top, p' = (x', y', 1)^\top$ and $p'' = (x'', y'', 1)^\top$ for views ψ_1, ψ_2, ψ_3, respectively. Let M denote the transformation from the (unknown) camera coordinate frame to the adopted coordinization of ψ_1, i.e., $zM^{-1}p$ is the location of the object point in the first camera coordinate frame. Similarly, M' and M'' denote the internal parameters of the cameras associated with views ψ_2 and ψ_3, respectively. Let R and T stand for the rotational and translational components from the first to the second camera coordinate frames, hence:

$$p' \cong M'RM^{-1}p + \frac{1}{z}M'T$$

is the transformation relating two corresponding points p, p'. Similarly, R'' and T'' are the rotational and translational components from the first to the third camera coordinate frames, and

$$p'' \cong M''R''M^{-1}p + \frac{1}{z}M''T''.$$

The vector $v' \cong M'T$ is known as the epipole (located on ψ_2) which is the point at the intersection of the line OO' with the image plane of the second camera. Similarly, $v'' \cong M''T''$ is the epipole on ψ_3.

The symbol $\cong$ denotes equality up to a scale, GL_n stands for the group of $n \times n$ matrices, and PGL_n is the group defined up to a scale. The skew-symmetric matrix associated with vector product $u \times w$ is denoted by $[u]$, i.e., $[u]w = u \times w$ for any vector w.

2.1 2-view Geometry via the Homography Matrix: Review

The intrinsic structures associated with two views, say between ψ_1 and ψ_2, can be conveniently described through a homography matrix, as follows. Let π be some arbitrary plane in $\mathcal{P}^3$ parameterized by normal vector n and distance to origin d with respect to the first camera frame, then the homography matrix A_π is described below [3, 25, 21]:

$$A_\pi \cong M'(R + \frac{1}{d}Tn^\top)M^{-1}. \tag{1}$$

The transformation A_π registers the points of π across the two images, i.e., if p, p' are projections of a point $P \in \pi$, then $A_\pi p \cong p'$. We have the following properties: (i) $A_\pi v \cong v'$, where $v \cong MR^\top T$ is the corresponding epipole on ψ_1, and (ii) $p'^\top [v'] A_\pi p = 0$ for any corresponding pair of image points. The first property is that the homography matrix aligns the epipoles, regardless of π. This property provides a simple method for recovering the homography matrix from corresponding pairs, once the epipoles v, v' are known: take three corresponding pairs p_i, p'_i, $i = 1, 2, 3$, (their corresponding object points determine the plane π), then the relation $A_\pi p_i \cong p'_i$, and $A_\pi v \cong v'$ provides a linear system of equations to solve for the homography matrix [20, 16]. The second property is the epipolar geometry: $F = [v']A_\pi \cong [v']M'RM^{-1}$ is known as the "fundamental" matrix [2], which we readily see can be recovered directly from eight corresponding image pairs (albeit is extremely numerically sensitive in the presence of measurement errors). Note that F is an intrinsic structure of two views because it does not depend on the choice of π. It also readily follows that for any π we have $A_\pi{}^\top F + F^\top A_\pi = 0$. For notation purposes, we denote F'' to be the fundamental matrix between ψ_1, ψ_3, i.e., $p''^\top F'' p = 0$.

Among the fiducial planes one can distinguish two intrinsic planes: the first readily follows from what we have so far and is the plane at infinity whose homography matrix is $A_\infty = M'RM^{-1}$ arises from setting $d = \infty$ in eqn. 1 (note that A_∞ cannot be recovered from image measurements alone without prior information, unlike A_π). The second plane, denoted by π_s, is coplanar with the center O' (thus the corresponding homography matrix is of rank 2, and $A_{\pi_s} v = 0$) and perpendicular to the line OO', thus in addition $A_{\pi_s}{}^\top v' = 0$. It then follows [12] that $A_{\pi_s} \cong [v']F$. We will refer to the plane π_s as the "singular intrinsic plane". The singular intrinsic plane plays a critical role in reconstruction from lines [5], which will be discussed later in this section.

Projective structure of the object can be described relative to two arbitrary fiducial planes, π_1 and π_2, whose homography matrices are recovered up to a mutual scale (this altogether requires five corresponding pairs and the epipoles), and which is captured by the equation: $p' \cong A_{\pi_1} p + \kappa A_{\pi_2} p$, where κ is invariant to the choice of two views [20] (i.e., to the camera transformations). A single fiducial plane π gives rise to an invariant, referred to as "relative affine structure" in [19, 21], and is captured by the equation: $p' \cong A_\pi p + kv'$, where A_π and v' are recovered up to a common scale (which all together requires four corresponding pairs and the epipoles), and k is invariant to the choice of the second view ψ_2 (the view ψ_1 is part of the projective reference frame by having the center O as part of the five fiducial points). For more details see [19, 21], and for related results and applications see [22, 17, 9].

2.2 Trilinearity: Review

The material presented in this paper relies critically on the existence of certain trilinear forms of three views with intrinsic coefficients [18]. Let $A_\pi : \psi_1 \mapsto \psi_2$ and $B_\pi : \psi_1 \mapsto \psi_3$ be homography matrices of some arbitrary plane π, recovered together with v', v'' up to a common scale, i.e., for every corresponding triplet p, p', p'', there exists a scalar k (relative affine structure) that satisfies:

$$p' \cong A_\pi p + kv'$$
$$p'' \cong B_\pi p + kv''$$

Then, there exists a tensor of 27 intrinsic (i.e., do not depend on the choice of π) entries:

$$\alpha_{ijk} = v'_i b_{jk} - v''_j a_{ik}. \qquad i,j,k = 1,2,3 \tag{2}$$

We use the standard notations for tensors, as follows. We identify vectors and matrices by fixing some of the indices while varying others. For example, α_{ijk} is a scalar, $\alpha_{ij\cdot}$ is a set of 9 vectors (k varies while i,j remain fixed); $\alpha_{i\cdot\cdot}$ is a set of 3 matrices ($\alpha_{1\cdot\cdot}$, $\alpha_{2\cdot\cdot}$ and $\alpha_{3\cdot\cdot}$), and so forth.

The tensor α forms the set of coefficients of certain trilinear forms that vanish on any corresponding triplet p, p', p'' (i.e., functions of views that are invariant to object structure). There are nine functions divided into sets of four linearly independent functions. For example, the following four forms are linearly independent invariant functions of three views:

$$\begin{aligned}
x''\boldsymbol{\alpha}_{13\cdot}^\top p - x''x'\boldsymbol{\alpha}_{33\cdot}^\top p + x'\boldsymbol{\alpha}_{31\cdot}^\top p - \boldsymbol{\alpha}_{11\cdot}^\top p &= 0,\\
y''\boldsymbol{\alpha}_{13\cdot}^\top p - y''x'\boldsymbol{\alpha}_{33\cdot}^\top p + x'\boldsymbol{\alpha}_{32\cdot}^\top p - \boldsymbol{\alpha}_{12\cdot}^\top p &= 0,\\
x''\boldsymbol{\alpha}_{23\cdot}^\top p - x''y'\boldsymbol{\alpha}_{33\cdot}^\top p + y'\boldsymbol{\alpha}_{31\cdot}^\top p - \boldsymbol{\alpha}_{21\cdot}^\top p &= 0,\\
y''\boldsymbol{\alpha}_{23\cdot}^\top p - y''y'\boldsymbol{\alpha}_{33\cdot}^\top p + y'\boldsymbol{\alpha}_{32\cdot}^\top p - \boldsymbol{\alpha}_{22\cdot}^\top p &= 0.
\end{aligned}$$

Since every corresponding triplet p, p', p'' contributes four linearly independent equations, then seven corresponding points across the three views uniquely determine (up to scale) the tensor α. More details and applications can be found in [18].

2.3 3D recovery from lines: Hartley's Observation

The tensor α can be decomposed into sets of three matrices in three ways represented by: $\alpha_{i\cdot\cdot}$, $\alpha_{\cdot j\cdot}$ and $\alpha_{\cdot\cdot k}$. Hartley [4] has observed that the set $\alpha_{\cdot\cdot k}$ corresponds to the three matrices T_1, T_2, T_3 of [5, 27, 23] associated with lines over three views. This connection implies that lines and points can be unified into one scheme for reconstruction from 3 views. Moreover, the methods for projective reconstruction from line correspondences across three views [5] can be applied directly to reconstruction from point correspondences. Reconstruction from lines, however, relies critically on the singular intrinsic plane π_s (coplanar with O'). In particular, from the identities $F = [v']A_\pi$ for any plane π and $A_{\pi_s} = [v']F$, it can be verified by direct substitution that the homography matrix B_{π_s} of the singular intrinsic plane π_s has the form

$$B_{\pi_s} = \frac{1}{v'^\top v'}[T_1 v', T_2 v', T_3 v'].$$

In other words, given the epipole $v' \in \psi_2$ and the tensor represented by $\alpha_{\cdot\cdot k}$ we can obtain the homography matrix registrating the plane π_s (coplanar with O') across views ψ_1, ψ_3. Similarly, from the identities $F = [v']A_\pi$ and $F^\top v' = 0$, one can verify (as shown in [5]) by direct substitution that v', v'' can be found as the common perpendiculars to the left (respectively, right) null-spaces of T_k.

However, these derivations do not reveal any intrinsic structure associated with three views (beyond the tensor itself). Moreover, with the equations described above we do not obtain the fundamental matrix F (and, F'') other than by round-about means such as by switching the order of views to obtain the epipole in $v \in \psi_1$, or by recovering A_{π_s} and solving for F from the identity $A_{\pi_s} = [v']F$. From a practical standpoint, there are several constraints which when not enforced explicitly may lead to instability in the presence of noise. For example, the matrices T_k are of rank 2 each and due to the use of the singular intrinsic plane we have $A_{\pi_s} v = 0$ and $A_{\pi_s}{}^\top v' = 0$. Indeed, Hartley in [4] resorts to constrained optimization methods (explicitly enforcing $A_{\pi_s}{}^\top v' = 0$) in order to achieve stability. Moreover, the coupling between the homographies A_{π_s}, B_{π_s} and the epipoles v', v'' imply a dominant role to the epipoles on the entire reconstruction process — like in the case of reconstruction from two views — thereby weakening the advantage of using the tensor α for reconstruction.

3 Theoretical Results

The tensor α of 27 coefficients is an intrinsic structure of 3 perspective views. What we are interested to find in this section are connections between the tensor and the 3D scene from which the views came from, and between the tensor and other familiar structures associated with 2 views. The following questions are addressed in this context: (i) can the fundamental matrix F, say between views ψ_1 and ψ_2, be recovered from α ? (ii) can the epipole, say $v' \in \psi_2$, be recovered directly from the tensor (i.e., without first recovering F)? (iii) what is the geometric picture underlying the tensor? (iv) can 3D reconstruction of projective structure, relative affine structure and Euclidean structure (the latter assuming calibrated cameras) be obtained (directly) from the tensor? and (v) if two views are in principle sufficient to completely determine the 3D world, what is then the role of the third view? The answers to these questions are detailed in the subsections below.

We unravel the properties of the tensor α from the substructure $\alpha_{\cdot j \cdot}$ (and its dual $\alpha_{i\cdot\cdot}$). This route leads to remarkably simple results and new intrinsic structures associated with three views. In particular, some of the results include (i) instead of the singular intrinsic plane we show the existence of three intrinsic reference planes (non-singular) whose corresponding homographies are the matrices $\alpha_{\cdot j \cdot}$, hence there is a complete decoupling between the epipoles and the intrinsic homographies, (ii) the fundamental matrix F (F'') is directly expressed as a solution of a system of equations in the matrices $\alpha_{\cdot j \cdot}$ (respectively, $\alpha_{i\cdot\cdot}$); (iii) the epipole v' (v'') can be recovered directly from the tensor, rather than through F (F''); (iv) the rotational component of camera motion R'' (R) (in case of calibrated cameras) can be expressed directly from the matrices $\alpha_{\cdot j \cdot}$ (respectively, $\alpha_{i\cdot\cdot}$); (v) there is a clear geometric interpretation for the role of the third view in the reconstruction process, and (vi) there are no constraints in the system, i.e., the matrices $\alpha_{\cdot j \cdot}$ have full rank, thus implying better behaved solutions in the presence of noise.

The details are described in the remaining of the paper. Taken together with Hartley's [5] results on $\alpha_{\cdot\cdot k}$, and our results on $\alpha_{\cdot j \cdot}$ ($\alpha_{i\cdot\cdot}$) we obtain most of the basics on the role of the tensor $\alpha_{\cdot\cdot\cdot}$ in capturing 3D information from three views.

3.1 Intrinsic Structures from 3 Views

Let E_1, E_2, E_3 be three matrices corresponding to $\alpha_{\cdot j \cdot}$, i.e., $E_1 = \alpha_{\cdot 1 \cdot}$, and so forth. The first result is remarkable in its simplicity and penetrates deep into various intrinsic structures associated with multiple views:

Theorem 1 *The matrices E_j corresponding to $\alpha_{.j.}$ are homography matrices of three distinct, non-singular, planes:*

$$E_j = \lambda v''_j M'(R + \frac{1}{v''_j} T {u_j}^\top) M^{-1}, \tag{3}$$

where u_j are the row vectors of $U = M''R''$ and λ is some fixed scale factor.

Proof. From the definition of the tensor in eqn. 2, we write E_j in terms of the components A, B, v', v'' as follows:

$$E_j = v' {b_j}^\top - v''_j A = M'T {b_j}^\top - v''_j A,$$

where b_j are the row vectors of B, and v' is scaled such that $v' = M'T$. Let v'' be scaled such that $v'' = M''T''$. Since A, B, v', v'' are defined up to a common scale, then the matrices A and B share a common scale λ:

$$A = \lambda M'(R + \frac{1}{d} T n^\top) M^{-1}$$
$$B = \lambda M''(R'' + \frac{1}{d} T'' n^\top) M^{-1},$$

where n and d are the parameters of some arbitrary plane π. By substituting the expressions of A and B in E_j we obtain:

$$E_j = \lambda M'T(u_j + v''_j \frac{n}{d})^\top M^{-1} - \lambda v''_j M'(R + \frac{1}{d} T n^\top) M^{-1},$$

where u_j are the row vectors of $U = M''R''$. By collecting terms we see that the contribution of π drops out and we are left with eqn. 3. $\square$

We see that the division of the tensor into $\alpha_{.j.}$ gives rise directly to homography matrices of some planes in space. These planes are not singular, i.e., they are not coplanar with any of the camera centers, and in case the third camera is calibrated (i.e., $M'' = I$), the planes are perpendicular to the main axes of the third camera coordinate frame, i.e., E_1 is the homography associated with the plane whose normal is the $x-axis$ of the third camera frame and whose distance to the origin O is $1/v''_1$, and so forth. We refer to the matrices E_1, E_2, E_3 as the *intrinsic homography matrices.* The following corollaries are a direct consequence of this result:

Corollary 1 *The intrinsic homography matrices E_1, E_2, E_3 are generically full rank. In the case the three camera centers O, O', O'' are collinear (i.e., $R''R^\top T = T''$) the intrinsic homography matrices have rank 2 (the corresponding fiducial planes are coplanar with O'). Finally, E_j has rank 1 iff $v''_j = 0$.*

Proof. Eqn. 3 is of the general form depicted by eqn. 1, rewritten below:

$$A_\pi \cong M'RM^{-1} + \frac{1}{d} v' n^\top M^{-1}. \tag{4}$$

The right hand terms of eqn. 4 represent a sum of two matrices, the first of rank 3 and the second of rank 1. Their sum is a matrix of deficient rank if and only if the scalar $\mu = 1/d$ is a generalized eigenvalue of the two matrices. There are two eigenvalues: one with multiplicity 2 making the sum a rank 1 matrix (the case when $v''_j = 0$), and the second eigenvalue making the sum a rank 2 matrix (homography matrix associated with the plane coplanar with O'). Since rank 2 homography matrices necessarily satisfy the constraint $A_\pi v = 0$, it can be readily calculated that the corresponding eigenvalue must equal:

$$\mu = \frac{1}{(M''R''M^{-1})_j{}^\top v}.$$

Therefore, if $M''R''M^{-1}v = v''$, then the rank of all three intrinsic homography matrices is 2. Since, $v = MR^\top T$ and $v'' = M''T''$, the condition reduces to $R''R^\top T = T''$, which in geometric terms means that the three camera centers are collinear. □

The finding that in general the intrinsic homography matrices are of full rank is encouraging for numerical stability reasons since otherwise any system of equations involving the E_js would ignore hidden constraints (or, constrained optimization techniques would have to be employed).

Corollary 2

$$[v']E_j \cong F$$

Proof. As discussed in Section 2, $[v']A_\pi \cong F$ for any plane π, thus this corollary is a direct consequence of the finding that each E_j represents a homography of some plane. The corollary can be shown directly as well: the columns of E_j are composed of the difference between v' scaled by an element of B and the corresponding column of A scaled by v''_j. Since $[v']v' = 0$, we have that $[v']E_j = v''_j[v']A$. □

The importance of the last corollary is that it provides a method for recovering the epipole v' from the tensor, as follows. Since $p'^\top Fp = 0$, we have $p'^\top[v']E_jp = 0$. Thus, for any choice of j, given the corresponding points we already used to recover the tensor we have a linear system for v' (over-determined if we use more than two pairs of corresponding points).

Corollary 3 *For every pair $i, j = 1, 2, 3$, $i \neq j$, the matrix $[v']E_iE_j^{-1}$ is skew-symmetric.*

Proof. From Corollary 2 we have $[v']E_i \cong [v']E_j$, thus

$$[v'] \cong [v']E_iE_j^{-1}.$$

It follows that every vector $y \in \mathcal{R}^3$ satisfies

$$y^\top[v']E_iE_j^{-1}y = 0,$$

thus $[v']E_iE_j^{-1}$ must be skew-symmetric. □

The last corollary provides another method to recover v' from the tensor. In particular, the diagonal of $[v']E_iE_j^{-1}$ vanishes, thus we have three homogeneous equation for v'. We may choose j such that E_j has the smallest condition number (more stable numerically), or simply use all the pairs $i \neq j$ to obtain a redundant system for a least-squares solution for v'.

Corollary 4 *The (over-determined) system of linear equations resulting from:*

$$F^\top E_j + E_j{}^\top F = 0 \qquad j = 1, 2, 3$$

provides a unique solution for F (up to scale). The solution is unique when at least two of the E_js are used in the system, i.e., $j = j_1, j_2$, $j_1 \neq j_2$, $j_1, j_2 = 1, 2, 3$.

Proof. The relation $F^{\top}E_j + E_j{}^{\top}F = 0$ is also a direct consequence of the finding that E_j is a homography of some plane. This identity can also be shown directly by observing that $F^T E_j$ is a skew-symmetric matrix. For a given j, the identity $F^{\top}E_j + E_j{}^{\top}F = 0$ provides only six equations for F, thus is not sufficient to uniquely determine F (this is also clear geometrically — we need at least two planes to determine F). The technical issues are (i) to show that the rank of the system, for a given j, is six (not lower), and (ii) with two homographies, E_{j_1} and E_{j_2}, we obtain a set (of 12 equations) whose rank is 8. These are proven in the appendix. $\square$

The importance of the last corollary is that it provides an algorithm for recovering the fundamental matrix F, associated with the first two views, from the tensor. Moreover, F can be obtained in a least-squares sense because we have a redundant set of equations (18 homogeneous equations for 9 unknowns). In Section 4 we will use this result to obtain a method for computing F given any number of views $m \geq 3$.

Corollary 5 *Let $(E_j)_m$ denote the mth column of matrix E_j. Then the vectors $(E_1)_m, (E_2)_m, (E_3)_m$ span a 2D space for $m = 1, 2, 3$.*

Proof. Follows immediately from the observation that $F^T E_j$ is a skew-symmetric matrix. $\square$

The importance of the last corollary is by showing that the set of independent and dependent parameters are distributed very conveniently for purposes of numerical stability. The 27 coefficients of the tensor are not all independent, and they are arranged such that 9 of them are linearly dependent on the remaining 18 parameters [18]. The corollary shows that any two out of the three matrices E_j contain 18 independent parameters. This is very convenient because all the results in this paper require the use of at most two out of the three matrices. Hence, there is no need to enforce hidden constraints in the corresponding numerical algorithms. This is unlike the decomposition $\alpha_{\cdot\cdot k}$ yielding the matrices T_k associated with lines: there each of the matrices contains a mixture of independent and depended parameters (rank of T_k is 2).

Corollary 6 *The projective structure of the scene is captured by the equation*

$$p' \cong E_1 p + \kappa E_2 p \tag{5}$$

where κ is a projective invariant of $\mathcal{P}^3$, referred to as "projective-depth" in [20]. The invariant κ does not depend on the choice of the first two views (ψ_1, ψ_2). The third view ψ_3 determines the projective reference frame.

Proof. It was shown in [20] that two homographies (of two distinct planes) recovered up to a common scale provide the projective structure of the 3D scene in the form represented in eqn. 5. By Theorem 1, E_1, E_2, E_3 are homographies of distinct planes, and are determined up to a common scale (because the tensor coefficients are recovered up to a common scale). $\square$

The last corollary contains several layers of significance. First, the projective structure of the scene is recovered directly from the tensor, without the need for additional steps of calculations. Second, and intriguing, is that structure is recovered without "fiducial" points. The projective structure of the scene requires the identification of five fiducial (reference/basis) points, and in the case of two views, can be achieved (linearly) by observing just the projections of those five points and the epipoles [2, 20]. For example, the projections of points 1,2,3 and the epipoles determine the homography due the corresponding face π_1 of the tetrahedron of reference. The projections of points 2,3,4 (and the epipoles) determine the homography due to another face π_2. The remaining fifth fiducial point is then used to achieve a common scale for both homographies, and the result is the projective invariance described in eqn. 5 [20]. Geometrically, the invariant κ is the cross ratio of the fifth fiducial point P_o, the point of interest P (projecting onto p, p') and the intersections of the line connecting P_o and P with the two planes π_1, π_2. As it turns out, with three views the two planes are provided "for free" without the need to select four fiducial points (and the epipoles). The fifth fiducial point is not required as well, because the homographies are already determined up to a common scale. To complete the geometric picture we may describe the location of the fifth fiducial point P_o relative to the planes associated with E_j, as follows. The points P_o is associated with $\kappa = 1$, i.e., $p'_o \cong (E_1 + E_2)p_o$. The matrix $E_1 + E_2$ is a homography of a plane π_{12}, because in general we have (from Theorem 1):

$$E_i + E_j = \lambda(v''_i + v''_j)M'(R + \frac{(u_i + u_j)^\top}{v''_i + v''_j})M^{-1}.$$

Therefore, P_o is coplanar with π_{12}. For similar reasons P_o is coplanar with π_{13} and π_{23}, thus P_o is at the intersection of the three planes associated with the homographies $E_1 + E_2$, $E_1 + E_3$ and $E_2 + E_3$.

What we may gather from this is that the third view provides the projective reference frame, i.e., we are replacing fiducial points with a "fiducial view". Projective structure is invariant to the choice of two views (i.e., choice of external and internal camera parameters) provided reconstruction is performed with respect to a fixed set of fiducial points. Similarly, in the context of a fiducial view, projective structure is invariant to the choice of views ψ_1, ψ_2 as long as the third view ψ_3 remains fixed (as it provides the reference frame). Changing the third view will result in reconstructions that are projectively related to one another. In practical terms, reconstruction using the tensor treats all points equally likely, i.e., the tensor does not depend on the choice of seven corresponding points, which is generally a desirable property in the presence of noise.

Corollary 7 *The relative affine structure of the scene is captured by the equation*

$$p' \cong E_1 p + kv'$$

where the scalar k is invariant, up to a common scale, to the choice of the second view ψ_2. The third view ψ_3 provides the reference plane (associated with the homography E_1). The common scale can be set by selecting an arbitrary corresponding pair p_o, p'_o and setting the scale of v' such that $p'_o \cong E_1 p_o + v'$. The projective coordinates of the scene are then $[x, y, 1, k]$.

The relative affine structure (see Section 2) requires only one fiducial point: the fixed point P_o for setting a common scale between the homography of the reference plane and the epipole v' (recall that v' is recovered from the tensor up to scale). Beside a more intuitive geometric description of the invariant k, compared to the projective invariant κ, the advantages lie with the projective invariant for two reasons: (i) nothing but the tensor is needed for the projective invariant, whereas for the relative affine the epipole v' is required in addition, and (ii) the projective invariant requires no fiducial points. This advantage is in contrast with reconstruction from two views where the epipole v' is required in both cases, and the relative affine invariant has the advantage because it relies on four fiducial points, rather than five.

3.2 The Dual Representation $\alpha_{i\cdot\cdot}$

Before we proceed to further results, we mention that the representation $\alpha_{i\cdot\cdot}$ is the dual of the representation $\alpha_{\cdot j\cdot}$ that we have been working with so far. In other words, if we denote by W_1, W_2, W_3 the three matrices corresponding to $\alpha_{i\cdot\cdot}$, then W_i are homographies from ψ_1 to ψ_3 corresponding to planes associated with the coordinate frame of the second view:

$$W_i = \lambda v'_i M''(R'' + \frac{1}{v'_i} T'' {g_j}^\top) M^{-1},$$

where $G = M'R$. The rest follow readily, and for example: $[v'']W_i \cong F''$ (recall that F'' is the fundamental matrix between views ψ_1 and ψ_3, and $v'' \in \psi_3$ is the epipole with the first view ψ_1); ${F''}^\top W_i + {W_i}^\top F'' = 0$. The second view ψ_2 is the fiducial view for reconstruction, rather than the third view.

3.3 Further Results and Euclidean Reconstruction

So far we have obtained results on recovering the epipoles and fundamental matrices from the tensor, and results on using the tensor for reconstruction of projective structure and relative affine structure. In this section we focus on recovering rigid motion parameters (assuming internal calibration) from the tensor, and as a byproduct obtain further results, including another method for recovering epipoles from the tensor.

Corollary 8

$$E_i - \frac{v''_i}{v''_j} E_j = \lambda v' (u_i - \frac{v''_i}{v''_j} u_j)^\top M^{-1}$$

Proof. Follows by substitution from Theorem 1. ∎

Corollary 9 *The matrix $E_j^{-1}E_i$ has an eigenvalue of multiplicity 2 and whose equal to $\frac{v''_i}{v''_j}$.*

Proof. Derives from Corollary 8 in the following way. Since $E_i - \frac{v''_i}{v''_j}E_j$ is a rank 1 matrix, then $\mu = v''_i / v''_j$ is one of the generalized eigenvalues of $(E_i - \mu E_j)x = 0$, where x is the corresponding generalized eigenvector. Also, since we know there exists μ that makes $E_i - \mu E_j$ rank 1, then μ must be with multiplicity 2. Finally, the generalized eigenvalues of $(E_i - \mu E_j)x = 0$ are the eigenvalues of $(E_j^{-1}E_i - \mu I)x = 0$, i.e., the eigenvalues of $E_j^{-1}E_i$. ∎

The last corollary provides another (third) method to recover epipoles from the tensor. The components v''_1/v''_3 and v''_2/v''_3 are simply the eigenvalue with multiplicity 2 of the matrices $E_3^{-1}E_1$ and $E_3^{-1}E_2$, respectively. Similarly, the eigenvalues with multiplicity 2 of $W_j^{-1}W_i$ provide the epipole v'.

In Corollary 8, the left-hand side of the equation is known (the tensor, and the epipole v'' recovered form the tensor); the right-hand side is a rank 1 matrix whose columns are multiples of v' who is also known (recovered from the tensor), and the scale factors of those columns depend on $\lambda M'' R'' M^{-1}$. The question, therefore, is whether by assuming $M = M'' = I$ (camera is internally calibrated) can we recover λ and R''? In other words, given the translational component of camera motion, recoverable from the tensor of three views, can we obtain a closed-form solution for the rotational component of camera motion from the tensor of three views?

Theorem 2 *Given the epipoles v', v'', then in the case the cameras are internally calibrated (M=M'=M"=I) the rotational component of camera motion R'' can be recovered (up to reflection) in closed-form from $\alpha_{\cdot j \cdot}$, and similarly R can be recovered from $\alpha_{i \cdot \cdot}$.*

Proof. Let h_1, h_2, h_3 be the vectors:

$$h_1 = \lambda u_1 - \frac{v_1''}{v_3''}\lambda u_3$$

$$h_2 = \lambda u_2 - \frac{v_2''}{v_3''}\lambda u_3 \tag{6}$$

$$h_3 = \lambda u_3 - \frac{v_1''}{v_2''}\lambda u_2 \tag{7}$$

Given v, v'', the vectors h_i are observable following Corollary 8, and u_i are the row vectors of R''. Our task is to recover λ and u_1, u_2, u_3 from h_1, h_2, h_3.

We note that since R'' is orthonormal, then $\lambda^2 = h_1^\top h_3$. Let $u_1 = u_2 \times u_3$, $u_2 = u_3 \times u_1$ and $u_3 = u_1 \times u_2$, then

$$h_1 \times h_3 = -\lambda\left(\frac{v_1''}{v_2''}\lambda u_3 + \frac{v_1''}{v_3''}\lambda u_2 + \frac{v_1''^2}{v_2'' v_3''}\lambda u_1\right).$$

By replacing λu_1, and λu_2 with their expression from eqn. 6 and 7, respectively, we obtain:

$$u_3 = \frac{-1}{h_1^\top h_3\left(\frac{v_1''}{v_2''} + \frac{v_1''^3}{v_3''^2 v_2''} + \frac{v_1'' v_2''}{v_3''^2}\right)}\left(h_1 \times h_3 + \frac{\lambda v_1''^2}{v_2'' v_3''} h_1 + \frac{\lambda v_1''}{v_3''} h_2\right).$$

Thus, we have two solutions for u_3 (because of two solutions for λ), and from there two solutions for R''. $\square$

4 New Algorithms for 3D Calculations from 2D Images

In this section we summarize the algorithmic aspect of the results presented in the previous section. We have seen (Corollary 6) that the tensor matrices $\alpha_{\cdot j \cdot}$ directly give rise to the projective structure of the scene. Thus, in principle we do not need to recover any 2-view structure such as the epipoles or the fundamental matrix. Nevertheless, the corollaries also provide a variety of means for recovering 2-view structures, which we will expand in this section.

Consider the problem of recovering the fundamental matrix F. The problem is known to suffer from numerical instability, and the best known algorithms (cf. [1]) are based on non-linear constrained optimization using many point matches along two views. Using the tensor, we can extend the envelope of algorithms for computing F by working not only with many points matches, but also with many views (three or more); this can be done as follows. Consider views $\psi_1, ..., \psi_m$, $m \geq 3$ with many (≥ 7) points matches across them. Assume we would like to compute F between views ψ_1, ψ_m. Let α^{1mj} denote the tensor of views ψ_1, ψ_m, ψ_j (in that order), where $j = 2, ..., m-1$. Each of these tensors can be recovered in a least-squares manner using all the available point matches across the appropriate three views. From Corollary 4 we have that each tensor contributes 18 linear equations for F. Hence, the more views we have the larger system of equations we have for F for a least-squares solution. Non-linear components may be added to this basic scheme by enforcing the rank constraint of F, and other constraints like in [1].

For the epipoles we have found a number of algorithms. The straightforward ones are either to recover F, and then solve for v' from the relation $F^\top v' = 0$ [2], or directly from one of the E matrices by using point matches (Corollary 2). The less obvious methods, that also allow the use of more than 3 views are derived from Corollaries 3 and 9, as follows.

Consider Corollary 9 that states that the eigenvalues of $E_j^{-1}E_i$ come with an eigenvalue of multiplicity 2 and which is equal to $\frac{v_i''}{v_j''}$. Assume we want to recover $(\frac{v_1''}{v_3''}, \frac{v_2''}{v_3''})$ between views ψ_1, ψ_m. We compute the tensors α^{1jm}, where $j = 2, ..., m-1$. We stack the matrix $E_3^{-1}E_1$ for each tensor and get a $(m-2) \times 3$ matrix whose singular values must have one with multiplicity 2 and whose square-root is equal to $\frac{v_1''}{v_3''}$, thus SVD can be applied onto multiple views (via the multiple tensors) to solve for epipoles (up to two solutions because of sign-flip).

Consider next Corollary 3 that states that $[v']E_iE_j^{-1}$ is skew-symmetric — thus providing three homogeneous equations for v' (because the diagonal of $E_iE_j^{-1}$ vanishes). Each tensor α^{1mj}, $j = 2, ..., m-1$, provides three equations for v', thus m views provide a redundant set of $3m-6$ linear equations for v'.

Similar logic applies to over-determining projective structure (Corollary 6) and relative affine structure (Corollary 7) and Euclidean structure (Theorem 2). The main point is that the basic ingredient is 3-view-geometry rather than 2-view geometry, and on top of that various different methods for using redundant equations can be applied — not necessarily least-squares methods.

5 Discussion and Future Research

The paper derives connections between the trilinear function invariants across three views and intrinsic structures and invariants of 3D space. The results show that the tensor of coefficients determined by three views replaces entirely the role of the fundamental matrix (and associated intrinsic structures of two views) in 3D tasks. In other words, the projective structure of the scene follows directly from the tensor (Corollary 6) without the need to recover any intrinsic structure associated with two views. In addition, the tensor encompasses 2-view structures in the sense that the fundamental matrix is readily expressed as a solution of a linear system determined by the tensor, the rotational component of camera motion is expressible in closed form by the tensor, and a variety of means exist for recovering the epipoles from the tensor.

The major result, from which all the rest follow readily as corollaries, is the existence of three intrinsic homography matrices (Theorem 1). In other words, there exists a decomposition of the tensor into three matrices that correspond to three intrinsic homography matrices of three distinct planes. The planes are associated with the camera coordinate frame of the third view and provide a reference basis for reconstruction of invariants. This provides a geometric intrinsic structure of three views; a complete decoupling between homographies matrices and epipoles (unlike 2-view geometry, and 3-view line geometry); and provides a very convenient decoupling of independent and depended parameters (Corollary 5), again unlike the situation in 2-view geometry (rank of F is 2), and 3-view line geometry (rank of T_k is 2).

The material presented in this paper contains a host of practical consequences — of which we mentioned only a few. For example, the envelope of techniques for recovering the fundamental matrix is now extended to multiple views (Section 4); projective reconstruction can proceed without fiducial points; and most importantly, the introduction of the third view has a potentially stabilizing effect on numerical algorithms for 3D calculations from 2D images (this has been experimentally verified, on real and synthetic examples, for purposes of image synthesis in [18], with results indicating an order of magnitude improvement compared to any other known technique).

Continuing on the practical motivation, three views admit three distinct fundamental matrices (pairs of views out of three), thus all together we have 27 homogeneous equations (from 9 points), or 24 inhomogeneous equations (from 8 points). However, these equations are decomposed into three independent sets, thus in practice we use only two views at a time. The trilinear coefficients comprising the tensor are determined also by 27 equations but with two important differences: (i) 7 points are needed, (ii) the set of equations is not decomposable linearly, thus all three views are used together. In other words, with the introduction of the tensor there is a better use of the information available in three views. This argument implies a clear superiority, regardless of statistical methods that may be used on top of the results presented here, of the techniques presented (especially those that follow directly from the tensor, such as projective reconstruction) over traditional 2-view techniques.

Regarding future research, it is safe to conjecture that the general structure of multiple views is contained in a multilinear tensor which follows from multilinear invariant functions of multiple views. We believe that a proper handle of these issues would comprise one of the major corner-stones of invariant research in Computer Vision, and would have great practical consequences on the design of algorithms for 3D calculations from 2D images.

Acknowledgments

We thank Shmuel Peleg for comments throughout this research and for contribution on simplifying the notations used in this paper. We also thank Nassir Navab for comments during early stages of this research.

References

1. R. Deriche, Z. Zhang, Q.T. Luong, and O.D. Faugeras. Robust recovery of the epipolar geometry for an uncalibrated stereo rig. In *Proceedings of the European Conference on Computer Vision*, pages 567–576, Stockholm, Sweden, May 1994. Springer-Verlag, LNCS 800.
2. O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proceedings of the European Conference on Computer Vision*, pages 563–578, Santa Margherita Ligure, Italy, June 1992.
3. O.D. Faugeras and F. Lustman. Let us suppose that the world is piecewise planar. In O. D. Faugeras and Georges Giralt, editors, *International Symposium on Robotics Research*, pages 33–40. MIT Press, Cambridge, MA, 1986.
4. R. Hartley. Lines and points in three views — a unified approach. In *Proceedings Image Understanding Workshop*, Monterey, CA, November 1994.
5. R. Hartley. Projective reconstruction from line correspondences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WI, June 1994. IEEE Computer Society Press.
6. R. Hartley and R. Gupta. Computing matched-epipolar projections. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 549–555, New York, NY, 1993.

7. R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 761–764, Champaign, IL., June 1992.
8. T.S. Huang and C.H. Lee. Motion and structure from orthographic projections. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-11:536–540, 1989.
9. M. Irani, B. Rousso, and S. Peleg. Recovery of ego-motion using image stabilization'. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 454–460, Seattle, Washington, June 1994.
10. J.J. Koenderink and A.J. Van Doorn. Affine structure from motion. *Journal of the Optical Society of America*, 8:377–385, 1991.
11. H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.
12. 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.
13. N. Navab, O. D. Faugeras, and T. Vieville. The critical sets of lines for camera displacement estimation: A mixed euclidean-projective and constructive approach. In *Proc. 4th International Conference on Computer Vision*, Berlin, Germany, May 1993. IEEE.
14. Q.T.Luong and O.D. Faugeras. Determining the fundamental matrix with planes: Instability and new algorithms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 489–494, New York, NY, June 1993.
15. L. Quan. Invariants of 6 points from 3 uncalibrated images. In *Proceedings of the European Conference on Computer Vision*, Stockholm, Sweden, May 1994. Springer-Verlag, LNCS 801.
16. L. Robert and O.D. Faugeras. Relative 3D positioning and 3D convex hull computation from a weakly calibrated stereo pair. In *Proceedings of the International Conference on Computer Vision*, pages 540–544, Berlin, Germany, May 1993.
17. H.S. Sawhney. 3D geometry from planar parallax. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 929–934, Seattle, Washington, June 1994.
18. A. Shashua. Algebraic functions for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1994. in press.
19. A. Shashua. On geometric and algebraic aspects of 3D affine and projective structures from perspective 2D views. In J.L. Mundy, A. Zisserman, and D. Forsyth, editors, *Applications of Invariance in Computer Vision.* Springer-Verlag, LNCS 825, 1994. Proc. of the 2nd European Workshop on Invariants, Ponta Delagada, Azores, Oct. 1993.
20. A. Shashua. Projective structure from uncalibrated images: structure from motion and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):778–790, 1994.
21. A. Shashua and N. Navab. Relative affine structure: Theory and application to 3D reconstruction from perspective views. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 483–489, Seattle, Washington, 1994.
22. G. Sparr. A common framework for kinetic depth, reconstruction and motion for deformable objects. In *Proceedings of the European Conference on Computer Vision*, pages 471–482, Stockholm, Sweden, May 1994.
23. M.E. Spetsakis and J. Aloimonos. Structure from motion using line correspondences. *International Journal of Computer Vision*, 4(3):171–183, 1990.
24. C. Tomasi. Pictures and trails: A new framework for the computation of shape and motion from perspective image sequences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 913–918, Seattle, Washington, June 1994. IEEE Computer Society Press.
25. R. Tsai and T.S. Huang. Estimating three-dimensional motion parameters of a rigid planar patch, II: singular value decomposition. *IEEE Trans. on Acoustic, Speech and Signal Processing*, 30, 1982.

26. S. Ullman and R. Basri. Recognition by linear combination of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-13:992—1006, 1991. Also in M.I.T AI Memo 1052, 1989.
27. J. Weng, T.S. Huang, and N. Ahuja. Motion anf structure from line correspondences: Closed form solution, uniqueness and optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(3), 1992.