Algebraic Functions For Recognition

Amnon Shashua

Abstract— **In the general case, a trilinear relationship between three perspective views is shown to exist. The *trilinearity* result is shown to be of much practical use in visual recognition by alignment — yielding a direct reprojection method that cuts through the computations of camera transformation, scene structure and epipolar geometry. Moreover, the direct method is linear and sets a new lower theoretical bound on the minimal number of points that are required for a linear solution for the task of reprojection. The proof of the central result may be of further interest as it demonstrates certain regularities across homographies of the plane and introduces new view invariants. Experiments on simulated and real image data were conducted, including a comparative analysis with epipolar intersection and the linear combination methods, with results indicating a greater degree of robustness in practice and a higher level of performance in re-projection tasks.**

Keywords— **Visual Recognition, Alignment, Re-projection, Projective Geometry, Algebraic and Geometric Invariants.**

I. Introduction

We establish a general result about algebraic connections across three perspective views of a 3D scene and demonstrate its application to visual recognition via alignment. We show that, in general, any three perspective views of a scene satisfy a pair of trilinear functions of image coordinates. In the limiting case, when all three views are orthographic, these functions become linear and reduce to the form discovered by [38]. Using the trilinear result one can manipulate views of an object (such as generate novel views from two model views) without recovering scene structure (metric or non-metric), camera transformation, or even the epipolar geometry. Moreover, the trilinear functions can be recovered by linear methods with a minimal configuration of seven points. The latter is shown to be new lower bound on the minimal configuration that is required for a general linear solution to the problem of re-projecting a 3D scene onto an arbitrary novel view given corresponding points across two reference views. Previous solutions rely on recovering the epipolar geometry which, in turn, requires a minimal configuration of eight points for a linear solution.

The central results in this paper are contained in Theorems 1, 2 and 3. The first theorem states that the variety of views ψ of a fixed 3D object obtained by an uncalibrated pin-hole camera satisfy a relation of the sort $F(\psi, \psi_1, \psi_2) = 0$, where ψ_1, ψ_2 are two arbitrary views of the object, and F has a special trilinear form. The coefficients of F can be recovered linearly without establishing first the epipolar geometry, 3D structure of the object, or camera motion. The auxiliary Lemmas required for the proof of Theorem 1 may be of interest on their own as they establish certain regularities across projective transformations of the plane and introduce new view invariants (Lemma 4).

Theorem 2 addresses the problem of recovering the coefficients of the trilinear functions in the most economical way. It is shown that among all possible trilinear functions across three views, there exists at most four linearly independent such functions. As a consequence, the coefficients of these functions can be recovered linearly from seven corresponding points across three views.

Theorem 3 is an obvious corollary of Theorem 1 but contains a significant practical aspect. It is shown that if the views ψ_1, ψ_2 are obtained by parallel projection, then F reduces to a special bilinear form — or, equivalently, that any perspective view ψ can be obtained by a rational linear function of two orthographic views. The reduction to a bilinear form implies that simpler recognition schemes are possible if the two reference views (model views) stored in memory are orthographic.

These results may have several applications (discussed in Section VI), but the one emphasized throughout this paper is for the task of recognition of 3D objects via alignment. The alignment approach for recognition ([37, 16], and references therein) is based on the result that the equivalence class of views of an object (ignoring self occlusions) undergoing 3D rigid, affine or projective transformations can be captured by storing a 3D model of the object, or simply by storing at least two arbitrary "model" views of the object — assuming that the correspondence problem between the model views can somehow be solved (cf. [27, 5, 33]). During recognition a small number of corresponding points between the novel input view and the model views of a particular candidate object are sufficient to "re-project" the model onto the novel viewing position. Recognition is achieved if the re-projected image is successfully matched against the input image. We refer to the problem of predicting a novel view from a set of model views using a limited number of corresponding points, as the problem of *re-projection*.

The problem of re-projection can in principal be dealt with via 3D reconstruction of shape and camera motion. This includes classical structure from motion methods for recovering rigid camera motion parameters and metric shape [36, 18, 35, 14, 15], and more recent methods for recovering non-metric structure, i.e., assuming

A. Shashua is with the Artificial Intelligence Laboratory and the Center for Biological Computational Learning, Massachusetts Institute of Technology, Cambridge, MA 02139.

the objects undergo 3D affine or projective transformations, or equivalently, that the cameras are uncalibrated [17, 25, 39, 10, 13, 30]. The classic approaches for perspective views are known to be unstable under errors in image measurements, narrow field of view, and internal camera calibration [3, 9, 12], and therefore, are unlikely to be of practical use for purposes of re-projection. The non-metric approaches, as a general concept, have not been fully tested on real images, but the methods proposed so far rely on recovering first the epipolar geometry — a process that is also known to be unstable in the presence of noise.

It is also known that the epipolar geometry alone is sufficient to achieve re-projection by means of intersecting epipolar lines [24, 6, 8, 26, 23, 11] using at least eight corresponding points across the three views. This, however, is possible only if the centers of the three cameras are non-collinear — which can lead to numerical instability unless the centers are far from collinear — and any object point on the tri-focal plane cannot be re-projected as well. Furthermore, as with the non-metric reconstruction methods, obtaining the epipolar geometry is at best a sensitive process even when dozens of corresponding points are used and with the state of the art methods (see Section V for more details and comparative analysis with simulated and real images).

For purposes of stability, therefore, it is worthwhile exploring more direct tools for achieving re-projection. For instance, instead of reconstruction of shape and invariants we would like to establish a direct connection between views expressed as a functions of image coordinates alone — which we call "algebraic functions of views". Such a result was established in the orthographic case by [38]. There it was shown that any three orthographic views of an object satisfy a linear function of the corresponding image coordinates — this we will show here is simply a limiting case of larger set of algebraic functions, that in general have a trilinear form. With these functions one can manipulate views of an object, such as create new views, without the need to recover shape or camera geometry as an intermediate step — all what is needed is to appropriately combine the image coordinates of two reference views. Also, with these functions, the epipolar geometries are intertwined, leading not only to absence of singularities, and a lower bound on the minimal configuration of points, but as we shall see in the experimental section to more accurate performance in the presence of errors in image measurements. Part of this work (Theorem 1 only) was presented in concise form in [31].

II. 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 $\mathcal{P}^2$. Let $\Phi \subset \mathcal{P}^3$ be a set of points standing for a 3D object, and let $\psi_i \subset \mathcal{P}^2$ denote views (arbitrary), indexed by i, of Φ. Given two cameras with centers located at $O, O' \in \mathcal{P}^3$, respectively, the epipoles are defined to be at the intersection of the line $\overline{OO'}$ with both image planes. Because the image plane is finite, we can assign, without loss of generality, the value 1 as the third homogeneous coordinate to every *observed* image point. That is, if (x, y) are the observed image coordinates of some point (with respect to some arbitrary origin — say the geometric center of the image), then $p = (x, y, 1)$ denotes the homogeneous coordinates of the image plane. Note that this convention ignores special views in which a point in Φ is at infinity in those views — these singular cases are not modeled here.

Since we will be working with at most three views at a time, we denote the relevant epipoles as follows: let $v \in \psi_1$ and $v' \in \psi_2$ be the corresponding epipoles between views ψ_1, ψ_2, and let $\bar{v} \in \psi_1$ and $v'' \in \psi_3$ the corresponding epipoles between views ψ_1, ψ_3. Likewise, corresponding image points across three views will be denoted by $p = (x, y, 1), p' = (x', y', 1)$ and $p'' = (x'', y'', 1)$. The term "image coordinates" will denote the non-homogeneous coordinate representation of $\mathcal{P}^2$, e.g., $(x, y), (x', y'), (x'', y'')$ for the three corresponding points.

Planes will be denoted by π_i, indexed by i, and just π if only one plane is discussed. All planes are assumed to be arbitrary and distinct from one another. 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.

III. The Trilinear Form

The central results of this paper are presented in the following two theorems. The remaining of the section is devoted to the proof of this result and its implications.

Theorem 1 (Trilinearity) *Let ψ_1, ψ_2, ψ_3 be three arbitrary perspective views of some object, modeled by a set of points in 3D. The image coordinates $(x, y) \in \psi_1$, $(x', y') \in \psi_2$ and $(x'', y'') \in \psi_3$ of three corresponding points across three views satisfy a pair of trilinear equations of the following form:*

$$x''(\alpha_1 x + \alpha_2 y + \alpha_3) + x''x'(\alpha_4 x + \alpha_5 y + \alpha_6) + x'(\alpha_7 x + \alpha_8 y + \alpha_9) + \alpha_{10} x + \alpha_{11} y + \alpha_{12} = 0,$$

and

$$y''(\beta_1 x + \beta_2 y + \beta_3) + y''x'(\beta_4 x + \beta_5 y + \beta_6) + x'(\beta_7 x + \beta_8 y + \beta_9) + \beta_{10} x + \beta_{11} y + \beta_{12} = 0,$$

where the coefficients α_j, β_j, $j = 1, ..., 12$, are fixed for all points, are uniquely defined up to an overall scale, and $\alpha_j = \beta_j$, $j = 1, ..., 6$.

The following auxiliary propositions are used as part of the proof.

Lemma 1 (Auxiliary - Existence) *Let $A \in PGL_3$ be the projective mapping (homography) $\psi_1 \mapsto \psi_2$ due to some plane π. Let A be scaled to satisfy $p'_o \cong A p_o + \boldsymbol{v}'$, where $p_o \in \psi_1$ and $p'_o \in \psi_2$ are corresponding points coming from an arbitrary point $P_o \notin \pi$. Then, for any corresponding pair $p \in \psi_1$ and $p' \in \psi_2$ coming from an arbitrary point $P \in \mathcal{P}^3$, we have*

$$p' \cong Ap + k\boldsymbol{v}'.$$

The coefficient k is independent of ψ_2, i.e., is invariant to the choice of the second view.

The lemma, its proof and its theoretical and practical implications are discussed in detail in [28, 32]. Note that the particular case where the homography A is affine, and the epipole v' is on the line at infinity, corresponds to the construction of affine structure from two orthographic views [17]. In a nutshell, a representation $\mathcal{R}_0$ of $\mathcal{P}^3$ (tetrad of coordinates) can always be chosen such that an arbitrary plane π is the plane at infinity. Then, a general uncalibrated camera motion generates representations $\mathcal{R}$ which can be shown to be related to $\mathcal{R}_0$ by an element of the affine group. Thus, the scalar k is an affine invariant within a projective framework, and is called a *relative affine invariant*. A ratio of two such invariants, each corresponding to a different reference plane, is a projective invariant [32]. For our purposes, there is no need to discuss the methods for recovering k — all we need is to use the existence of a relative affine invariant k associated with some arbitrary reference plane π which, in turn, gives rise to a homography A.

Definition 1 *Homographies $A_i \in PGL_3$ from $\psi_1 \mapsto \psi_i$ due to the same plane π, are said to be scale-compatible if they are scaled to satisfy Lemma 1, i.e., for any point $P \in \Phi$ projecting onto $p \in \psi_1$ and $p^i \in \psi_i$, there exists a scalar k that satisfies*

$$p^i \cong A_i p + k\boldsymbol{v}^i,$$

for any view ψ_i, where $v^i \in \psi_i$ is the epipole with ψ_1 (scaled arbitrarily).

Lemma 2 (Auxiliary — Uniqueness) *Let $A, A' \in PGL_3$ be two homographies of $\psi_1 \mapsto \psi_2$ due to planes π_1, π_2, respectively. Then, there exists a scalar s, that satisfies the equation:*

$$A - sA' = [\alpha\boldsymbol{v}', \beta\boldsymbol{v}', \gamma\boldsymbol{v}'],$$

for some coefficients α, β, γ.

Proof: Let $q \in \psi_1$ be any point in the first view. There exists a scalar s_q that satisfies $v' \cong Aq - s_q A'q$. Let $H = A - s_q A'$, and we have $Hq \cong v'$. But, as shown in [29], $Av \cong v'$ for any homography $\psi_1 \mapsto \psi_2$ due to any plane. Therefore, $Hv \cong v'$ as well. The mapping of two distinct points q, v onto the same point v' could happen only if H is the homography due to the meridian plane (coplanar with the projection center O), thus $Hp \cong v'$ for all $p \in \psi_1$, and s_q is a fixed scalar s. The latter, in turn, implies that H is a matrix whose columns are multiples of v'. □

Lemma 3 (Auxiliary for Lemma 4)
Let $A, A' \in PGL_3$ be homographies from $\psi_1 \mapsto \psi_2$ due to distinct planes π_1, π_2, respectively, and $B, B' \in PGL_3$ be homographies from $\psi_1 \mapsto \psi_3$ due to π_1, π_2, respectively. Then, $A' = AT$ for some $T \in PGL_3$, and $B = BCTC^{-1}$, where $Cv \cong \bar{v}$.

Proof: Let $A = A_2^{-1}A_1$, where A_1, A_2 are homographies from ψ_1, ψ_2 onto π_1, respectively. Similarly $B = B_2^{-1}B_1$, where B_1, B_2 are homographies from ψ_1, ψ_3 onto π_1, respectively. Let $A_1\bar{v} = (c_1, c_2, c_3)^T$, and let $C \cong A_1^{-1} diag(c_1, c_2, c_3) A_1$. Then, $B_1 \cong A_1 C^{-1}$, and thus, we have $B \cong B_2^{-1} A_1 C^{-1}$. Note that the only difference between A_1 and B_1 is due to the different location of the epipoles $v, \bar{v}$, which is compensated by C ($Cv \cong \bar{v}$). Let $E_1 \in PGL_3$ be the homography from ψ_1 to π_2, and $E_2 \in PGL_3$ the homography from π_2 to π_1. Then with proper scaling of E_1 and E_2 we have

$$A' = A_2^{-1} E_2 E_1 = A A_1^{-1} E_2 E_1 = AT,$$

and with proper scaling of C we have,

$$B' = B_2^{-1} E_2 E_1 C^{-1} = BCA_1^{-1} E_2 E_1 C^{-1} = BCTC^{-1}.$$

□

Lemma 4 (Auxiliary — Uniqueness)
For scale-compatible homographies, the scalars s, α, β, γ of Lemma 2 are invariants indexed by ψ_1, π_1, π_2. That is, given an arbitrary third view ψ_3, let B, B' be the homographies from $\psi_1 \mapsto \psi_3$ due to π_1, π_2, respectively. Let B be scale-compatible with A, and B' be scale-compatible with A'. Then,

$$B - sB' = [\alpha\boldsymbol{v}'', \beta\boldsymbol{v}'', \gamma\boldsymbol{v}''].$$

Proof: We show first that s is invariant, i.e., that $B - sB'$ is a matrix whose columns are multiples of v''. From Lemma 2, and Lemma 3 there exists a matrix H, whose columns are multiples of v', a matrix T that satisfies $A' = AT$, and a scalar s such that $I - sT = A^{-1}H$. After multiplying both sides by BC, and then pre-multiplying by C^{-1} we obtain

$$B - sBCTC^{-1} = BCA^{-1}HC^{-1}.$$

From Lemma 3, we have $B' = BCTC^{-1}$. The matrix $A^{-1}H$ has columns which are multiples of v (because $A^{-1}v' \cong v$), $CA^{-1}H$ is a matrix whose columns are multiple of $\bar{v}$, and $BCA^{-1}H$ is a matrix whose columns are multiples of v''. Pre-multiplying $BCA^{-1}H$ by C^{-1} does not change its form because every column of $BCA^{-1}HC^{-1}$ is simply a linear combination of the columns of $BCA^{-1}H$. As a result, $B - sB'$ is a matrix whose columns are multiples of v''.

Let $H = A - sA'$ and $\hat{H} = B - sB'$. Since the homographies are scale compatible, we have from Lemma 1 the existence of invariants k, k' associated with an arbitrary $p \in \psi_1$, where k is due to π_1, and k' is due to π_2: $p' \cong Ap + k\boldsymbol{v}' \cong A'p + k'\boldsymbol{v}'$ and $p'' \cong Bp + k\boldsymbol{v}'' \cong B'p + k'\boldsymbol{v}''$. Then from Lemma 2 we have $Hp = (sk' - k)\boldsymbol{v}'$ and $\hat{H}p = (sk' - k)\boldsymbol{v}''$. Since p is arbitrary, this could happen only if the coefficients of the multiples of v' in H and the coefficients of the multiples of v'' in $\hat{H}$, coincide. □

Proof of Theorem: Lemma 1 provides the existence part of theorem, as follows. Since Lemma 1 holds for any plane, choose a plane π_1 and let A, B be the scale-compatible homographies $\psi_1 \mapsto \psi_2$ and $\psi_1 \mapsto \psi_3$, respectively. Then, for every point $p \in \psi_1$, with corresponding points $p' \in \psi_2, p'' \in \psi_3$, there exists a scalar k that satisfies: $p' \cong$

$Ap + k\boldsymbol{v}'$, and $p'' \cong Bp + k\boldsymbol{v}''$. We can isolate k from both equations and obtain:

$$k = \frac{v_1' - x'v_3'}{(x'\boldsymbol{a}_3 - \boldsymbol{a}_1)^T p} = \frac{v_2' - y'v_3'}{(y'\boldsymbol{a}_3 - \boldsymbol{a}_2)^T p} = \frac{y'v_1' - x'v_2}{(x'\boldsymbol{a}_2 - y'\boldsymbol{a}_1)^T p}, \qquad (1)$$

$$k = \frac{v_1'' - x''v_3''}{(x''\boldsymbol{b}_3 - \boldsymbol{b}_1)^T p} = \frac{v_2'' - y''v_3''}{(y''\boldsymbol{b}_3 - \boldsymbol{b}_2)^T p} = \frac{y''v_1'' - x''v_2''}{(x''\boldsymbol{b}_2 - y''\boldsymbol{b}_1)^T p}, \qquad (2)$$

where $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3$ and $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ are the row vectors of A and B and $v' = (v_1', v_2', v_3')$, $v'' = (v_1'', v_2'', v_3'')$. Because of the invariance of k we can equate terms of (1) with terms of (2) and obtain trilinear functions of image coordinates across three views. For example, by equating the first two terms in each of the equations, we obtain:

$$x''(v_1'\boldsymbol{b}_3 - v_3''\boldsymbol{a}_1)^T p - x''x'(v_3'\boldsymbol{b}_3 - v_3''\boldsymbol{a}_3)^T p + x'(v_3'\boldsymbol{b}_1 - v_1''\boldsymbol{a}_3)^T p - (v_1'\boldsymbol{b}_1 - v_1''\boldsymbol{a}_1)^T p = 0, \qquad (3)$$

In a similar fashion, after equating the first term of (1) with the second term of (2), we obtain:

$$y''(v_1'\boldsymbol{b}_3 - v_3''\boldsymbol{a}_1)^T p - y''x'(v_3'\boldsymbol{b}_3 - v_3''\boldsymbol{a}_3)^T p + x'(v_3'\boldsymbol{b}_2 - v_2''\boldsymbol{a}_3)^T p - (v_1'\boldsymbol{b}_2 - v_2''\boldsymbol{a}_1)^T p = 0. \qquad (4)$$

Both equations are of the desired form, with the first six coefficients identical across both equations.

The question of uniqueness arises because Lemma 1 holds for any plane. If we choose a different plane, say π_2, with homographies A', B', then we must show that the new homographies give rise to the same coefficients (up to an overall scale). The parenthesized terms in (3) and (4) have the general form: $v_i'\boldsymbol{b}_j - v_j''\boldsymbol{a}_i$, for some i and j. Thus, we need to show that there exists a scalar s that satisfies

$$v_j''(\boldsymbol{a}_i - s\boldsymbol{a}_i') = v_i'(\boldsymbol{b}_j - s\boldsymbol{b}_j').$$

This, however, follows directly from Lemmas 2 and 4. □

The direct implication of the theorem is that one can generate a novel view (ψ_3) by simply combining two model views (ψ_1, ψ_2). The coefficients α_j and β_j of the combination can be recovered together as a solution of a linear system of 17 equations $(24 - 6 - 1)$ given nine corresponding points across the three views (more than nine points can be used for a least-squares solution).

In the next theorem we obtain the lower bound on the number of points required for solving for the coefficients of the trilinear functions. The existence part of the proof of Theorem 1 indicates that there exists nine trilinear functions of that type, with coefficients having the general form $v_i'\boldsymbol{b}_j - v_j''\boldsymbol{a}_i$. Thus, we have at most 27 distinct coefficients (up to a uniform scale), and thus, if more than two of the nine trilinear functions are linearly independent, we may solve for the coefficients using less than nine points. The next theorem shows that at most four of the trilinear functions are linearly independent and consequently seven points are sufficient to solve for the coefficients.

Theorem 2 *There exists nine distinct trilinear forms of the type described in Theorem 1, of which at most four are linearly independent. The coefficients of the four trilinear forms can be recovered linearly with seven corresponding points across the three views.*

Proof: The existence of nine trilinear forms follow directly from (1) and (2). Let $\boldsymbol{\alpha}_{ij} = v_i'\boldsymbol{b}_j - v_j''\boldsymbol{a}_i$. The nine forms are given below (the first two are (3) and (4) repeated for convenience):

$$x''\boldsymbol{\alpha}_{13}^\top p - x''x'\boldsymbol{\alpha}_{33}^\top p + x'\boldsymbol{\alpha}_{31}^\top p - \boldsymbol{\alpha}_{11}^\top p = 0,$$
$$y''\boldsymbol{\alpha}_{13}^\top p - y''x'\boldsymbol{\alpha}_{33}^\top p + x'\boldsymbol{\alpha}_{32}^\top p - \boldsymbol{\alpha}_{12}^\top p = 0,$$
$$x''\boldsymbol{\alpha}_{23}^\top p - x''y'\boldsymbol{\alpha}_{33}^\top p + y'\boldsymbol{\alpha}_{31}^\top p - \boldsymbol{\alpha}_{21}^\top p = 0, \qquad (5)$$
$$y''\boldsymbol{\alpha}_{23}^\top p - y''y'\boldsymbol{\alpha}_{33}^\top p + y'\boldsymbol{\alpha}_{32}^\top p - \boldsymbol{\alpha}_{22}^\top p = 0, \qquad (6)$$
$$y''x'\boldsymbol{\alpha}_{31}^\top p - x''x'\boldsymbol{\alpha}_{32}^\top p + x''\boldsymbol{\alpha}_{12}^\top p - y''\boldsymbol{\alpha}_{11}^\top p = 0, \qquad (7)$$
$$y''y'\boldsymbol{\alpha}_{31}^\top p - x''y'\boldsymbol{\alpha}_{32}^\top p + x''\boldsymbol{\alpha}_{22}^\top p - y''\boldsymbol{\alpha}_{21}^\top p = 0, \qquad (8)$$
$$x''y'\boldsymbol{\alpha}_{13}^\top p - x''x'\boldsymbol{\alpha}_{23}^\top p + x'\boldsymbol{\alpha}_{21}^\top p - y'\boldsymbol{\alpha}_{11}^\top p = 0, \qquad (9)$$
$$y''y'\boldsymbol{\alpha}_{13}^\top p - y''x'\boldsymbol{\alpha}_{23}^\top p + x'\boldsymbol{\alpha}_{22}^\top p - y'\boldsymbol{\alpha}_{12}^\top p = 0, \qquad (10)$$
$$x''y'\boldsymbol{\alpha}_{12}^\top p - x''x'\boldsymbol{\alpha}_{22}^\top p + y''x'\boldsymbol{\alpha}_{21}^\top p - y''y'\boldsymbol{\alpha}_{11}^\top p = 0 \qquad (11)$$

For a given triplet p, p', p'' the first four functions on the list produce a 4×27 matrix. The rank of the matrix is four because it contains four orthogonal columns (columns associated with $\boldsymbol{\alpha}_{11}, \boldsymbol{\alpha}_{12}, \boldsymbol{\alpha}_{21}$ and $\boldsymbol{\alpha}_{22}$), therefore these functions are linearly independent. Since we have 27 coefficients, and each triplet p, p', p'' contributes four linear equations, then seven corresponding points across the three views provide a sufficient number of equations for a linear solution for the coefficients (given that the system is determined up to a common scale, seven points produce two extra equations which can be used for consistency checking or for obtaining a least squares solution).

The remaining trilinear forms are linearly spanned by the first four, as follows:

$$(7) = y''(3) - x''(4)$$
$$(8) = y''(5) - x''(6)$$
$$(9) = y'(3) - x'(5)$$
$$(10) = y'(4) - x'(6)$$
$$(11) = y''y'(3) - x''y'(4) + x''x'(6) - y''x'(5)$$

where the numbers in parenthesis represent the equation numbers of the various trilinear functions. □

Taken together, both theorems provide a constructive means for solving for the positions x'', y'' in a novel view given the correspondences p, p' across two model views. This process of generating a novel view can be easily accomplished without the need to explicitly recover structure, camera transformation, or even just the epipolar geometry — and requires fewer corresponding points than any other known alternative.

The solution for x'', y'' is unique without constraints on the allowed camera transformations. There are, however, certain camera configurations that require a different set of four trilinear functions from the one suggested in the proof of Theorem 2. For example, the set of equations (5), (6),(9) and (10) are also linearly independent. Thus, for example, in case v_1' and v_3' vanish simultaneously, i.e., $v' \cong (0, 1, 0)$, then that set should be used instead. Similarly, equations (3), (4),(9) and (10) are linearly independent, and should be used in case $v' \cong (1, 0, 0)$. Similar situations arise with

$v'' \cong (1,0,0)$ and $v'' \cong (0,1,0)$ which can be dealt by choosing the appropriate basis of four functions from the six discussed above. Note that we have not addressed the problem of singular configurations of seven points. For example, its clear that if the seven points are coplanar, then their correspondences across the three views could not possibly yield a unique solution to the problem of recovering the coefficients. The matter of singular surfaces has been studied for the eight-point case necessary for recovering the epipolar geometry [19, 14, 22]. The same matter concerning the results presented in this paper is an open problem.

Moving away from the need to recover the epipolar geometry carries distinct and significant advantages. To get a better idea of these advantages, we consider briefly the process of re-projection using epipolar geometry. The epipolar intersection method can be described succinctly (see [11]) as follows. Let F_{13} and F_{23} be the matrices ("fundamental" matrices in recent terminology [10]) that satisfy $p''F_{13}p = 0$, and $p''F_{23}p' = 0$. Then, by incidence of p'' with its epipolar line, we have:

$$p'' \cong F_{13}p \times F_{23}p'. \tag{12}$$

Therefore, eight corresponding points across the three views are sufficient for a linear solution of the two fundamental matrices, and then all other object points can be re-projected onto the third view. Equation (12) is also a trilinear form, but not of the type introduced in Theorem 1. The differences include (i) epipolar intersection requires the correspondences coming from eight points, rather than seven, (ii) the position of p'' is solved by a line intersection process which is singular in the case the three camera centers are collinear; in the trilinearity result the components of p'' are solved separately and the situation of three collinear cameras is admissible, (iii) the epipolar intersection process is decomposable, i.e., only two views are used at a time; whereas the epipolar geometries in the trilinearity result are intertwined and are not recoverable separately. The latter implies a better numerically behaved method in the presence of noise as well, and as will be shown later, the performance, even using the minimal number of required points, far exceeds the performance of epipolar intersection using many more points. In other words, by avoiding the need to recover the epipolar geometry we obtain a significant practical advantage as well, since the epipolar geometry is the most error-sensitive component when working with perspective views.

The connection between the general result of trilinear functions of views and the "linear combination of views" result [38] for orthographic views, can easily be seen by setting A and B to be affine in $\mathcal{P}^2$, and $v'_3 = v''_3 = 0$. For example, (3) reduces to

$$v'_1 x'' - v''_1 x' + (v''_1 \boldsymbol{a}_1 - v'_1 \boldsymbol{b}_1)^T p = 0,$$

which is of the form

$$\alpha_1 x'' + \alpha_2 x' + \alpha_3 x + \alpha_4 y + \alpha_5 = 0.$$

As in the perspective case, each point contributes four equations, but here there is no advantage for using all four of them to recover the coefficients, therefore we may use only two out of the four equations, and require four corresponding points to recover the coefficients. Thus, in the case where all three views are orthographic, x'' (y'') is expressed as a linear combination of image coordinates of the two other views — as discovered by [38].

IV. The Bilinear Form

Consider the case for which the two reference (model) views of an object are taken orthographically (using a tele lens would provide a reasonable approximation), but during recognition any perspective view of the object is allowed. It can easily be shown that the three views are then connected via bilinear functions (instead of trilinear):

Theorem 3 (Bilinearity) *Within the conditions of Theorem 1, in case the views ψ_1 and ψ_2 are obtained by parallel projection, then the pair of trilinear forms of Theorem 1 reduce to the following pair of bilinear equations:*

$$x''(\alpha_1 x + \alpha_2 y + \alpha_3) + \alpha_4 x'' x' + \alpha_5 x' + \alpha_6 x + \alpha_7 y + \alpha_8 = 0,$$

and

$$y''(\beta_1 x + \beta_2 y + \beta_3) + \beta_4 y'' x' + \beta_5 x' + \beta_6 x + \beta_7 y + \beta_8 = 0,$$

where $\alpha_j = \beta_j$, $j = 1, ..., 4$.

Proof: Under these conditions we have from Lemma 1 that A is affine in $\mathcal{P}^2$ and $v'_3 = 0$, therefore (3) reduces to:

$$x''(v'_1 \boldsymbol{b}_3 - v''_3 \boldsymbol{a}_1)^T p + v''_3 x'' x' - v''_1 x' + (v''_1 \boldsymbol{a}_1 - v'_1 \boldsymbol{b}_1)^T p = 0.$$

Similarly, (4) reduces to:

$$y''(v'_1 \boldsymbol{b}_3 - v''_3 \boldsymbol{a}_1)^T p + v''_3 y'' x' - v''_2 x' + (v''_2 \boldsymbol{a}_1 - v'_1 \boldsymbol{b}_2)^T p = 0.$$

Both equations are of the desired form, with the first four coefficients identical across both equations. □

The remaining trilinear forms undergo a similar reduction, and Theorem 2 still holds, i.e., we still have four linearly independent bilinear forms. Consequently, we have 21 coefficients up to a common scale (instead of 27) and four equations per point, thus five corresponding points (instead of seven) are sufficient for a linear solution.

A bilinear function of three views has two advantages over the general trilinear function. First, as mentioned above, only five corresponding points (instead of seven) across three views are required for solving for the coefficients. Second, the lower the degree of the algebraic function, the less sensitive the solution may be in the presence of errors in measuring correspondences. In other words, it is likely (though not necessary) that the higher order terms, such as the term $x''x'x$ in Equation 3, will have a higher contribution to the overall error sensitivity of the system.

Compared to the case when all views are assumed orthographic, this case is much less of an approximation. Since the model views are taken only once, it is not unreasonable to require that they be taken in a special way, namely, with a tele lens (assuming we are dealing with object recognition, rather than scene recognition). If this requirement is satisfied, then the recognition task is general since we allow any perspective view to be taken during the recognition process.

V. Experimental Data

The experiments described in this section were done in order to evaluate the practical aspect of using the trilinear result for re-projection compared to using epipolar intersection and the linear combination result of [38] (the latter we have shown is simply a limiting case of the trilinear result).

The epipolar intersection method was implemented as described in Section III by recovering first the fundamental matrices. Although eight corresponding points are sufficient for a linear solution, in practice one would use more than eight points for recovering the fundamental matrices in a linear or non-linear squares method. Since linear least squares methods are still sensitive to image noise, we used the implementation of a non-linear method described in [20] which was kindly provided by T. Luong and L. Quan (these were two implementations of the method proposed in [20] — in each case, the implementation that provided the better results was adopted).

The first experiment is with simulation data showing that even when the epipolar geometry is recovered accurately, it is still significantly better to use the trilinear result which avoids the process of line intersection. The second experiment is done on a real set of images, comparing the performance of the various methods and the number of corresponding points that are needed in practice to achieve reasonable re-projection results.

A. Computer Simulations

We used an object of 46 points placed randomly with z coordinates between 100 units and 120 units, and x, y coordinates ranging randomly between -125 and +125. Focal length was of 50 units and the first view was obtained by $fx/z, fy/z$. The second view (ψ_2) was generated by a rotation around the point $(0, 0, 100)$ with axis $(0.14, 0.7, 0.7)$ and by an angle of 0.3 radians. The third view (ψ_3) was generated by a rotation around an axis $(0, 1, 0)$ with the same translation and angle. Various amounts of random noise was applied to all points that were to be re-projected onto a third view, but not to the eight or seven points that were used for recovering the parameters (fundamental matrices, or trilinear coefficients). The noise was random, added separately to each coordinate and with varying levels from 0.5 to 2.5 pixel error. We have done 1000 trials as follows: 20 random objects were created, and for each degree of error the simulation was ran 10 times per object. We collected the maximal re-projection error (in pixels) and the average re-projection error (averaged of all the points that were re-projected). These numbers were collected separately for each degree of error by averaging over all trials (200 of them) and recording the standard deviation as well. Since no error were added to the eight or seven points that were used to determine the epipolar geometry and the trilinear coefficients, we simply solved the associated linear systems of equations required to obtain the fundamental matrices or the trilinear coefficients.

The results are shown in Figure 1. The graph on the left shows the performance of both algorithms for each level of image noise by measuring the maximal re-projection error. We see that under all noise levels, the trilinear method is significantly better and also has a smaller standard deviation. Similarly for the average re-projection error shown in the graph on the right.

This difference in performance is expected, as the trilinear method takes all three views together, rather than every pair separately, and thus avoids line intersections.

B. Experiments On Real Images

Figure 2 shows three views of the object we selected for the experiment. The object is a sports shoe with added texture to facilitate the correspondence process. This object was chosen because of its complexity, i.e., it has a shape of a natural object and cannot easily be described parametrically (as a collection of planes or algebraic surfaces). Note that the situation depicted here is challenging because the re-projected view is not in-between the two model views, i.e., one should expect a larger sensitivity to image noise than in-between situations. A set of 34 points were manually selected on one of the frames, ψ_1, and their correspondences were automatically obtained along all other frames used in this experiment. The correspondence process is based on an implementation of a coarse-to-fine optical-flow algorithm described in [7]. To achieve accurate correspondences across distant views, intermediate in-between frames were taken and the displacements across consecutive frames were added. The overall displacement field was then used to push ("warp") the first frame towards the target frame and thus create a synthetic image. Optical-flow was applied again between the synthetic frame and the target frame and the resulting displacement was added to the overall displacement obtained earlier. This process provides a dense displacement field which is then sampled to obtain the correspondences of the 34 points initially chosen in the first frame. The results of this process are shown in Figure 2 by displaying squares centered around the computed locations of the corresponding points. One can see that the correspondences obtained in this manner are reasonable, and in most cases to sub-pixel accuracy. One can readily automate further this process by selecting points in the first frame for which the Hessian matrix of spatial derivatives is well conditioned — similar to the confidence values suggested in the implementations of [4, 7, 34] — however, the intention here was not so much as to build a complete system but to test the performance of the trilinear re-projection method and compare it to the performance of epipolar intersection and the linear combination methods.

The trilinear method requires at least seven corresponding points across the three views (we need 26 equation, and seven points provide 28 equations), whereas epipolar intersection can be done (in principle) with eight points.

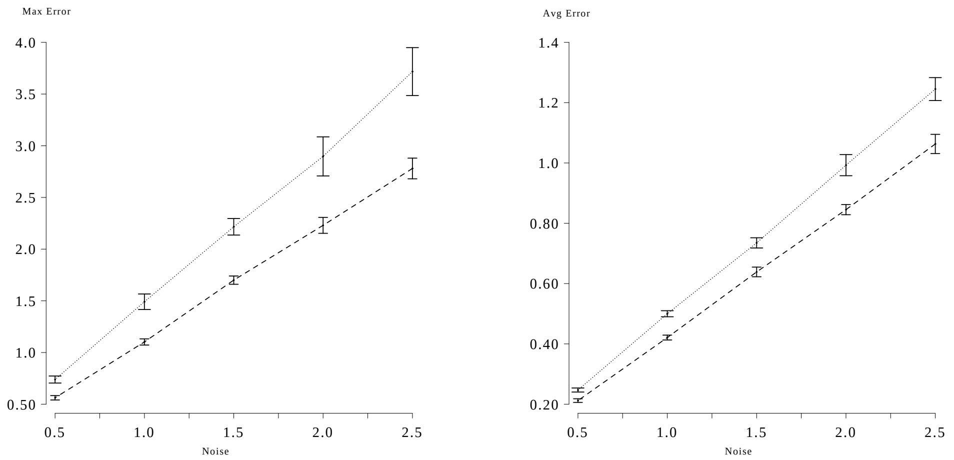

Fig. 1. Comparing the performance of the epipolar intersection method (the dotted line) and the trilinear functions method (dashed line) in the presence of image noise. The graph on the left shows the maximal re-projection error averaged over 200 trials per noise level (bars represent standard deviation). Graph on the right displays the average re-projection error averaged over all re-projected points averaged over the 200 trials per noise level.

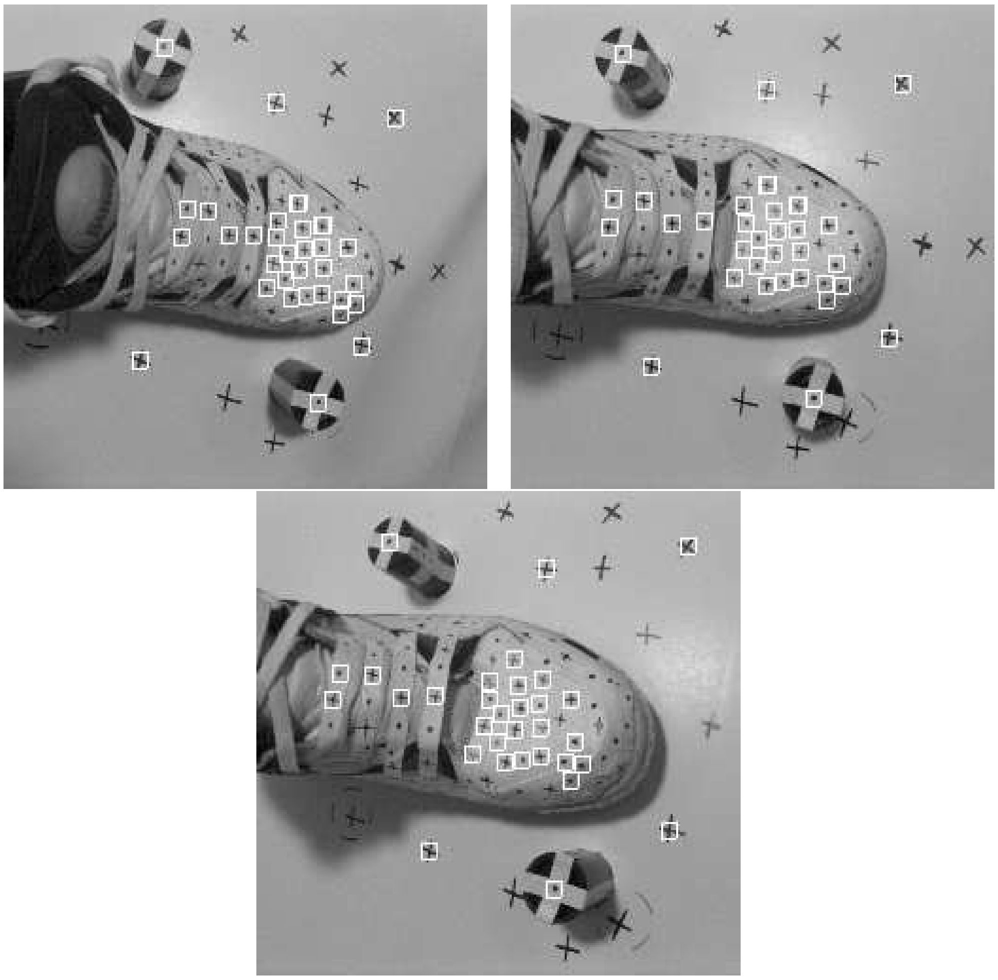

Fig. 2. *Top Row:* Two model views, ψ_1 on the left and ψ_2 on the right (image size are 256×240). The overlayed squares illustrate the corresponding points (34 points). *Bottom Row:* Third view ψ_3. Note that ψ_3 is not in-between ψ_1 and ψ_2, making the re-projection problem more challenging (i.e., performance is more sensitive to image noise than in-between situations).

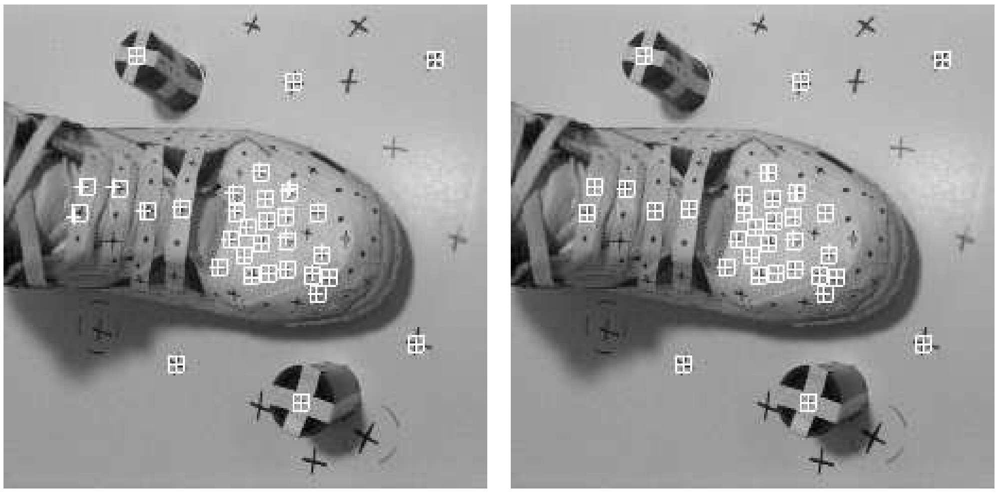

Fig. 3. Re-projection onto ψ_3 using the trilinear result. The re-projected points are marked as crosses, therefore should be at the center of the squares for accurate re-projection. On the left, the minimal number of points were used for recovering the trilinear coefficients (seven points); the average pixel error between the true an estimated locations is 0.98, and the maximal error is 3.3. On the right 10 points were used in a least squares fit; average error is 0.44 and maximal error is 1.44.

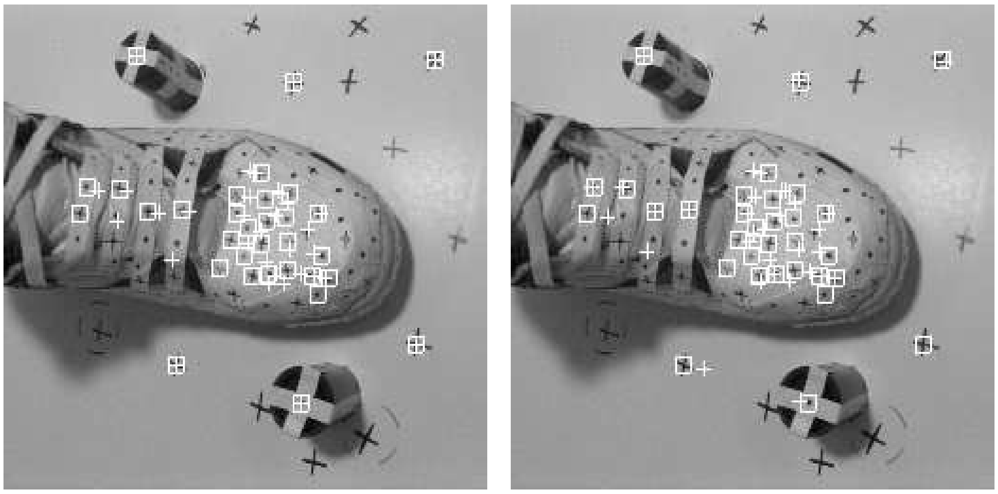

Fig. 4. Results of re-projection using intersection of epipolar lines. In the lefthand display the ground plane points were used for recovering the fundamental matrix (see text), and in the righthand display the fundamental matrices were recovered from the implementation of [20] using all 34 points across the three views. Maximum displacement error in the lefthand display is 25.7 pixels and average error is 7.7 pixels. Maximal error in the righthand display is 43.4 pixels and average error is 9.58 pixels.

The question we are about to address is what is the number of points that are required in practice (due to errors in correspondence, lens distortions and other effects that are not adequately modeled by the pin-hole camera model) to achieve reasonable performance?

The trilinear result was first applied with the minimal number of points (seven) for solving for the coefficients, and then applied with 8,9, and 10 points using a linear least-squares solution (note that in general, better solutions may be obtained by using SVD or Jacobi methods instead of linear least-squares, but that was not attempted here). The results are shown in Figure 3. Seven points provide a re-projection with maximal error of 3.3 pixels and average error of 0.98 pixels. The solution using 10 points provided an improvement with maximal error of 1.44 and average error of 0.44 pixels. The performance using eight and nine points was reasonably in-between the performances above. Using more points did not improve significantly the results; for example, when all 34 points were used the maximal error went down to 1.14 pixels and average error stayed at 0.42 pixels.

Next the epipolar intersection method was applied. We used two methods for recovering the fundamental matrices. One method is by using the implementation of [20], and the other is by taking advantage that four of the corresponding points are coming from a plane (the ground plane). In the former case, much more than eight points were required in order to achieve reasonable results. For example, when using all the 34 points, the maximal error was 43.4 pixels and the average error was 9.58 pixels. In the latter case, we recovered first the homography B due to the ground plane and then the epipole v'' using two additional points (those on the film cartridges). It is then known (see [28, 21, 32]) that $F_{13} = [v'']B$, where $[v'']$ is the anti-symmetric matrix of v''. A similar procedure was used to recover F_{23}. Therefore, only six points were used for re-projection,

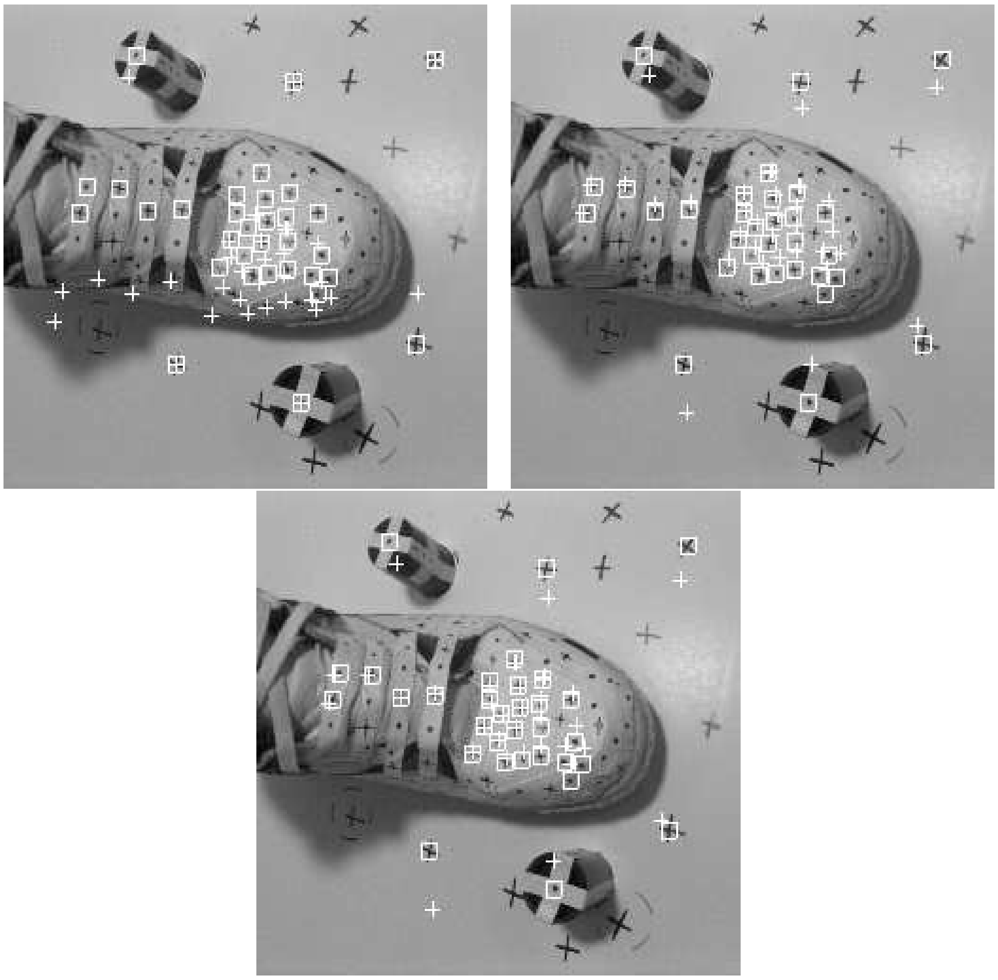

Fig. 5. Results of re-projection using the linear combination of views method proposed by [38] (applicable to parallel projection). *Top Row:* In the lefthand display the linear coefficients were recovered from four corresponding points; maximal error is 56.7 pixels and average error is 20.3 pixels. In the righthand display the coefficients were recovered using 10 points in a linear least squares fashion; maximal error is 24.3 pixels and average error is 6.8 pixels. *Bottom Row:* The coefficients were recovered using all 34 points across the three views. Maximal error is 29.4 pixels and average error is 5.03 pixels.

but nevertheless, the results were slightly better: maximal error of 25.7 pixels and average error of 7.7 pixels. Figure 4 shows these results.

Finally, we tested the performance of re-projection using the linear combination method. Since the linear combination method holds only for orthographic views, we are actually testing the orthographic assumption under a perspective situation, or in other words, whether the higher (bilinear and trilinear) order terms of the trilinear equations are significant or not. The linear combination method requires at least four corresponding points across the three views. We applied the method with four, 10 (for comparison with the trilinear case shown in Figure 3), and all 34 points (the latter two using linear least squares). The results are displayed in Figure 5. The performance in all cases are significantly poorer than when using the trilinear functions, but better than the epipolar intersection method.

VI. Discussion

We have seen that any view of a fixed 3D object can be expressed as a trilinear function with two reference views in the general case, or as a bilinear function when the reference views are created by means of parallel projection. These functions provide alternative, much simpler, means for manipulating views of a scene than other methods. Moreover, they require fewer corresponding points in theory, and much fewer in practice. Experimental results show that the trilinear functions are also useful in practice yielding performance that is significantly better than epipolar intersection or the linear combination method (although we emphasize that the linear combination was tested just to provide a base-line for comparison, i.e., to verify that the extra bilinear and trilinear terms indeed contribute to better performance).

In general two views admit a "fundamental" matrix (cf. [10]) representing the epipolar geometry between the two views, and whose elements are subject to a cubic constraint (rank of the matrix is 2). The trilinearity results (Theorems 1,2) imply, first, that three views admit a "fundamental" tensor with 27 distinct elements. Second, the robustness of the re-projection results may indicate that the elements of this tensor, are either independent, or constrained by a second order polynomial. In other words, in the two-view case, the elements of the fundamental matrix lie on third-degree hypersurface in $\mathcal{P}^9$. In the error-free case, a linear solution, ignoring the cubic constraint, is valid. However, in the presence of errors, the linear solution does not guarantee that the point in $\mathcal{P}^9$ (representing the solution) will lie on the hypersurface, hence a non-admissible solution is obtained. The non-linear solutions proposed by [20], do not address this problem directly, but in practice yield better behaved solutions than the linear ones. Since the trilinear case yields good performance in practice using linear methods, we arrive to the conjecture above. The notion of the "fundamental" tensor, its properties, relation to the geometry of three views, and applications to 3D reconstruction from multiple views, constitutes (in my mind) an important future direction.

The application that was emphasized throughout the paper is visual recognition via alignment. Reasonable performance was obtained with the minimal number of required points (seven) with the novel view (ψ_3) — which may be too many if the image to model matching is done by trying all possible combinations of point matches. The existence of bilinear functions in the special case where the model is orthographic, but the novel view is perspective, is more encouraging from the standpoint of counting points. Here we have the result that only five corresponding points are required to obtain recognition of perspective views (provided we can satisfy the requirement that the model is orthographic). We have not experimented with bilinear functions to see how many points would be needed in practice, but plan to do that in the future. Because of their simplicity, one may speculate that these algebraic functions will find uses in tasks other than visual recognition — some of those are discussed below.

There may exist other applications where simplicity is of major importance, whereas the number of points is less of a concern. Consider for example, the application of model-based compression. With the trilinear functions we need 17 parameters to represent a view as a function of two reference views in full correspondence (recall, 27 coefficients were used in order to reduce the number of corresponding points from nine to seven). Assume both the sender and the receiver have the two reference views and apply the same algorithm for obtaining correspondences between the two views. To send a third view (ignoring problems of self occlusions that may be dealt with separately) the sender can solve for the 17 parameters using many points, but eventually send only the 17 parameters. The receiver then simply combines the two reference views in a "trilinear way" given the received parameters. This is clearly a domain where the number of points is not a major concern, whereas simplicity, and robustness (as shown above) due to the short-cut in the computations, is of great importance.

Related to image coding, an approach of image decomposition into "layers" was recently proposed by [1, 2]. In this approach, a sequence of views is divided up into regions, whose motion of each is described approximately by a 2D affine transformation. The sender sends the first image followed only by the six affine parameters for each region for each subsequent frame. The use of algebraic functions of views can potentially make this approach more powerful because instead of dividing up the scene into planes one can attempt to divide the scene into objects, each carries the 17 parameters describing its displacement onto the subsequent frame.

Another area of application may be in computer graphics. Re-projection techniques provide a short-cut for image rendering. Given two fully rendered views of some 3D object, other views (again ignoring self-occlusions) can be rendered by simply "combining" the reference views. Again, the number of corresponding points is less of a concern here.

Acknowledgments

I acknowledge ONR grants N00014-92-J-1879 and N00014-93-1-0385, NSF grant ASC-9217041, and ARPA grant N00014-91-J-4038 as a source of funding for the Artificial Intelligence laboratory and for the Center for Biological Computational Learning. Also acknowledged is the McDonnell-Pew postdoctoral fellowship that has been my direct source of funding for the duration of this work. I thank T. Luong an L. Quan for providing their implementation for recovering fundamental matrices and epipoles. Thanks to N. Navab and A. Azarbayejani for assistance in capturing the image sequence (equipment courtesy of MIT Media Laboratory).

References

[1] E.H. Adelson. Layered representations for image coding. Technical Report 181, Media Laboratory, Massachusetts Institute of Technology, 1991.

[2] E.H. Adelson and J.Y.A. Wang. Layered representation for motion analysis. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 361–366, New York, NY, June 1993.

[3] G. Adiv. Inherent ambiguities in recovering 3D motion and structure from a noisy flow field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-11(5):477–489, 1989.

[4] P. Anandan. A unified perspective on computational techniques for the measurement of visual motion. In *Proceedings Image Understanding Workshop*, pages 219–230, Los Angeles, CA, February 1987. Morgan Kaufmann, San Mateo, CA.

[5] I.A. Bachelder and S. Ullman. Contour matching using local affine transformations. In *Proceedings Image Understanding Workshop*. Morgan Kaufmann, San Mateo, CA, 1992.

[6] E.B. Barrett, M.H. Brill, N.N. Haag, and P.M. Payton. Invariant linear methods in photogrammetry and model-matching. In J.L. Mundy and A. Zisserman, editors, *Applications of invariances in computer vision*. MIT Press, 1992.

[7] J.R. Bergen and R. Hingorani. Hierarchical motion-based frame rate conversion. Technical report, David Sarnoff Research Center, 1990.

[8] S. Demey, A. Zisserman, and P. Beardsley. Affine and projective structure from motion. In *Proceedings of the British Machine Vision Conference*, October 1992.

[9] R. Dutta and M.A. Synder. Robustness of correspondence based structure from motion. In *Proceedings of the International Conference on Computer Vision*, pages 106–110, Osaka, Japan, December 1990.

[10] 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.

[11] O.D. Faugeras and L. Robert. What can two images tell us about a third one? Technical Report INRIA, France, 1993.

[12] W.E.L. Grimson. Why stereo vision is not always about 3D reconstruction. A.I. Memo No. 1435, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, July 1993.

[13] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 761–764, Champaign, IL., June 1992.

[14] B.K.P. Horn. Relative orientation. *International Journal of Computer Vision*, 4:59–78, 1990.

[15] B.K.P. Horn. Relative orientation revisited. *Journal of the Optical Society of America*, 8:1630–1638, 1991.

[16] D.P. Huttenlocher and S. Ullman. Recognizing solid objects by alignment with an image. *International Journal of Computer Vision*, 5(2):195–212, 1990.

[17] J.J. Koenderink and A.J. Van Doorn. Affine structure from motion. *Journal of the Optical Society of America*, 8:377–385, 1991.

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

[19] H.C. Longuet-Higgins. the reconstruction of a scene from two projections – configurations that defeat the 8-point algorithm. In *Proc. of the 1st Conf. AI applications*, pages 395–397, Denver, December 1984.

[20] Q.T. Luong, R. Deriche, O.D. Faugeras, and T. Papadopoulo. On determining the fundamental matrix: Analysis of different methods and experimental results. Technical Report INRIA, France, 1993.

[21] Q.T. Luong and T. Vieville. Canonical representations for the geometries of multiple projective views. Technical Report INRIA, France, fall 1993.

[22] S.J. Maybank. The projective geometry of ambiguous surfaces. *Proceedings of the Royal Society of London*, 332:1–47, 1990.

[23] J. Mundy and A. Zisserman. Appendix — projective geometry for machine vision. In J. Mundy and A. Zisserman, editors, *Geometric invariances in computer vision*. MIT Press, Cambridge, 1992.

[24] J.L. Mundy, R.P. Welty, M.H. Brill, P.M. Payton, and E.B. Barrett. 3-D model alignment without computing pose. In *Proceedings Image Understanding Workshop*, pages 727–735. Morgan Kaufmann, San Mateo, CA, January 1992.

[25] A. Shashua. Correspondence and affine shape from two orthographic views: Motion and Recognition. A.I. Memo No. 1327, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, December 1991.

[26] A. Shashua. *Geometry and Photometry in 3D visual recognition*. PhD thesis, M.I.T Artificial Intelligence Laboratory, AI-TR-1401, November 1992.

[27] A. Shashua. Illumination and view position in 3D visual recognition. In S.J. Hanson J.E. Moody and R.P. Lippmann, editors, *Advances in Neural Information Processing Systems 4*, pages 404–411. San Mateo, CA: Morgan Kaufmann Publishers, 1992. Proceedings of the fourth annual conference NIPS, Dec. 1991, Denver, CO.

[28] A. Shashua. On geometric and algebraic aspects of 3D affine and projective structures from perspective 2D views. In *The 2nd European Workshop on Invariants*, Ponta Delagada, Azores, October 1993. Also in MIT AI memo No. 1405, July 1993.

[29] A. Shashua. Projective depth: A geometric invariant for 3D reconstruction from two perspective/orthographic views and for visual recognition. In *Proceedings of the International Conference on Computer Vision*, pages 583–590, Berlin, Germany, May 1993.

[30] A. Shashua. Projective structure from uncalibrated images: structure from motion and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1994. in press.

[31] A. Shashua. Trilinearity in visual recognition by alignment. In *Proceedings of the European Conference on Computer Vision*, Stockholm, Sweden, May 1994.

[32] A. Shashua and N. Navab. Relative affine structure: Theory and application to 3d reconstruction from perspective views. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, Seattle, Washington, 1994.

[33] A. Shashua and S. Toelg. The quadric reference surface: Applications in registering views of complex 3d objects. In *Proceedings of the European Conference on Computer Vision*, Stockholm, Sweden, May 1994.

[34] C. Tomasi and T. Kanade. Factoring image sequences into shape and motion. In *IEEE Workshop on Visual Motion*, pages 21–29, Princeton, NJ, September 1991.

[35] R.Y. Tsai and T.S. Huang. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surface. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6:13–26, 1984.

[36] S. Ullman. *The Interpretation of Visual Motion*. MIT Press, Cambridge and London, 1979.

[37] S. Ullman. Aligning pictorial descriptions: an approach to object recognition. *Cognition*, 32:193–254, 1989. Also: in MIT AI Memo 931, Dec. 1986.

[38] 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.

[39] D. Weinshall. Model based invariants for 3-D vision. *International Journal of Computer Vision*, 10(1):27–42, 1993.

Amnon Shashua received the B.Sc. degree in Mathematics and Computer Science from Tel-Aviv University, Tel-Aviv, Israel, in 1986; the M.Sc. degree in Mathematics and Computer Science from the Weizmann Institute of Science, Rehovot, Israel, in 1989; and the Ph.D. degree in Computational Neuroscience, working at the Artificial Intelligence Laboratory, from the Massachusetts Institute of Technology, in 1993. In 1993 he received a McDonnell-Pew postdoctoral fellowship and is currently at the Artificial Intelligence Laboratory and at the Center for Biological Computational Learning, MIT.