Trilinearity in Visual Recognition by Alignment

Amnon Shashua

Massachusetts Institute of Technology
Artificial Intelligence Laboratory
Department of Brain and Cognitive Sciences
Cambridge, MA 02139

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 method superior to the conventional epipolar line intersection method. The proof of the central result may be of further interest as it demonstrates certain regularities across homographies of the plane.

1 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, 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 [11]. 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.

The central theorem and a complete proof is presented in this paper. The proof itself may be of interest on its own because it reveals certain regularities across homographies of the plane. The trilinear result is demonstrated on real images with a comparison to other methods for achieving the same task (epipolar intersection and the linear combination of views methods). For more details on theoretical and practical aspects of this work, the reader is referred to [7].

2 The Trilinear Form

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

Lecture Notes in Computer Science, Vol. 800
Jan-Olof Eklundh (Ed.)
Computer Vision – ECCV '94

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.

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, undergoing at most a 3D affine transformation between views. 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:*

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

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

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

Lemma 2 (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 $p'_o \cong Ap_o + v'$, 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 $p' \cong Ap + kv'$. 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 [9]. 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 [3].

Definition 3. 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 2, 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 $p^i \cong A_i p + kv^i$, for any view ψ_i, where $v^i \in \psi_i$ is the epipole with ψ_1 (scaled arbitrarily).

Lemma 4 (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 $A - sA' = [\alpha v', \beta v', \gamma v']$, 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 [10], $Av \cong v'$ 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 $Hp \cong v'$ for all $p \in \psi_1$, and s_q is a fixed scalar s. This, in turn, implies that H is a matrix whose columns are multiples of v'. $\square$

Lemma 5 (Auxiliary for Lemma 6). *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, $B_1 \cong A_1C^{-1}$, and thus, we have $B \cong B_2^{-1}A_1C^{-1}$. 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 $A' = A_2^{-1}E_2E_1 = AA_1^{-1}E_2E_1 = AT$, and with proper scaling of C we have, $B' = B_2^{-1}E_2E_1C^{-1} = BCA_1^{-1}E_2E_1C^{-1} = BCTC^{-1}$. $\square$

Lemma 6 (Auxiliary — Uniqueness). *For scale-compatible homographies, the scalars s, α, β, γ of Lemma 4 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, $B - sB' = [\alpha v'', \beta v'', \gamma v'']$.*

Proof. We show first that s is invariant, i.e., that $B - sB'$ is a matrix whose columns are multiples of v''. Let H be a matrix whose columns are multiples of v'. From Lemma 4, and Lemma 5 we have $I - sT = A^{-1}H$, for some scalar s, and where $A' = AT$. After multiplying both sides by BC, and then pre-multiplying by C^{-1} we obtain $B - sBCTC^{-1} = BCA^{-1}HC^{-1}$. From Lemma 5, 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 2 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: $p' \cong Ap + kv' \cong A'p + k'v'$ and $p'' \cong Bp + kv'' \cong B'p + k'v''$. Then from Lemma 4 we have $Hp = (sk' - k)v'$ and $\hat{H}p = (sk' - k)v''$. 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. $\square$

Proof of Theorem: Lemma 2 provides the existence part of theorem, as follows. Since Lemma 2 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: $p' \cong Ap + kv'$, and $p'' \cong Bp + kv''$. By isolating k from both equations, and following some simple re-arrangements we obtain:

$$x''(v_1'b_3 - v_3''a_1)^Tp + x''x'(v_3''a_3 - v_3'b_3)^Tp + x'(v_3'b_1 - v_1''a_3)^Tp + (v_1''a_1 - v_1'b_1)^Tp = 0, \quad (1)$$

where b_1, b_2, b_3 and a_1, a_2, a_3 are the row vectors of A and B and $v' = (v_1', v_2', v_3')$, $v'' = (v_1'', v_2'', v_3'')$. In a similar fashion, following a different re-arrangement, we obtain:

$$y''(v_1'b_3 - v_3''a_1)^T p + y''x'(v_3''a_3 - v_3'b_3)^T p + x'(v_3'b_2 - v_2''a_3)^T p + (v_2''a_1 - v_1'b_2)^T p = 0. \quad (2)$$

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

The question of uniqueness arises because Lemma 2 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 Equations 1 and 2 have the general form: $v_j' \boldsymbol{b}_i \pm v_i'' \boldsymbol{a}_j$, for some i and j. Thus, we need to show that there exists a scalar s that satisfies $v_i''(\boldsymbol{a}_j - s\boldsymbol{a}_j') = v_j'(\boldsymbol{b}_i - s\boldsymbol{b}_i')$. This, however, follows directly from Lemmas 4 and 6. $\square$

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).

Taken together, the process of generating a novel view can be easily accomplished without the need to explicitly recover structure, camera transformation, or just the epipolar geometry. The process described here is fundamentally different from intersecting epipolar lines [6, 1, 8, 5, 2] in the following ways: first, we use the three views together, instead of pairs of views separately; second, there is no process of line intersection, i.e., the x and y coordinates of ψ_3 are obtained separately as a solution of a single equation in coordinates of the other two views; and thirdly, the process is well defined in cases where intersecting epipolar lines becomes singular (e.g., when the three camera centers are collinear). Furthermore, by avoiding the need to recover the epipolar geometry (e.g., the epipoles v, v', v'' or the matrices F_{13} and F_{23} satisfying $p''F_{13}p = 0$, and $p''F_{23}p' = 0$, respectively), we obtain a significant practical advantage, 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 to the "linear combination of views" result [11] 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, Equation 1 reduces to $v_1'x'' - v_1''x' + (v_1''\boldsymbol{a}_1 \cdot p - v_1'\boldsymbol{b}_1 \cdot p) = 0$, which is of the form $\alpha_1 x'' + \alpha_2 x' + \alpha_3 x + \alpha_4 y + \alpha_5 = 0$. In the case where all three views are orthographic, then x'' is expressed as a linear combination of image coordinates of the two other views — as discovered by [11].

3 Experimental Data

Figure 1 demonstrates re-projection using the trilinear result on a triple of three real images. The re-projection result was also compared with the methods of epipolar intersection and the linear combination of views (which, as shown here, is a limiting case of the trilinear result). The epipolar intersection was obtained in

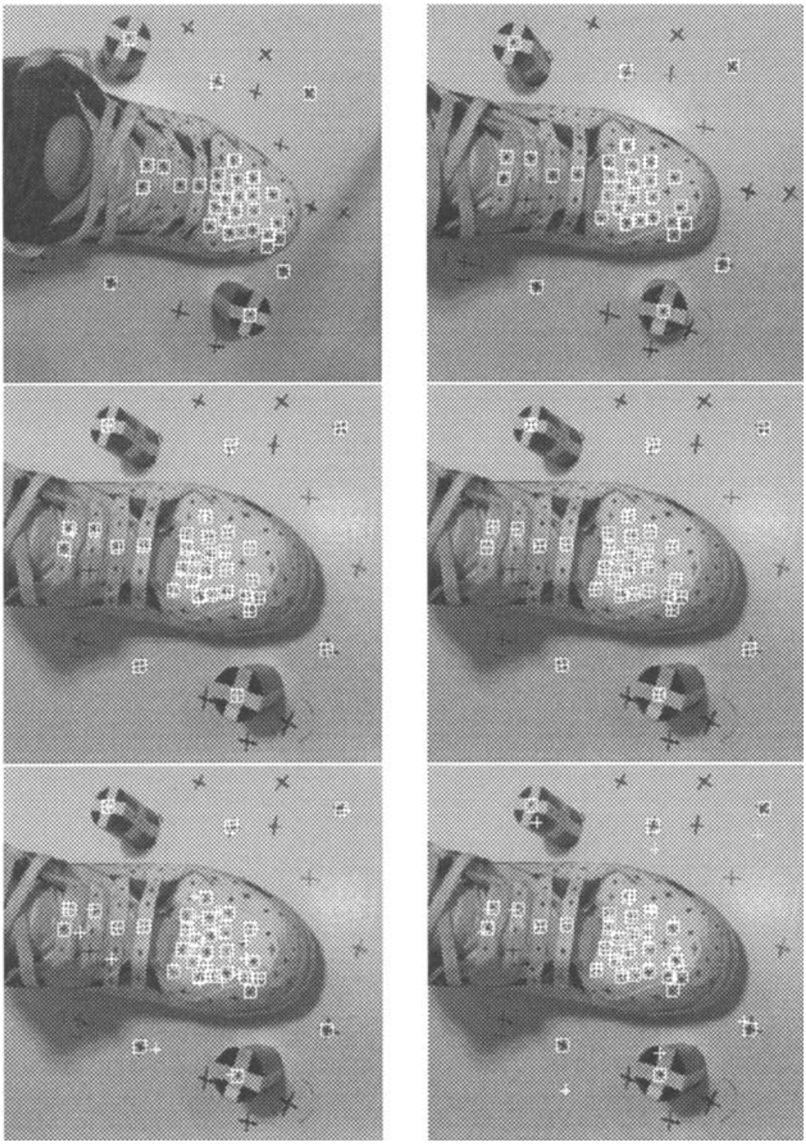

Fig. 1. *Top Row:* Two model views, ψ_1 on the left and ψ_2 on the right. The overlayed squares illustrate the corresponding points (34 points). *Middle Row:* Re-projection onto ψ_3 using the trilinear result. The overlayed squares illustrate the true location of the corresponding points (p''), and the crosses illustrate the estimated locations. On the left only nine points were used; the average pixel error between the true an estimated locations is 1.4, and the maximal error is 5.7. On the right 12 points were used in a least squares fit; average error is 0.4 and maximal error is 1.4. *Bottom Row:* On the left the epipolar intersection method was applied (using all 34 points); average error is 9.58 and maximal error is 43.4. On the right the linear combination method was applied (using all 34 points); average error is 5.03 and maximal error is 29.4.

the standard way by recovering the matrices F_{13} and F_{23} satisfying $p''F_{13}p = 0$, and $p''F_{23}p' = 0$, respectively. Those matrices were recovered using *all* the available points using the non-linear method (currently the state-of-the-art) proposed by [4] (code was kindly provided by T. Luong and L. Quan). Re-projection is obtained by $p'' \cong F_{13}p \times F_{23}p'$.

Note that the situation depicted here is challenging because the re-projected view is not in-between the two model views. The trilinear result was first applied with the minimal number of points (nine) for solving for the coefficients, and then applied with twelve points using a linear least-squares solution. This is compared to using 34 points for the epipolar intersection and the linear combination methods.

References

1. 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.
2. O.D. Faugeras and L. Robert. What can two images tell us about a third one? Technical Report INRIA, France, 1993.
3. J.J. Koenderink and A.J. Van Doorn. Affine structure from motion. *Journal of the Optical Society of America*, 8:377–385, 1991.
4. 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.
5. 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.
6. 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.
7. A. Shashua. Algebraic functions for recognition. Submitted for publication, Jan. 1994. Also in MIT AI Memo No. 1452, Jan. 1994.
8. A. Shashua. *Geometry and Photometry in 3D visual recognition.* PhD thesis, M.I.T Artificial Intelligence Laboratory, AI-TR-1401, November 1992.
9. A. Shashua. On geometric and algebraic aspects of 3D affine and projective structures from perspective 2D views. In *The 2nd European Workshop on Invariants*, Azores Islands, Portugal, October 1993. Also in MIT AI memo No. 1405, July 1993.
10. 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.
11. 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.