Relative Affine Structure: Theory and Application to 3D Reconstruction From Perspective Views

Amnon Shashua

Artificial Intelligence Laboratory
Center for Biol. and Comp. Learning
Massachusetts Institute of Technology
Cambridge, MA 02139

Nassir Navab

The Media Laboratory
Perceptual Computing Group
Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract

We propose an affine framework for perspective views, captured by a single extremely simple equation based on a viewer-centered invariant we call relative affine structure. Via a number of corollaries of our main results we show that our framework unifies previous work — including Euclidean, projective and affine — in a natural and simple way. Finally, the main results were applied to a real image sequence for purpose of 3D reconstruction from 2D views.

1 Introduction

The introduction of affine and projective tools into the field of computer vision have brought increased activity in the fields of structure from motion and recognition by alignment in the recent few years. The emerging realization is that non-metric information, although weaker than the information provided by depth maps and rigid camera geometries, is nonetheless useful in the sense that the framework may provide simpler algorithms, camera calibration is not required, more freedom in picture-taking is allowed — such as taking pictures of pictures of objects — and there is no need to make a distinction between orthographic and perspective projections.

The material introduced so far in the literature focuses on the projective framework [4, 8, 25, 15], or the affine framework. The latter requires either assuming parallel projection (cf. [10, 32, 31, 22]), or certain apriori assumptions on object structure (for determining the location of the plane at infinity [4, 19]), or assuming purely translational camera motion [16].

In this paper, we propose a unified framework that includes by generalization and specialization the Euclidean, projective and affine frameworks. The basic idea is to choose a representation of projective space in which an arbitrarily chosen reference plane becomes the plane at infinity. What is left to show is that under general, uncalibrated, camera motion, the resulting new representations can be described by an element of the affine group applied to the initial representation. As a result, we obtain an affine invariant, we call *relative affine structure*, relative to the initial representation. Via several corollaries of this basic result we show, among other things, that the invariant is a generalization of the affine structure under parallel projection [10] and is a specialization of the projective structure (projective structure can be described as a ratio of two relative affine structures). Furthermore, in computational terms the relative affine result requires fewer corresponding points and fewer calculations than the projective framework, and is the only next general framework after projective when working with perspective views.

2 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 (or scene) is modeled by a set of points and let $\psi_i \subset \mathcal{P}^2$ denote views (arbitrary), indexed by i, of the object. Given two views with projection centers $O, O' \in \mathcal{P}^3$, respectively, the epipoles are defined as 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. When only two views ψ_o, ψ_1 are discussed, then points in ψ_o are denoted by p, their corresponding points in ψ_1 are denoted by p', and the epipoles are $v \in \psi_o$ and $v' \in \psi_1$. 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.

3 Relative Affine Structure

The following theorems and corollaries introduce our main results which are then followed by explanatory text.

Theorem 1 (Relative Affine Structure[23]) *Let π be some arbitrary plane and let $P_j \in \pi$, $j = 1, 2, 3$ projecting onto p_j, p'_j in views ψ_o, ψ_1, respectively. Let $p_o \in \psi_o$ and $p'_o \in \psi_1$ be projections of $P_o \notin \pi$. Let $A \in PGL_3$ be a homography of $\mathcal{P}^2$ determined by the equations $Ap_j \cong p'_j$, $j = 1, 2, 3$, and $Av \cong v'$, scaled to satisfy the equation $p'_o \cong Ap_o + v'$. Then, for any point $P \in \mathcal{P}^3$ projecting onto $p \in \psi_o$ and $p' \in \psi_1$, we*

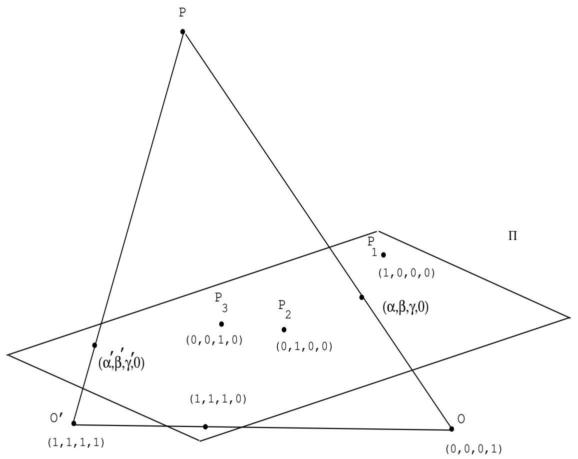


Figure 1: See proof of Theorem 1.

have

$$\boxed{p' \cong Ap + k\boldsymbol{v}'} \tag{1}$$

The coefficient $k = k(p)$ is independent of ψ_1, i.e., is invariant to the choice of the second view, and the coordinates of P are $(x, y, 1, k)$.

Proof. We assign the coordinates $(1,0,0,0), (0,1,0,0), (0,0,1,0)$ to P_1, P_2, P_3, respectively. Let O and O' be the projection centers associated with the views ψ_o and ψ_1, respectively, and let their coordinates be $(0,0,0,1), (1,1,1,1)$, respectively (see Figure 1). This choice of representation is always possible because the two cameras are part of $\mathcal{P}^3$. By construction, the point of intersection of the line $\overline{OO'}$ with π has the coordinates $(1,1,1,0)$.

Let P be some object point projecting onto p, p'. The line $\overline{OP}$ intersects π at the point $(\alpha, \beta, \gamma, 0)$. The coordinates α, β, γ can be recovered by projecting the image plane onto π, as follows. Given the epipoles $v \in \psi_o$ and $v' \in \psi_1$, we have by our choice of coordinates that p_1, p_2, p_3 and v are projectively (in $\mathcal{P}^2$) mapped onto $e_1 = (1,0,0), e_2 = (0,1,0), e_3 = (0,0,1)$ and $e_4 = (1,1,1)$, respectively. Therefore, there exists a unique element $A_1 \in PGL_3$ that satisfies $A_1 p_j \cong e_j$, $j = 1,2,3$, and $A_1 v = e_4$. Note that we have made a choice of scale by setting $A_1 v$ to e_4, this is simply for convenience as will be clear later on. Let $A_1 p = (\alpha, \beta, \gamma)$.

Similarly, the line $\overline{O'P}$ intersects π at $(\alpha', \beta', \gamma', 0)$. Let $A_2 \in PGL_3$ be defined by $A_2 p'_j \cong e_j$, $j = 1,2,3$, and $A_2 v' = e_4$. Let $A_2 p' = (\alpha', \beta', \gamma')$. Since P can be described as a linear combination of two points along each of the lines $\overline{OP}$, and $\overline{O'P}$, we have the following equation:

$$P \cong \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ 0 \end{pmatrix} - k \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \mu \begin{pmatrix} \alpha' \\ \beta' \\ \gamma' \\ 0 \end{pmatrix} - s \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix},$$

from which it readily follows that $k = s$ (i.e., the transformation between the two representations of $\mathcal{P}^3$ is affine). Note that since only ratios of coordinates are significant in $\mathcal{P}^n$, k is determined up to a uniform scale, and any point $P_o \notin \pi$ can be used to set a mutual scale for all views — by setting an appropriate scale for A, for example. The value of k can easily be determined from image measurements as follows: we have $\mu(\alpha', \beta', \gamma')^\top = (\alpha, \beta, \gamma)^\top + k(1,1,1)^\top$. Multiply both sides by A_2^{-1} to obtain $\mu p' = Ap + k\boldsymbol{v}'$, where $A = A_2^{-1} A_1$. Note that $A \in PGL_3$ is a homography between the two image planes, due to π, determined by $p'_j \cong A p_j$, $j = 1,2,3$, and $Av \cong v'$ (therefore, can be recovered directly without going through A_1, A_2). Similar proofs that a homography of a plane can be recovered from three points and the epipoles are found in [24, 20]. Since k is determined up to a uniform scale, we need a fourth correspondence p_o, p'_o, and let A, or v', be scaled such that $p'_o \cong A p_o + v'$. Finally, $(x, y, 1, k)$ are the homogeneous coordinates representation of P, and the 3×4 matrix $[A, v']$ is a camera transformation matrix between the two views. ∎

Theorem 2 (Algebraic Description of Relative Affine Depth[18]) *Let the coordinate transform from $P = zp$ to $P' = z'p'$ be described by $P' = M'RMP + M'T$, where R, T are the rotational and translational parameters of the relative camera displacement, and M, M' are the internal camera parameters (the transformation from camera coordinate systems to the observed coordinate systems in which $p = (x, y, 1)^\top$ and $p' = (x', y', 1)$). Given A, π, k defined in Theorem 1, let n be the unit normal to the plane π, and d_π the perpendicular distance of the origin to π, both in the first camera coordinate frame. Then,*

$$A \cong M'(R + \frac{Tn^\top}{d_\pi})M^{-1}, \tag{2}$$

and

$$k = \frac{z}{z_o}\alpha,$$

where z_o is the depth of $P_o \notin \pi$, and $\alpha = \alpha(p)$ is the affine structure of P in the case of parallel projection (the ratio of normal distances of P and P_o from π).

Proof. Let $\tilde{P}$ be at the intersection of the ray $\overline{OP}$ with π. Then $\tilde{P}' = M'RM^{-1}\tilde{P} + M'T$. Since $n^\top M^{-1}\tilde{P} = d_\pi$, we have: $\tilde{P}' = M'(R + \frac{Tn^\top}{d_\pi})M^{-1}\tilde{P}$. Since the term in parentheses describes the homography due to π, we have $A \cong M'(R + \frac{Tn^\top}{d_\pi})M^{-1}$ — which is the generalization of the classical motion of planes in the calibrated case [6, 29]. For the point P we have:

$$\begin{aligned} \frac{z'}{z}p' &= M'RMp + \frac{1}{z}M'T \\ &= Ap + \left[\frac{1}{z} - \frac{n^\top(M^{-1}p)}{d_\pi}\right] M'T \\ &\cong Ap + \left[\frac{d_\pi - n^\top(M^{-1}P)}{zd_\pi}\right] v'. \end{aligned}$$

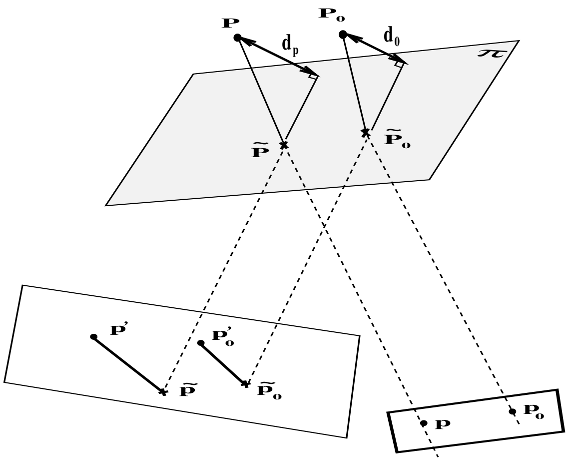


Figure 2: Affine structure under parallel projection is d_p/d_o. This can be seen from the similarity of trapezoids followed by the similarity of triangles: $\frac{p'-\tilde{p}}{p'_o-\tilde{p}'_o} = \frac{P'-\tilde{P}}{P'_o-\tilde{P}'_o} = \frac{d_p}{d_o}$.

Denote $d_p = d_\pi - n^\top(M^{-1}P)$ the perpendicular distance from P to π. We thus have

$$k = \frac{z_o}{z}\frac{d_p}{d_o},$$

where d_o is the perpendicular distance of P_o from π. Finally, note that the ratio $\alpha = d_p/d_o$ of the perpendicular distances of P and P_o from π is the affine structure when the projection is parallel (see Figure 2). □

Corollary 1 *Relative affine structure k approaches affine structure under parallel projection when O goes to infinity, i.e., $k \longrightarrow \alpha$ when $O \longrightarrow \infty$.*

Proof. When $O \longrightarrow \infty$, then $z, z_o \longrightarrow \infty$. Thus $k = \frac{z_o}{z}\frac{d_p}{d_o} \longrightarrow \frac{d_p}{d_o}$. □

Corollary 2 *When the plane π is at infinity, then relative affine structure k is affine structure under perspective $k = z_o/z$, $A = M'RM^{-1}$, and, if in addition, the cameras are internally calibrated as $M = M' = I$, then $A = R$.*

Proof. When π is at infinity, then $d_p, d_o \longrightarrow \infty$. Thus $k = \frac{z_o}{z}\frac{d_p}{d_o} \longrightarrow \frac{z_o}{z}$. Also, $d_\pi \longrightarrow \infty$, thus $A \longrightarrow M'RM^{-1}$. □

Corollary 3 *The "essential" matrix $E = [v']R$ is a particular case of a generalized matrix $F = [v']A$. The matrix F, referred to as "fundamental" matrix in [4], is unique and does not depend on the plane π.*

Proof. Let $p \in \psi_o, p' \in \psi_1$ be two corresponding points, and let l, l' be their corresponding epipolar lines, i.e., $l \cong p \times v$ and $l' \cong p' \times v'$. Since lines are projective invariants, then any point along l is mapped by A to some point along l'. Thus, $l' \cong v' \times Ap$, and because p' is incident to l', we have ${p'}^\top(v' \times Ap) = 0$, or equivalently: ${p'}^\top[v']Ap = 0$, or ${p'}^\top Fp = 0$, where $F = [v']A$. From Corollary 2, $A = R$ in the special case where the plane π is at infinity and the cameras are internally calibrated as $M = M' = I$, thus $E = [v']R$ is a special case of F. The uniqueness of F follows from substitution of A with Equation 2 and noting that $[v']T = 0$, thus $F = [v']M'RM^{-1}$. □

Corollary 4 *The projective structure of the scene can be described as the ratio of two relative affine structures each with respect to a distinct reference plane π_1, π_2, respectively, which in turn can be described as the ratio of affine structures under parallel projection with respect to the same two planes.*

Proof. Let k_{π_1} and k_{π_2} be the relative affine structures with respect to planes π_1 and π_2, respectively. From Theorem 2 we have that $k_{\pi_1} = \frac{z}{z_o}\frac{d_p^{\pi_1}}{d_o^{\pi_1}}$ and $k_{\pi_2} = \frac{z}{z_o}\frac{d_p^{\pi_2}}{d_o^{\pi_2}}$. The ratio k_{π_1}/k_{π_2} removes the dependence on the projection center O (z/z_o cancels out) and is therefore a projective invariant. This projective invariant is the ratio of cross-ratios of the rays $\overline{OP}$ and $\overline{OP_o}$ with their intersections with the two planes π_1 and π_2, which was introduced in [24, 25] as "projective depth". Because $d_p^{\pi_i}/d_o^{\pi_i}$ is the affine structure under parallel projection with respect to plane π_i (the ratio of perpendicular distances of P and P_o from π_i), then the projective structure of the scene can be described (up to a uniform scale factor) as the ratio of perpendicular distances of P from two reference planes[18]. □

Corollary 5 (stream of views) *Given $m \geq 2$ views, let A_j and v'_j be the homographies of π and the epipoles, respectively, from view ψ_o to view ψ_j, and let the views of an object point P be p_j where the index j ranges over the m views. Then, the least squares solution for k is given by*

$$k = \frac{\sum_j (p_j \times v'_j)^T (A_j p_o \times p_j)}{\sum_j \parallel p_j \times v'_j \parallel^2}. \quad (3)$$

Proof. This is simply a calculation based on the observation that given a general equation of the type $\boldsymbol{a} \cong \boldsymbol{b} + k\boldsymbol{c}$, then by performing a cross product with $\boldsymbol{a}$ on both sides we get: $k(\boldsymbol{a} \times \boldsymbol{c}) = \boldsymbol{b} \times \boldsymbol{a}$. The value of k can be found using the normal equations (treating k as a vector of dimension 1):

$$k = \frac{(b \times a)^T (a \times c)}{\|a \times c\|^2}.$$

Similarly, if in addition we have $\boldsymbol{a}' \cong \boldsymbol{b}' + k\boldsymbol{c}'$, then the overall least squares solution is given by

$$k = \frac{(b \times a)^T (a \times c) + (b' \times a')^T (a' \times c')}{\|a \times c\|^2 + \|a' \times c'\|^2}. \square$$

The key idea in Theorem 1 was to use both camera centers as part of the reference frame in order to show that the transformation between an arbitrary representation $\mathcal{R}_o$ of space as seen from the first camera and the representation $\mathcal{R}$ as seen from any other camera position, can be described by an element of the affine group. In other words, we have chosen an arbitrary plane π and made a choice of representation $\mathcal{R}_o$ in which π is the plane at infinity (i.e., π was mapped to infinity — not an unfamiliar trick, especially in computer graphics). The representation $\mathcal{R}_o$ is associated with $[x, y, 1, k]$ where k vanishes for all points coplanar with π, which means that π is the plane at infinity under the representation $\mathcal{R}_o$. What was left to show is that π remains the plane at infinity under all subsequent camera transformations, and therefore k is an affine invariant. Because k is invariant relative to the representation $\mathcal{R}_o$ we named it "relative affine structure"; this should not be confused with the term "relative invariants" used in classical invariant theory (multiplied by a power of the transformation determinant, as opposed to "absolute invariants").

In practical terms, the difference between a full projective framework (like in [4, 8, 25]) and the relative affine framework can be described as follows. In a full projective framework, if we denote by f the invariance function acting on a pair of views indexed by a fixed set of five corresponding points, then $f(\psi_i, \psi_j)$ is fixed for all i, j. In a relative affine framework, if we denote f_o as the invariance function acting on a fixed view ψ_o and an arbitrary view ψ_i and indexed by a fixed set of four corresponding points, then $f_o(\psi_o, \psi_i)$ is fixed for all i.

The remaining theorem 2 and corollaries put the relative affine framework within the familiar context of affine structure under parallel and perspective projections, Euclidean structure and projective structure. The homography A due to the plane π was described as a product of the rigid camera motion parameters, the parameters of π, and the internal camera parameters of both cameras. This result was similarly described in [14]. The relative affine structure k was described as a product of the affine structure under parallel projection and a term that contains the location of the camera center of the reference view. Geometrically, k is the product of two ratios, the first being the ratio of the perpendicular distance of a point P to the plane π and the depth z to the reference camera, and the second ratio is of the same form but applied to a fixed point P_o which is used to set a uniform scale to the system. Therefore, when the depth goes to infinity (projection approaches orthography), then k approaches the ratio of the perpendicular distances of P from π and the perpendicular distance of P_o from π — which is precisely the affine structure under parallel projection [10]. Thus, relative affine structure is a generalization in the sense of including the center of projection of an arbitrary camera, and when the camera center goes to infinity we obtain an affine structure which becomes independent of the reference camera.

Another specialization of relative affine structure was shown in Corollary 2 by considering the case when π is at infinity with respect to our Euclidean frame (i.e., really at infinity). In that case k is simply inverse depth (up to a uniform scale factor), and the homography A is the familiar rotational component of camera motion (orthogonal matrix R) in the case of calibrated cameras, or a product of R with the internal calibration parameters (see also [14]). In other words, when π is at infinity also with respect to our world coordinate frame, then relative affine becomes affine (the plane at infinity is preserved under all representations [4]). In practice, of course, we cannot recover the homography due to the plane at infinity unless we are given prior information on the nature of the scene structure [19], or the camera motion is purely translational [16]. Thus in the general case, we can realize either the relative affine framework or the projective framework.

The connection between the relative affine structure and projective structure was shown in Corollary 4. Projective invariants are necessarily described with reference to five scene points [4], or equivalently, with reference to two planes and a point laying outside of them both [25, 24]. Corollary 4 shows that by taking the ratio of two relative affine structures, each relative to a different reference plane, then the dependence on the camera center (the term z_o/z) drops and we are left with the projective invariant described in [25], which is the ratio of the perpendicular distance of a point to two planes [18](up to a uniform scale factor).

Corollary 3 unifies previous results on the nature of what is known by now as the "fundamental matrix" [4, 5]. It is shown, that for any plane π and its corresponding homography A we have $F = [v']A$. First, we see that given a homography, the epipole v' follows by having two corresponding points coming from scene points not coplanar with π — an observation that was originally made by [11]. Second, F is fixed, regardless of the choice of π, which was shown by using the result of Theorem 2. As a particular case, the product $[v']R$ remains fixed if we add to R a element that vanishes as a product with $[v']$ — an observation that was made previously by [8]. Thirdly, the "essential" matrix [12], $E = [v']R$, is shown to be a specialization of F in the case π is at infinity with respect to the world coordinate frame and the cameras are internally calibrated as $M = M' = I$.

Finally, Corollary 5 provides a practical formula for obtaining a least-squares estimation of relative affine structure which also applies for the case where a stream of views is available — in the spirit of [32, 28, 15, 27, 1, 2]. In the next section we apply these results to obtain a simple algorithm for relative affine reconstruction from multiple $m \geq 2$ views and multiple points.

3.1 Application I: Reconstruction from a Stream of Views

Taken together, the results above demonstrate the ability to compute relative affine structure using many points over many views in a least squares manner (see also [18]). At minimum we need two views and four corresponding points and the corresponding epipoles to recover k for all other points of the scene whose projections onto the two views are given. Let p_{ij}, $i = 0, ..., n$ and $j = 0, ..., m$ denote the i'th image point on frame j. Let A_j denote the homography from frame 0 to frame j, v_j, v'_j the corresponding epipoles such that $A_j v_j \cong v'_j$, and let k_i denote the relative affine structure of point i. We follow these steps:

1. Compute epipoles v_j, v'_j using the relation $p_{ij} F_j p_{io} = 0$, over all i. Eight corresponding points (frame 0 and frame j) are needed for a linear solution, and a least-squares solution is possible if more points are available. In practice the best results were obtained using the non-linear algorithm of [13]. The epipoles follow by $F_j v_j = 0$ and $F^\top v'_j = 0$ [4]. The latter readily follows from Corollary 3 as $[v'_j] A_j v_j \cong [v'_j] v'_j = 0$ and $A_j^\top [v'_j]^\top v'_j = -A_j^\top [v'_j] v'_j = 0$.
2. Compute A_j from the equations $A_j p_{io} \cong p_{ij}$, $i = 1, 2, 3$, and $A_j v_j \cong v'_j$. This leads to a linear set of eight equations for solving for A_j up to a scale. A least squares solution is available from the equation $p_{ij} [v'_j] A_j p_{io} = 0$ for all additional points (Corollary 3). Scale A_j to satisfy $p_{oj} \cong A_j p_{oo} + v'_j$.
3. Relative affine structure k_i is given by (3).

3.2 Application II: Recognition by Alignment

The relative affine invariance relation, captured by Theorem 1, can be used for visual recognition by alignment ([30, 9], and references therein). In other words, the invariance of k can be used to "re-project" the object onto any third view p'', as follows. Given two "model" views in full correspondence $p_i \longleftrightarrow p'_i$, $i = 1, ..., n$, we recover the epipoles and homography A from $Ap_i \cong p'_i$, $i = 1, 2, 3$, and $Av \cong v'$. Then the corresponding points p''_i in any third view satisfy $p'' \cong Bp + k\boldsymbol{v}''$, for some matrix B and epipole v''. One can solve for B and v'' by observing six corresponding points between the first and third view. Once B, v'' are recovered, we can find the estimated location of p''_i for the remaining points p_i, $i = 7, ..., n$, by first solving for k_i from the equation $p'_i \cong Ap_i + k_i v'$, and then substituting the result in the equation $\hat{p}''_i \cong Bp_i + k_i \boldsymbol{v}''$. Recognition is achieved if the distance between p''_i and $\hat{p}''_i$, $i = 7, ..., n$, is sufficiently small. Other methods for achieving reprojection include the epipolar intersection method (cf. [17, 3, 7]), or by using projective structure instead of the relative affine structure. In all the above methods the epipolar geometry plays a key and preconditioned role. More direct methods, that do not require the epipolar geometry can be found in [21, 26].

4 Experimental Results

We show here some of the real image experiments that were conducted — more details can be found in [18]. We used three views of a sport shoe (Figure 3) and established point correspondences over the three frames (marked by squares).

Epipoles were recovered by the following two methods. First, by using the four ground points to recover the homography A, and then by Corollary 3 to compute the epipoles using all the remaining points in a least squares manner. Second, using the non-linear algorithm proposed by [13]. The two methods gave rise to very similar results for reconstruction, and slightly different results for reprojection (see later).

In the reconstruction application (Section 3.1), the relative affine structure was recovered from the two extreme views (displays (a) and (d)). The transformation to Euclidean coordinates was done for purposes of display by assuming that the ground plane is parallel to the image plane (it actually is not) and that the camera is calibrated (there was no calibration attempt made). The 3D coordinates were shown from various views, and two of them are shown in displays (c) and (e). The side view (display (e)) illustrates the robustness of the reconstruction process, as it was obtained by rotating the object around a different axis than the one used for capturing the images. When multiple frames were used for reconstruction (via Corollary 5), the results where very similar (not shown here).

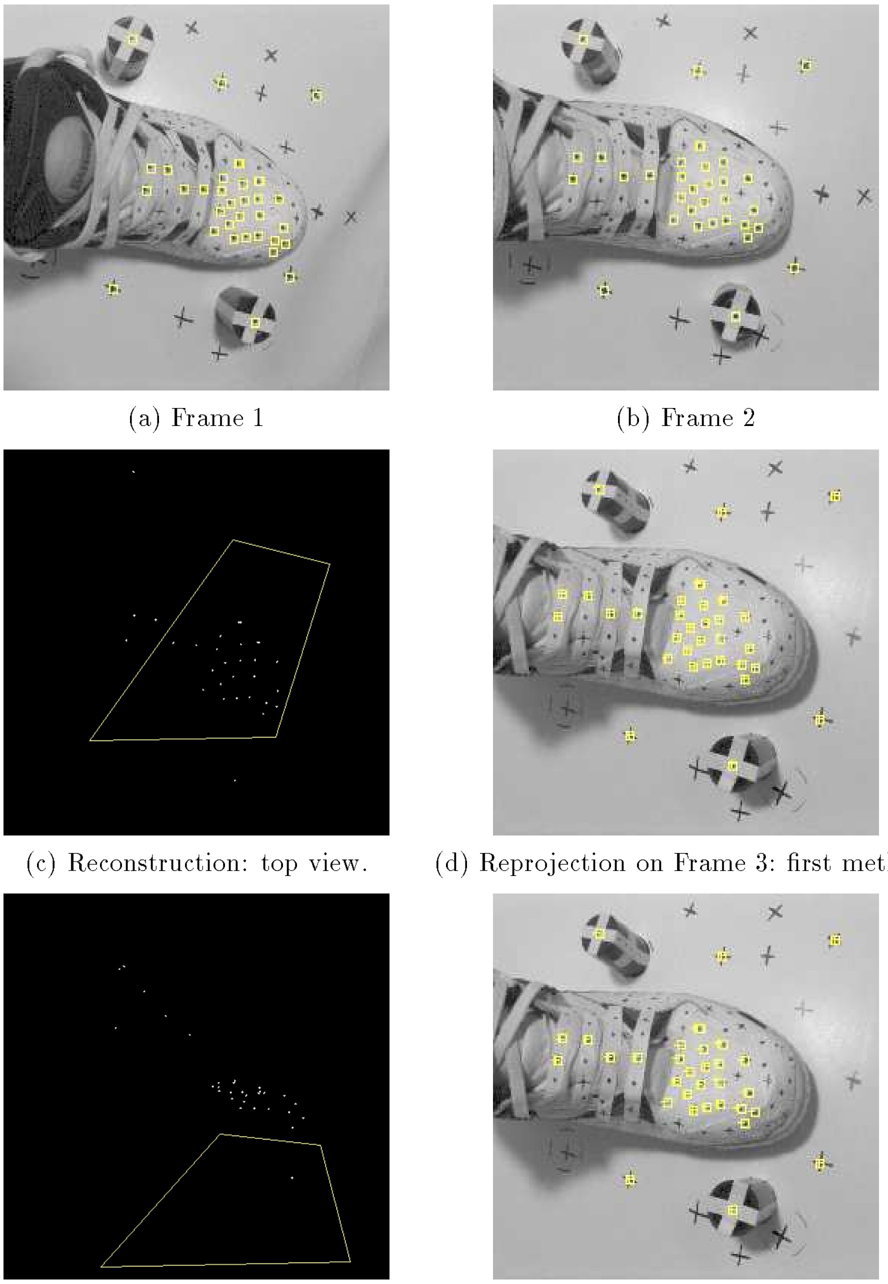
In the reprojection application (Section 3.2), relative affine structure was recovered using the first and in-between views (displays (a) and (b)), and reprojected onto the extreme view (display (d)). In the case where the ground plane was used for recovering the epipoles, the average error of reprojection was 0.7 pixels and maximal error was 2.1 pixels (display (d)). In the other case the average error was 1.7 and the maximal error was 3.9 pixels (display (f)). Note that because all points were used for recovering the epipoles, the reprojection performance only indicates the level of accuracy one can obtain when all the information is being used. In practice we would like to use much fewer points from the reprojected view, and therefore, reprojection methods that avoid the epipoles all together would be preferred — an example of such a method can be found in [21, 26].

5 Summary

We have proposed the computation of "relative affine structure" which is the next general framework after projective. The equations are extremely simple; are useful for applications of 3D reconstruction from point matches over two or more views and applications of recognition; and form a bridge unifying projective, affine and Euclidean frameworks.

Acknowledgments

We acknowledge ONR grants N00014-92-J-1879 and N00014-93-1-0385, NSF grant ASC-9217041, and ARPA grant N00014-91-J-4038. N. Navab thanks Aaron Bobick for financial support, through the Nichidai Foundation fellowship, and for his encouragement throughout this project.



(a) Frame 1 (b) Frame 2

(c) Reconstruction: top view. (d) Reprojection on Frame 3: first method.

(e) Reconstruction: side view. (f) Reprojection on Frame 3: second method.

Figure 3: *Reconstruction:* Relative affine structure was recovered from views (a) and (d). A top view and side view of the reconstructed shoe are shown in (c) and (e), respectively. *Reprojection:* Relative affine structure was recovered from views (a) and (b) and reprojected onto (d) (shown by overlayed crosses). The crosses on (d) were obtained by computing the epipoles using the ground-plane homography, recovered from the four ground points, and the remaining points in a least squares manner (Corollary3). The crosses on (f) were obtained by recovering the epipoles using the non-linear algorithm proposed by [13].

References

[1] Ali J. Azarbayejani, Bradley Horowitz, and Alex Pentland. Recursive estimation of structure and motion using relative orientation constraints. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 294–299, New York, NY, June 1993.

[2] T.J. Broida and R. Chellappa Kinematics and structure of a rigid object from a sequence of noisy images: A batch approach. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 176–182, Miami, FL, June 1986.

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

[4] 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.

[5] O.D. Faugeras, Q.T. Luong, and S.J. Maybank. Camera self calibration: Theory and experiments. In *Proceedings of the European Conference on Computer Vision*, pages 321–334, Santa Margherita Ligure, Italy, June 1992.

[6] 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, 1986.

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

[8] 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.

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

[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] C.H. Lee. Structure and motion from two perspective views via planar patch. In *Proceedings of the International Conference on Computer Vision*, pages 158–164, Tampa, FL, December 1988.

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

[13] 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.

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

[15] R. Mohr, F. Veillon, ans L. Quan. Relative 3D Reconstruction Using Multiple Uncalibrated Images. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 543–548, New York, NY, June 1993.

[16] T. Moons, L. Van Gool, M. Van Diest, and E. Pauwels. Affine reconstruction from perspective image pairs. In *The 2nd European Workshop on Invariants*, Ponta Delagada, Azores, October 1993.

[17] 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.

[18] N. Navab, and A. Shashua. Algebraic Description of Relative Affine Structure : Connections to Euclidean, Affine and Projective Structure. Media Lab Memo No. 270, The Media Laboratory, Massachusetts Institute of Technology, March 1994.

[19] L. Quan. Affine stereo calibration for relative affine shape reconstruction. In *Proceedings of the British Machine Vision Conference*, pages 659–668, 1993.

[20] 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.

[21] A. Shashua. Algebraic functions for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1994. in press.

[22] 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.

[23] 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.

[24] 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.

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

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

[27] R. Szeliski and S.B. Kang. Recovering 3D shape and motion from image streams using non-linear least squares. Technical Report D.E.C., December 1992.

[28] 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.

[29] R.Y. Tsai, and T.S. Huang, Estimating 3-D motion parameters of a rigid planar patch. IEEE Trans. on Acoustic, Speech and Signal Processing. 29(6):1147–1152, December 1981.

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

[31] 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.

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