M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 270
March 1994, A similar version of this paper appears in ISPRS Conference, Munich, Sept. 1994.

Algebraic Derivations of Relative Affine Structure and Applications to 3D Reconstruction from 2D Views

Nassir Navab

The Media Laboratory
Perceptual Computing Group
Massachusetts Institute of Technology
Cambridge, MA 02139
U.S.A

Amnon Shashua

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

Key Words: 3D Reconstruction From 2D Views, Visual Recognition, Projective Geometry, Structure Representation, Invariants.

Abstract

We present an algebraic description of relative affine structure —an invariant structure affinely related to the structure seen from one of the cameras[16]. This algebraic description yields a simple canonical framework that unifies results in projective, affine and Euclidean structure from motion. We then introduce a method for capturing the redundancy in recovering relative affine structure from a stream of perspective views. We propose a certain decomposition that on one hand involves an optimal projection of the contribution of each point at each frame onto a single bilinear equation; and on the other hand reveals a connection (bilinear) between the homography of an arbitrary plane and the translational component of motion. Given an estimation of the epipoles, which can be computed in a least squares manner for each frame separately, the decomposition equations yield a linear least squares method for solving for scene structure. The main results were applied a real image sequence for purpose of 3D reconstruction from 2D views, visual recognition by alignment, and image coding.

1 Introduction

The geometric relation between 3D objects and their views is a key component for various applications in computer vision, image coding, and animation. For example, the change in the 2D projection of a moving 3D object is a source of information for 3D reconstruction, and for visual recognition applications — in the former case the retinal changes produce the cues for 3D recovery, and in the latter case the retinal changes provide the cues for factoring out the effects of changing viewing positions on the recognition process.

In the recent few years there is a growing realization that affine and projective tools can be useful by obtaining deeper understandings of the underlying geometric issues — such as the effects of internal camera calibration, the kind of structure representations associated with allowable groups of transformations, and issues of stability — and simpler, better behaved, algorithms. The material introduced so far in the literature focuses on the projective framework [3, 5, 10], or the affine framework. The latter requires either assuming parallel projection [7], or certain apriori assumptions on object structure (for determining the location of the plane at infinity [3, 14]), or assuming purely translational camera motion [2]. The framework of relative affine, in which the invariant structure is affinely related to the structure seen from one of the cameras, was presented in [16]. The framework is captured by a single equation relating an invariant structure, a homography of some arbitrary plane, the epipoles, and image correspondences.

In this paper, we give an algebraic description of "relative affine structure". Using this description we show that the framework of relative affine structure is a unified framework that includes by generalization and specialization the Euclidean, projective and affine frameworks. We show that the relative affine structure is a generalization of the affine structure under parallel projection [7] 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.

We then propose a linear algorithm for capturing redundancy of structure and motion of a stream of perspective views. We show that the contribution of every image point at each frame can be decomposed into two bilinear equations, one representing motion (in an affine sense, to be described later) and the other representing both motion and relative affine structure. This decomposition gives rise to a linear over-determination of relative affine shape and motion over many points and many frames. Instead of requiring an initial solution for bootstrapping we require the epipoles, which can be computed using many points between any pair of views (cf.[9]).

Finally, experimental results on real data demonstrate significant robustness of our approach as applied to problems of reconstruction, recognition and image coding.

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

Lemma 1 *(Relative Affine Structure)*

Let p_0, p_1, p_2, p_3 and p'_0, p'_1, p'_2, p'_3 be four corresponding points coming from four non-coplanar points in space. Let A be a homography of P^2 determined by the equations $Ap_j \cong p'_j$, $j = 1, 2, 3$, and $Av \cong v'$. Finally let v' be scaled such that $p'_0 \cong Ap_0 + v'$. Then, for any point $P \in P^3$ projecting onto p and p', we have

$$p' \cong Ap + kv' \tag{1}$$

The coefficient $k = k(p)$ is independent of Ψ_2, i.e., is invariant of the choice of the second view, and the homogeneous coordinates of M are $(x, y, 1, k)$.

The lemma, its proof and its theoretical implications are discussed in detail in [16]. The scalar k is called a *relative affine invariant.*

Theorem 1 (Algebraic Description of Relative Affine Depth) *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 Lemma 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 [4, 18] (see appendix for homography from line features). For the point P we have:

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

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. $\square$

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 $k = \frac{z_o}{z}\frac{d_p}{d_o} \longrightarrow \frac{d_p}{d_o}$. 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 1-a). $\square$

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 $A \longrightarrow M'RM^{-1}$ (see Figure 1-c). $\square$

Corollary 3 *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 1 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 [17] 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. $\square$

3.1 The Decomposition Method: Reconstruction from a Stream of Views

Theorem 2 (Decomposition Equations) *Given two arbitrary views $\psi_1,\ \psi_2 \subset P^2$ of a set of points $\Theta \subset P^3$. Let $A \in PGL_3$ be a homography, due to some plane π, as in Lemma 1. Every pair of corresponding points p and p' satisfy two bilinear equations of the form:*

$$\mathbf{e}_1^t A p = 0 \qquad (3)$$

$$\mathbf{e}_2^t A p + k \mathbf{e}_2^t v' = 0 \qquad (4)$$

where

$$\mathbf{e}_1 = v' \times p' \ , \qquad \mathbf{e}_2 = \mathbf{e}_1 \times p',$$

and $k = k(p)$ is the scalar defined in Lemma 1.

Proof of Theorem: The two bilinear equations (3) and (4) are obtained by a projection of equation (1) onto two directions. One is orthogonal to v' and p', and the other is orthogonal to p' and as close as possible to v'. Letting $\mathbf{e}_1$ and $\mathbf{e}_2$ be these directions we obtain the bilinear system of equations (3,4). $\square$

Corollary 4 *The matrix F, referred to as "fundamental matrix" in [3], that satisfies $p'Fp = 0$ for all corresponding pairs p and p', can be decomposed in the form of $F = [v']_x A$, where A is a homography due to some plane (defined in Lemma 1), and $[v']_x$ is the skew-symmetric matrix associated with v', i.e., $[v']_x w = v' \times w$ for any vector w.*

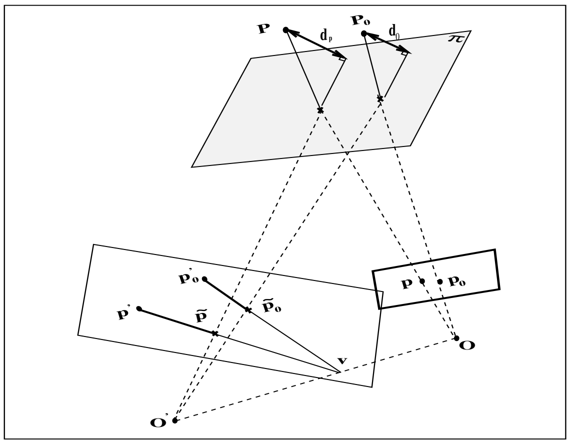
Proof: Follows immediately from theorem 2 by expanding equation (3):

$$\mathbf{e}_1^t A p = {p'}^t [v']_x A p = 0, \qquad (5)$$

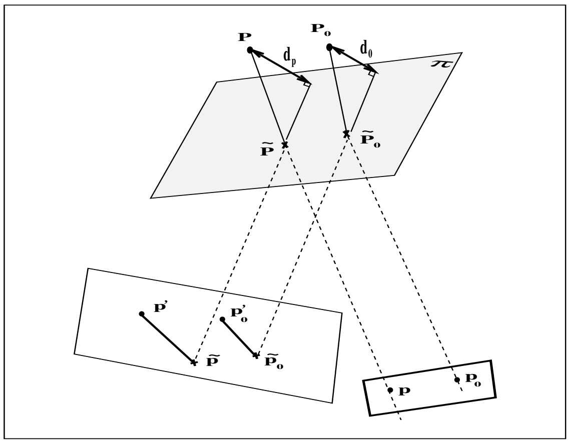
$\square$

The decomposition of the Fundamental matrix, equation 5 in theorem 4, provides a simple generalization of the essential matrix. Note that the essential matrix E can be decomposed as $E = [v']_x R$ where R is the rotational component of camera motion. Note that $M'RM^{-1}$ is the homography due to the plane at infinity (Corollary 2). Thus, in case the cameras are internally calibrated, i.e. $M = M' = I$, then R is the homography due to the plane at infinity. Note that this decomposition of the fundamental matrix is also a generalization of previous representations, such as in [5].

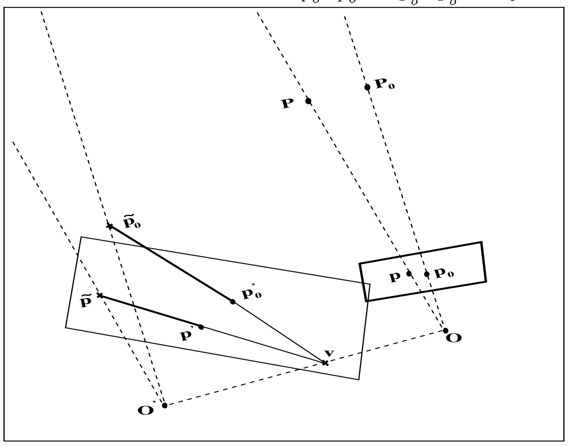
The decomposition $F = [v']_x A$ gives rise to practical results as well. For example, the homography A is



(a)



(b) 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: $\frac{p'-\tilde{p}}{p'_o-\tilde{p}'_o} = \frac{P'-\tilde{P}}{P'_o-\tilde{P}'_o} = \frac{d_p}{d_o}$



(c)

Figure 1: (a) Relative affine Structure: $k = \frac{z_o}{z}\frac{d_p}{d_o}$. (b) Affine structure under parallel projection: when $O \longrightarrow \infty$ then $z, z_o \longrightarrow \infty$. Thus $k \longrightarrow \frac{d_p}{d_o}$. (c) Affine structure under perspective (when π is at infinity). Note that the rays $\overline{OP}$ and $\overline{O'\tilde{p}}$ are parallel, thus the homography is the rotational component of motion.

uniquely determined from the epipoles and three corresponding points [17]. The equation ${p'}^t[v']_x Ap = 0$ provides, in addition, the means for finding a least squares estimate of A given additional corresponding pairs (see also [15] for alternative methods for over-determinating the homography).

Theorem 3 *Among all possible projections of $p' \cong Ap + kv'$ onto some direction (yielding a single equation), the projection onto $\mathbf{e}_2$, defined in Theorem 2, is unique in the sense that it provides the least squares solution for k.*

Proof: First the least squares solution for k can be obtained from equation (1) as follows: perform a cross product with p' on both sides of equation (1) to get $k(v' \times p') = p' \times Ap$, the least squares solution can be found using the normal equations (treating k as a vector of dimension 1):

$$k = \frac{(p' \times Ap)^t (v' \times p')}{\|v' \times p'\|^2}$$

This solution is the same as the one obtained by solving equation (4):

$$k = -\frac{\mathbf{e}_2^t Ap}{\mathbf{e}_2^t v'}$$

Replace $\mathbf{e}_2$ by its value and following some simple algebraic manipulations. we have:

$$\begin{aligned} \mathbf{e}_2^t Ap &= [(v' \times p') \times p']^t Ap \\ &= (p' \times Ap)^t (v' \times p') \end{aligned}$$

and also:

$$\begin{aligned} \mathbf{e}_2^t v' &= [(v' \times p') \times p']^t v' \\ &= -\|v' \times p'\|^2 \end{aligned}$$

□

Corollary 5 *Given $n \geq 2$ views, let A_i and v_i' be the homographies of π and the epipoles, respectively, from view ψ_o to view ψ_i, and let the views of an object point P be p_i where the index i ranges over the n views. Then, the least squares solution for k is given by*

$$k = \frac{\sum_{i=1}^n (p_i \times A_i p_0)^t (v_i \times p_i)}{\sum_{i=1}^n \|v_i \times p_i\|^2}. \tag{6}$$

Taken together, the algorithm for recovering a least squares solution for relative affine structure from a stream of views $\psi_0, ..., \psi_n$ can be summarized as follows:

- compute $v_i \in \psi_i$, $i = 1..n$, the epipole with ψ_0,
- compute homographies A_i, $i = 1..n$, in a least squares fashion using 3 arbitrary points to define the plane of homography (equation 1 with k=0) and all remaining points using equation 3 (the first projection equation).
- for each set of correspondences (p_0, p_i), $i = 1..n$, k follows from equation 6.

4 Experimental Results

The following experiments were conducted to illustrate the applications that arise from the relative affine framework (reconstruction, recognition by alignment, and image coding) and to test the algorithms on real data. The performance under real imaging situations is interesting, in particular, because of the presence of deviations from the pinhole camera model (radial distortions, decentering, and other effects), and due to errors in obtaining image correspondences.

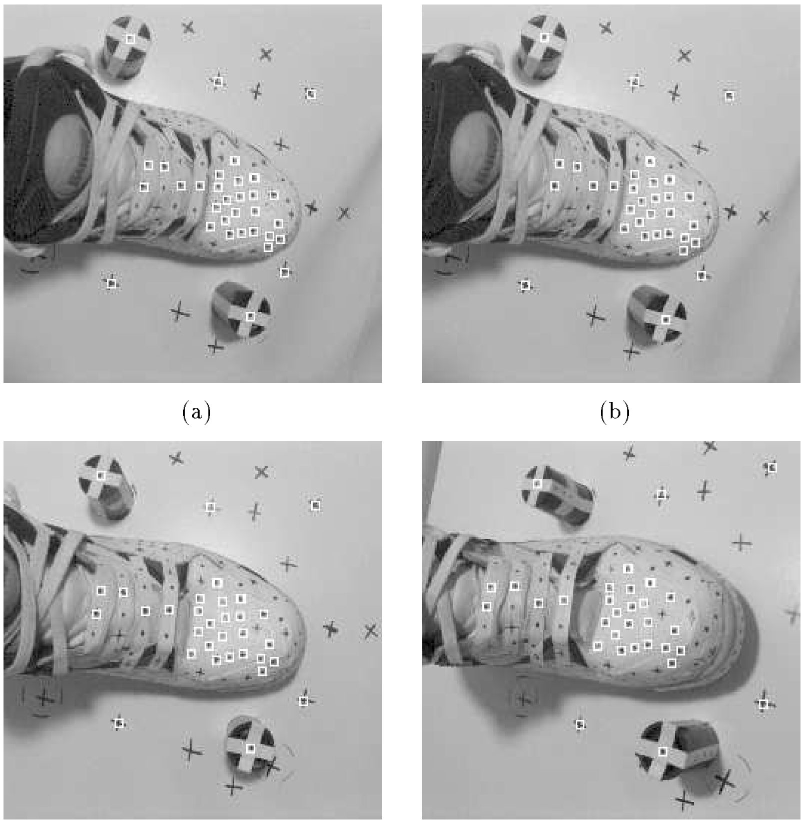
Fig. 2 shows four views, out of a sequence of ten views, of the object we selected for experiments. The object is a sneaker with added texture to facilitate the correspondence process. This object was chosen because of its complexity, i.e., it has a shape of a natural object and cannot easily be described parameterically (as a collection of planes or algebraic surfaces). A set of thirty-four points were manually selected on one of the frames, referred to as the first frame, and their correspondences were automatically obtained along all other frames used in this experiment (corresponding points are marked by overlapping squares in Fig. 2). The correspondence process is based on an implementation of a coarse-to-fine optical-flow algorithm based on [8] and described in [1].

Epipoles were recovered by either one of the following two methods. First, by using the four ground points to recover the homography A, and then by Corollary 4 to compute the epipoles using all the remaining points in a least squares manner. Second, using the non-linear algorithm proposed by [9]. The two methods gave rise to very similar results for reconstruction, and slightly different results for re-projection (see later).

4.1 Application I: Reconstruction From Two or More Views

We recovered relative affine structure from two views and multiple views. In the two-view case we used either a small base-line (the first two views of the sequence) or a large base-line (the first and last views of the sequence). In the multiple view case, we used all ten views of the sequence (Corollary 5). 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 are shown in Fig. 3. Display (a) shows a frontal view (in order to visually align the display with the image of the sneaker). Other displays show a side view of the reconstructed sneaker under the following experimental situations. Display (b) is due to reconstruction under large base-line situation (the two methods for obtaining the epipoles produced very similar results; the multiple-view case produced very similar results as well). The side view 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. Display (c) is due to reconstruction under small base-line situation (both methods for obtaining the epipoles produced very similar results). The quality of reconstruction in the latter case is not as good as



(a) (b) (c) (d)

Figure 2: Four views, out of a sequence of ten views, of a sneaker. The frames shown here are the first, second, fifth and tenth of the sequence (top-bottom, left-to-right). The overlayed squares mark the corresponding points that were tracked and subsequently used for our experiments.

in the former, as should be expected. Nevertheless, the system does not totally brake-down under relatively small base-line situations and produces a reasonable result under these circumstances.

4.2 Application II: Re-projection

Another way for testing the robustness of the relative affine framework, is by applying it to the problem of visual recognition by alignment ([6], and references therein). Since k remains fixed under changes of viewing positions, any new view can be synthesized given correspondences across two "model" views. Lemma 1 indicates that in order to synthesize a new view, one must recover first the epipole and the homography (due to the same plane π). We refer to the process of synthesizing new views by "re-projection".

In the re-projection application, relative affine structure was recovered using the first and in-between views, and re-projected onto the last view of the sequence. Note that this is an extrapolation example, thereby performance is expected to be poorer than interpolation examples, i.e., when the re-projected view is in-between the model views. The interpolation case will be discussed in the next section, where relevance to image coding applications is argued for.

In general, the performance was better when the ground plane was used for recovering the epipoles. When the intermediate view was the fifth in the sequence (Fig. 2, display (c)), the average error in re-projection was 1.1 pixels (with standard deviation of 0.98 pixels). When the intermediate view was the second frame in the sequence (Fig. 2, display (b)), the results were poorer (due to small base-line and large extrapolation) with average error of 7.81 pixels (standard deviation of 6.5). These two cases are displayed in Fig. 4. The re-projected points are represented by crosses overlayed on the last frame (the re-projected view).

When the second method for computing the epipoles was used (more general, but generally less accurate), the results were as follows. With the fifth frame, the average error was 1.62 pixels (standard deviation of 1.2); and with the second frame (small base-line situation) the average error was 13.87 pixels (standard deviation of 9.47). These two cases are displayed in Fig. 5. Note that because all points were used for recovering the epipoles, the re-projection performance, only indicates the level of accuracy one can obtain when all the information is being used.

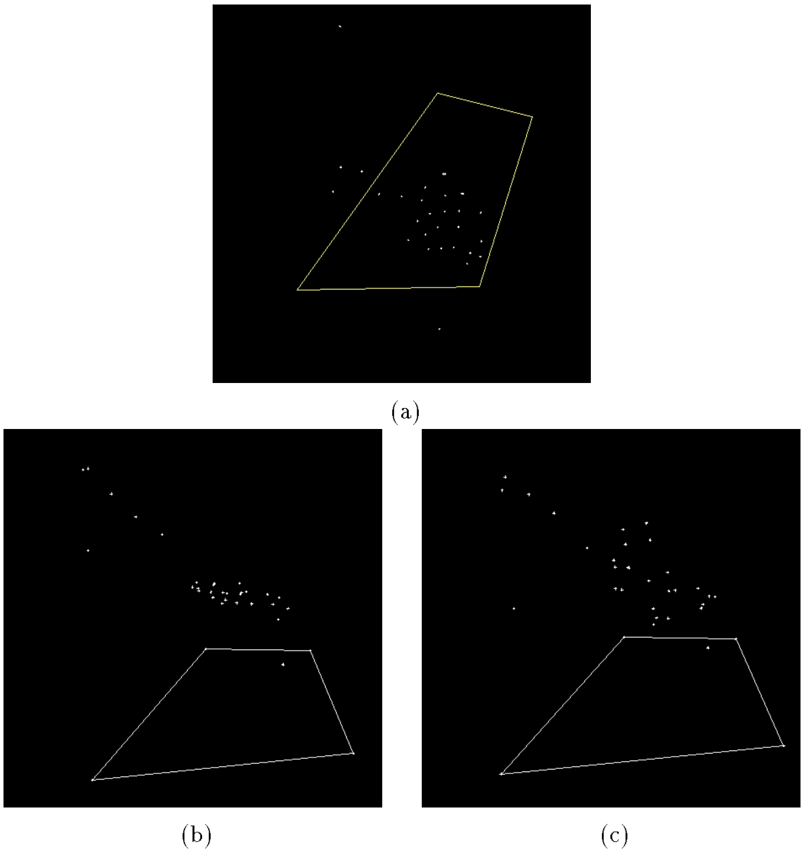


Figure 3: Results of 3D reconstruction of the collection of sample points. (a) Frontal view (aligned with the first frame of the sneaker). The two bottom displays show a side view of the sample. (b) Result of recovering structure between the first and tenth frame (large base-line); (c) Result of recovery between the first and second frames (small base-line).

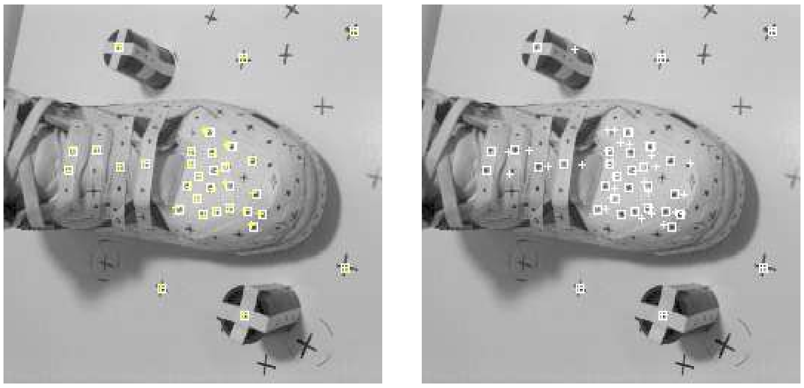
4.3 Application III: Image Coding

The re-projection paradigm can serve as a principle for image coding. In a sender/receiver mode, the sender computes the relative affine structure between two extreme views of a sequence, and sends the first view, the relative affine scalars, and the homographies and epipoles between the first frame and all the intermediate frames. The intermediate frames can be reconstructed by re-projection.

Relative affine structure of the 34 sample points was computed between the first and last frame of the ten frame sequence (displays (a) and (d) in Fig. 2). Display (a) in Fig. 6 shows a graph of the average re-projection error for all the intermediate frames (from second to ninth frames). Display (b) shows the relative error normalized by the distance between corresponding points across the sequence. We see that the relative error generally goes down as the re-projected frame is farther from the first frame (increase of base-line). In all frames, the average error is less than 1 pixel, indicating a relatively robust performance in practice.

5 Summary

We gave an algebraic description for relative affine structure. 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. We have also shown the equation of relative affine motion (1) can be projected onto two unique directions such that on one direction we obtain a decomposition of the Fundamental matrix as a product of a homography, due to an arbitrary plane, and the skew matrix of the epipole. The second direction yields a projection of the relative affine equation onto a simple equation of structure giving rise to a least squares solution. A corollary of these results gives rise to a extremely simple method for capturing the redundancy in recovering structure from a stream of views. This work can be easily extended to recovery of relative affine structure without use of explicit correspondences (see [13]).



(a) (b)

Figure 4: Results of re-projection onto the tenth frame. Epipoles were recovered using the ground plane homography (see text). The re-projected points are marked by crosses, and should be in the center of their corresponding square for accurate re-projection. (a) Structure was recovered between the first and fifth frames, then re-projected onto the tenth frame (large base-line). Average error is 1.1 pixels with std of 0.98. (b) Structure was recovered between the first and second frames (small base-line situation) and then re-projected onto the tenth frame. Average error is 7.81 pixels with std of 6.5.

6 Appendix: Homography from Line Correspondences

In many application it is more robust to use line features instead of points. In the case of lines, the relation between the Plücker coordinates of a line in the first and the second camera, $(\mathbf{N},\mathbf{l})$ and $(\mathbf{N}',\mathbf{l}')$, is expressed by:

$$\begin{bmatrix} \mathbf{l}' \\ \mathbf{N}' \end{bmatrix} = \mathbf{D} \begin{bmatrix} \mathbf{l} \\ \mathbf{N} \end{bmatrix}$$

where the matrix $\mathbf{D}$ is defined as [11]:

$$\mathbf{D} = \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{E} & \mathbf{R} \end{bmatrix}$$

If a line $D : (\mathbf{N},\mathbf{l})$ belongs to the plane $\pi : (\mathbf{n}_\pi, 1)$, then [12]:

$$\mathbf{n}_\pi \wedge \mathbf{N} = \mathbf{l}$$

We can then write:

$$\begin{aligned} \mathbf{N}' &= \mathbf{RN} + \mathbf{El} \\ &= \mathbf{RN} + \mathbf{E}\tilde{\mathbf{n}}_\pi\mathbf{N} \\ &= (\mathbf{R} + \mathbf{E}\tilde{\mathbf{n}}_\pi)\mathbf{N} \end{aligned}$$

Therefore if $\mathbf{n}$ and $\mathbf{n}'$ are the images of a line lying in the plane π, we have:

$$\mathbf{A}_l\mathbf{n} \cong \mathbf{n}' \quad (7)$$

where $\mathbf{A}_l = \mathbf{R} + \mathbf{E}\tilde{\mathbf{n}}_\pi$.

It is easy to verify the follwoing realtionship between homographic transformation of points $\mathbf{A}$, and lines $\mathbf{A}_l$ due to the plane π:

$$\mathbf{A}_l = \mathbf{A}^* = (\mathbf{A}^t)^{-1} \quad (8)$$

For uncalibrated cameras (unknown intrinsic parameters), where M, M' are the internal camera parameters (the transformation from camera coordinate systems to the observed coordinate systems), the homographic transformations $\mathbf{A}$ and $\mathbf{A}_1$ can be expressed as:

$$\mathbf{A} = \mathbf{M}(\mathbf{R} + \mathbf{T}\mathbf{n}^t)\mathbf{M}^{-1}$$

and

$$\mathbf{A}_l = \mathbf{M}^*(\mathbf{R} + \mathbf{E}\tilde{\mathbf{n}})\mathbf{M}^{-1^*}$$

Acknowledgments

N. Navab would like to thank Prof. Aaron Bobick for his support and helpful comments.

References

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

[2] T. Moons L. Van Gool M. Van Diest and E. Pauwels. Affine reconstruction from perspective image pairs. October 1993.

[3] O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In Giulio Sandini, editor, *Proc. Second European Conf. on Comput. Vision*, pages 563–578. Springer-Verlag, Lecture Notes in Computer Science 588, May 1992.

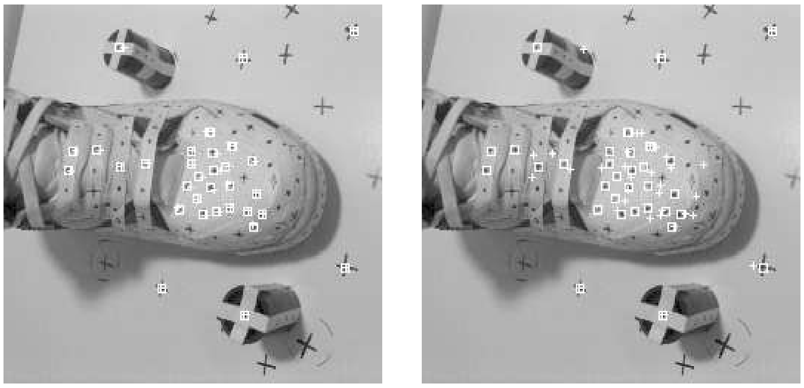
[4] O.D. Faugeras and F. Lustman. Let us suppose that the world is piecewise planar. In O. D. Faugeras and Georges Giralt, editors, *Proc. Int'l Symposium Robotics Res.*, pages 33–40. MIT Press, Cambridge, MA, 1986.

[5] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, pages 761–764, Urbana Champaign, IL., June 1992.

[6] D.P. Huttenlocher and S. Ullman. Recognizing solid objects by alignment with an image. *Int'l J. Comput. Vision*, 5(2):195–212, 1990.

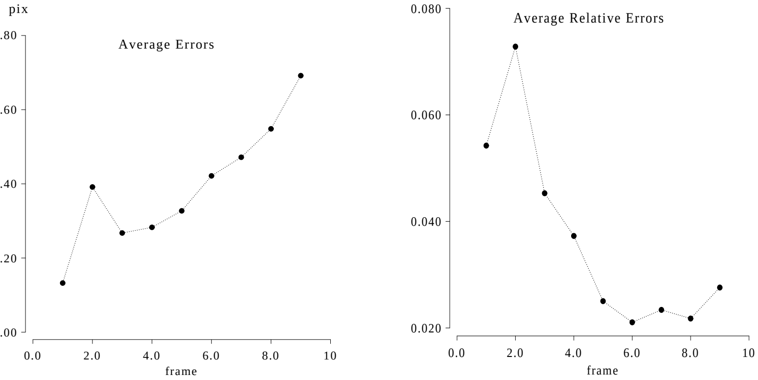
[7] J.J. Koenderink and A.J. Van Doorn. Affine structure from motion. *Journal of the Optical Society of America A*, 8:377–385, 1991.

[8] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. Int'l Joint Conf. Artif. Intell.*, pages 674–679, Vancouver, 1981.



(a) (b)

Figure 5: Re-projection onto the tenth frame. Epipoles are computed via fundamental matrix (see text) using the implementation of [9]. (a) Large base situation (structure computed between first and fifth frames): average error 1.62 with std of 1.2. (b) Small base-line situation (structure computed between first and second frames): average error 13.87 with std of 9.47.



(a) (b)

Figure 6: Error in re-projection onto the intermediate frames (2–9). Structure was computed between frames one and ten. (a) average error in pixels, (b) relative error normalized by the displacement between corresponding points.

[9] 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.

[10] R. Mohr, F. Veillon, and L. Quan. Relative 3d reconstruction using multiple uncalibrated images. pages 543–548, June 1993.

[11] N. Navab. *Visual motion of lines, and Cooperation between motion and stereo.* Dissertation, University of Paris XI, INRIA Sophia-Antipolis, France, January 1993. in English.

[12] N. Navab, O. D. Faugeras, and T. Vieville. The critical sets of lines for camera displacement estimation:a mixed euclidean-projective and constructive approach. In *Proc. Fourth Int'l Conf. Comput. Vision*, Berlin, Germany, May 1993. IEEE.

[13] N. Navab and S. Mann. Recovery of relative affine structure using the motion flow field of a rigid planar patch. In *Mustererkennung 1994, Tagungsband*, pages 186–196.

[14] L. Quan. Affine stereo calibration for relative affine shape reconstruction. In *Proc. British Machine Vision Conf.*, pages 659–668, September 1993.

[15] L. Robert and O.D. Faugeras. Relative 3D positioning and 3D convex hull computation from a weakly calibrated stereo pair. In *Proc. Fourth Int'l Conf. Comput. Vision*, pages 540–544, Berlin, Germany, May 1993. IEEE.

[16] 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.

[17] A. Shashua. Projective depth: A geometric invariant for 3D reconstruction from two perspective/orthographic views and for visual recognition. *Proc. Fourth Int'l Conf. Comput. Vision*, May 1993.

[18] R. Tsai and T.S. Huang. Estimating three-dimensional motion parameters of a rigid planar patch, II: singular value decomposition. *IEEE Trans. ASSP*, 30, 1982.