Relative Affine Structure: Canonical Model for 3D From 2D Geometry and Applications

Amnon Shashua and Nassir Navab

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, and introduces new, extremely simple, algorithms for the tasks of reconstruction from multiple views, recognition by alignment, and certain image coding applications.

Index Terms—Structure from motion, visual recognition, alignment, reprojection, projective reometry, algebraic and geometric invariants.

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 in the recognition process.

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 in the recent few years. The emerging realization is that nonmetric 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.

In this paper, we propose a unified framework that includes by generalization and specialization the Euclidean, projective and affine frameworks. The framework, we call "relative affine," gives rise to an equation that captures most of the spectrum of previous results related to 3D-from-2D geometry, and introduces new, extremely simple, algorithms for the tasks of reconstruction from multiple views, recognition by alignment, and certain image coding applications. For example, previous results in these areas—such as affine structure from orthographic views, projective structure from perspective views, the use of the plane at infinity for reconstruction (obtaining affine structure from perspective views), epipolar-geometry related results—are often reduced to a single-line proof under the new framework (see Corollaries 1 to 4).

The basic idea is to choose a representation of projective space in which an arbitrarily chosen reference plane becomes the plane at infinity. We then show 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 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. Parts of this work, as it evolved, have been presented in the meetings found in [35], [38], and in [28].

2 Related Work

The introduction of nonmetric reconstruction from a set of cameras was first introduced, under the special case of parallel projection, by [18] and later work on various aspects of this framework can be found in [44], [45], [34], [17], [33].

The extension to the full projective framework was later introduced by [39], [9], [13], [36] (see [9] for additional review) with a sample of other relevant work on the topic in [10], [31], [25], [24], [41], [2].

Independently of us [35], [38], Kumar and Anandan [19], Sparr [40], and Sawhney [32] have developed the theory of representing projective space with reference to a planar

• *A. Shashua is with the Computer Science Department, Technion, Haifa 32000, Israel. E-mail: shashua@cs.technion.ac.il.*
• *N. Navab is with Siemens Corporate Research, Princeton, NJ 08540. E-mail: navab@scr.siemens.com.*

Manuscript received June 20, 1994; revised May 1, 1996. Recommended for acceptance by Y. Shirai.
For information on obtaining reprints of this article, please send e-mail to: transpami@computer.org, and reference IEEECS Log Number P96048.

surface by a parallax term, also coined by Kumar and Anandan as "plane + parallax."

3 NOTATION

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 $\psi \subset \mathcal{P}^2$ denote a view (arbitrary) of the object. Given two views ψ and ψ' 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. A set of numbers defined up to scale are enclosed by brackets, a set of numbers enclosed by parentheses define a vector in the usual way. 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 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.

When only two views ψ, ψ' are discussed, then points in ψ are denoted by p, their corresponding points in ψ' are denoted by p', and the epipoles are $v \in \psi$ and $v' \in \psi'$. 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.

A camera coordinate system is an Euclidean frame describing the actual internal geometry of the camera (position of the image plane relative to the camera center). If $p = (x, y, 1)^\top$ is a point in the projective coordinate representation, then $M^{-1}p$ represents the camera coordinates, where M is an upper-diagonal matrix containing the internal parameters of the camera. When M is known the camera is said to be internally calibrated. The material presented in this paper does not require further details of internal calibration—such as its decomposition into the components of principle point, image plane aspect ratios and skew—only the mere existence of M is required for the remainder of this paper. We will refer to the coordinates $p = [x, y, 1]$ as projective image coordinates, whereas $M^{-1}p$ will be referred to as camera coordinates.

4 PRELIMINARIES

The classic equation of camera motion, taking into consideration the internal parameters of the cameras, can be described as follows. Let z and z' be the depth of a point P with respect to the first and second camera positions. Let R, T be the rotational and translational components of camera displacement ($R \in O_3^+$), and let M and M' describe the internal camera parameters of the first and second cameras, respectively. Since $zM^{-1}p$ represents the space coordinates of P with respect to the first camera frame, and $z'M'^{-1}p'$ are the coordinates of P with respect to the second camera frame, we obtain the following equation:

$$z'M'^{-1}p' = zRM^{-1}p + T,$$

or

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

Often it is assumed that $M = M'$, and, furthermore, that $M = I$, otherwise M, M' are recovered off-line during a calibration procedure. The location of the epipole v' in the second image plane is defined (up to scale) by $M'T$ (which in a calibrated frame is simply the translational component of camera motion), and the epipole v in the first image plane is defined by $MR^\top T$.

Part of what will be shown later in the paper is that R, M, M' can be replaced by a single entity, a projective transformation (homography) due to some arbitrary plane captured by three point matches across the two views, and z is correspondingly replaced by a "relative affine" invariant. We now describe the concept of a homography and its connection to 3D-from-2D geometry.

Assume a coplanar configuration of points in space projecting onto corresponding points p_i, p_i', $i = 1, ..., n$ $(n \geq 4)$ across the two views. The fundamental theorem of plane projectivity states that the entire mapping from one view to the other is determined by four of the corresponding points (assuming a general configuration of the four points on the plane). The mapping is a projective transformation, a homography, described by a matrix $A \in PGL_3$. In other words, $Ap_i \cong p_i'$, $i = 1, ..., n$, and since A has eight parameters (set $A_{33} = 1$, for example), and each point match contributes two linear equations, we can recover A linearly using four point matches.

In general, however, we would like to work with 3D point configurations rather than coplanar sets, and since the homography A will play a crucial role in the 3D-from-2D geometry, we must first find a general way to get around the problem of recovering A from a coplanar set of four points (in a 3D configuration, it is unlikely that four point matches are coming from a coplanar set in space). The following proposition is the result of a simple observation that the epipoles can be used instead of a fourth pair of corresponding points, for establishing the homography due to any arbitrary plane in space.

PROPOSITION 1. *A projective transformation, A, which is determined from three arbitrary, noncollinear, corresponding points and the corresponding epipoles, is a projective transformation of the plane passing through the three object points which project onto the corresponding image points.*

The proof is straightforward and can be found in [36]. Given the epipoles, therefore, we need just three points to determine the correspondences of all other points coplanar with the plane passing through the three corresponding object points. Note that in a 3D configuration of points, the plane is a virtual plane, i.e., may not correspond to any physical planar facet of the object. We will address later how the epipoles can be obtained, and for now we assume they are given.

The concept of the "relative affine" framework described below is based on, first, connecting the matrix A (determined by three arbitrary point matches), the epipole v' (translational component of motion) and a structure parameter (Theorem 1). Second, relating A to the parameters

R, T, M, M' and the parameters of the virtual plane determined by the three point matches (Theorem 2). Third, unifying by generalization and specialization of the two theorems most of the previous and current results related to 3D-from-2D geometry.

5 Relative Affine Structure

For the sake of clarity, we describe the general result as one that naturally extends from affine structure under parallel projection. Koenderink and Van-Doorn [18] have described the 3D affine geometry resulting from two views obtained by parallel projection from the 3D object onto the 2D image plane. Their approach leads to a "plane + parallax" paradigm, described as follows.

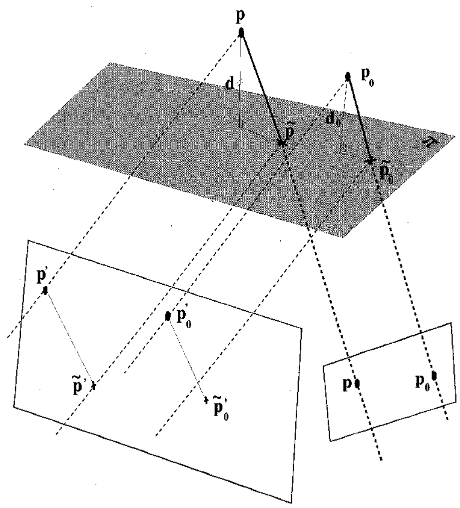


Fig. 1. Affine structure under parallel projection is d/d_o. This can be seen from the similarity of trapezoids followed by the similarity of triangles: $\frac{p'-\tilde{p}'}{p'_o-\tilde{p}'_o} = \frac{P'-\tilde{P}}{P_o-\tilde{P}_o} = \frac{d}{d_o}$.

Consider Fig. 1, and assume that three arbitrary (non-collinear) matching points were identified in both views ψ and ψ', and let the plane π be the plane defined by the three corresponding object points in 3D. The triplet of matching image points define a unique affine transformation A (a 2×3 matrix) mapping all points of π from image ψ to image ψ'. For any point $p = (x, y) \in \psi$ let $\tilde{p}' = Ap$. Let P be some 3D point not lying on π and let $\tilde{P}$ be at the intersection of the line of sight (the projecting ray) with π. Similarly, let $P_o \notin \pi$ and $\tilde{P}_o$ be some fixed point and its projection onto π, respectively. One can clearly see that the trapezoids $P, \tilde{P}, p', \tilde{p}'$ and $P_o, \tilde{P}_o, p'_o, \tilde{p}'_o$ are similar. Hence the ratio,

$$\frac{P' - \tilde{P}}{P_o - \tilde{P}_o}$$

is measurable from the image planes and is equal to,

$$\frac{p' - \tilde{p}'}{p'_o - \tilde{p}'_o},$$

which, in turn, is equal to (by similarity of triangles) to the ratio d/d_o of the perpendicular distances of P and P_o from the plane π. In other words, an affine invariant in space whose value is equal to d/d_o is recovered from image measurements alone—three matching points to recover the affine transformation associated with some plane, call it the "reference plane" and a fourth point P_o to set up a mutual scale.

In the general perspective case, a projective invariant (or projective coordinates) can be measured from two views, but instead of a basis of four points, one would need five [9], [13], [36].

In what follows, we approach the general perspective case via an intermediate structure representation, referred to as a "relative affine structure," which is a natural extension of the affine construction we have seen above. We will continue to use the language of a homography matrix (which is the generalization of a 2D affine transformation) and a parallax term from some reference plane, but now the results will hold for perspective views. This approach has the advantage of unifying the affine and projective cases in a very simple an intuitive way, and produces a simple method for computing invariants from perspective views.

Theorem 1 (Relative Affine Structure). *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, ψ', respectively. Let $p_o \in \psi$ and $p'_o \in \psi'$ 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 $p \in \psi$ and $p' \in \psi'$, we have*

$$\boxed{p' \cong Ap + k\,v'} \qquad (1)$$

The coefficient $k = k(p)$ is independent of ψ', 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 and ψ', respectively, and let their coordinates be (0, 0, 0, 1), (1, 1, 1, 1), respectively (see Fig. 2). 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 $v \in \psi$ and $v' \in \psi'$, 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

$$\mu \begin{pmatrix} \alpha' \\ \beta' \\ \gamma' \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} + k \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Multiply both sides by A_2^{-1} to obtain $\mu p' = Ap + kv'$, 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 $p'_j \cong Ap_j$, $j = 1, 2, 3$, 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 [36], [31]. 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 $p'_o \cong Ap_o + v'$. 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. □

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 $\mathcal{R}_0$ 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 $\mathcal{R}_0$ 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 $\mathcal{R}_0$ 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 $\mathcal{R}_0$. 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 $\mathcal{R}_0$ we named it "relative affine structure"; this should not be confused with the term "relative invariants" used in classical invariant theory (invariants multiplied by a power of the transformation determinant, as opposed to "absolute invariants").

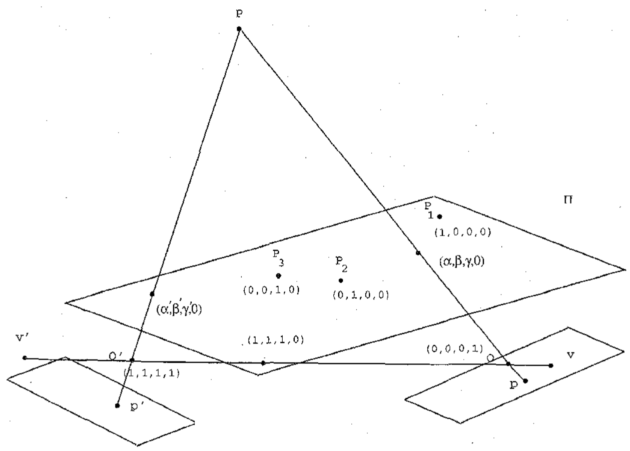


Fig. 2. See proof of Theorem 1.

In practical terms, the difference between a full projective framework (like in [9], [13], [36]) 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 and an arbitrary view ψ_i and indexed by a fixed set of four corresponding points, then $f_o(\psi, \psi_i)$ is fixed for all i.

With comparison with the affine construction described in Fig. 1, we see that the same ingredients are used. A homography matrix of some plane π replaces the 2D affine transformation and a parallax term k replaces d/d_o, and the fixed point P_o is used to set a mutual scale. The difference is that the invariance is with respect to the camera parameters of view ψ' alone, and in order to compute the homography matrix we need in addition the epipolar points. The number of image matching points we use for a basis remains four. We will see the relationship to the projective invariance later in this section.

To complete the geometric picture we derive next the value of the parallax term k as a function of distances from the reference plane π.

THEOREM 2. *Let π denote a reference plane and $P_o \notin \pi$ be some fixed point, and P be a point whose corresponding points p, p' in both views satisfy (1). Let d, d_o denote the perpendicular distances of P and P_o from π, and let z, z_o denote the distances of P and P_o from the first camera center O. Then the parallax term k is:*

$$k = \frac{z_o}{z} \frac{d}{d_o}.$$

PROOF. Let M, M' be the transformation matrices from camera coordinate systems to standard coordinates of first and second camera, respectively. Namely, $zM^{-1}p$ and $z'M'^{-1}p'$ are the Euclidean coordinates of the point P in the first and second camera frames. Let R, T be the rotational and translational components of the coordinate change between the two cameras, thus we have:

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

In the case $P \in \pi$, then $n^\top(zM^{-1}p) = d_\pi$ where n (normal vector) and d_π (scalar) are the parameters of π in the first camera coordinate system. We obtain,

$$p' \cong M'RM^{-1}p + \frac{1}{z}\frac{1}{d_\pi}M'T\left(zn^\top M^{-1}p\right)$$
$$= M'\left(R + \frac{1}{d_\pi}Tn^\top\right)M^{-1}p = Ap.$$

In other words, the homography matrix associated with π is

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

which is the generalization of the classical motion of planes in the calibrated case [11], [42].

Let $v' = M'T$, then for a general point P we have:

$$p' \cong M'RM^{-1}p + \frac{1}{z}v'$$
$$= Ap - \frac{1}{d_\pi}v'n^\top M^{-1}p + \frac{1}{z}v'$$
$$= Ap + \left[\frac{1}{z} - \frac{n^\top\left(M^{-1}p\right)}{d_\pi}\right]v'$$
$$= Ap + \left[\frac{d_\pi - n^\top\left(zM^{-1}p\right)}{zd_\pi}\right]v'$$
$$= Ap + \left[\frac{d}{zd_\pi}\right]v',$$

where $d = d_\pi - n^\top(zM^{-1}p)$ the (perpendicular) signed distance from P to π. Since the value of the parallax term k in (1) is normalized by the parallax term of P_o (by appropriately scaling v'), then d_π drops out and we are left with,

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

where d_o is the (perpendicular) signed distance of P_o from π (see Fig. 3a). □

With the corollaries below we put the relative affine framework within the familiar context of affine structure under parallel and perspective projections, Euclidean structure and projective structure.

COROLLARY 1. Relative affine structure k approaches affine structure under parallel projection when O goes to infinity, i.e., $k \to \frac{d}{d_o}$ when $O \to \infty$.

PROOF. When $O \to \infty$, then $z, z_o \to \infty$, and $\frac{z_o}{z}$ tends to 1. Thus $k = \frac{z_o}{z}\frac{d}{d_o} \to \frac{d}{d_o}$ (see Fig. 1). □

COROLLARY 2. *When the plane π is at infinity (with respect to the camera coordinate frame), 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 $d, d_o \to \infty$, and $\frac{d}{d_o}$ tends to 1. Thus $k = \frac{z_o}{z}\frac{d}{d_o} \to \frac{z_o}{z}$. Also, $d_\pi \to \infty$, thus $A \to M'RM^{-1}$ (see Fig. 3b). □

In Theorem 2, the homography A due to the plane π is described (by (2)) as a product of the rigid camera motion parameters, the parameters of π, and the internal camera parameters of both cameras. This result is a natural extension of the classical motion of planes found in [11], [42], and also in [24]. Equation (2) also implies that any homography matrix A can be described by some other homography matrix, the epipole and four parameters:

$$A = A_\pi + \frac{1}{d}v'n^\top,$$

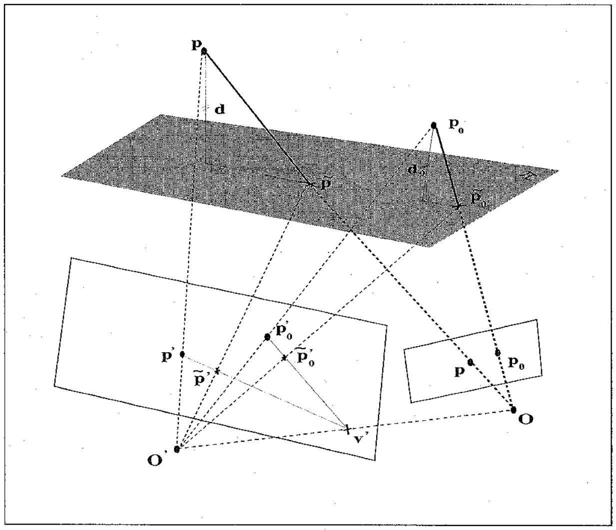
where A_π is a homography matrix associated with some plane π, and n, d are the parameters describing the plane associated with the homography A. A similar form of this equation was shown also in [14].

The relative affine structure k is described as a product of the affine structure under parallel projection (d/d_o) 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[1] 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 we saw earlier in this section. 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.

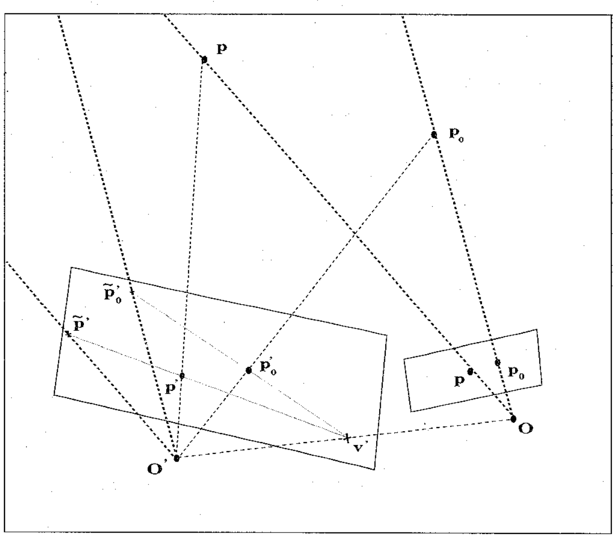
Another specialization of relative affine structure was shown in Corollary 2 by considering the case when π is at infinity with respect to our Euclidean frame (i.e., really at infinity). In that case k is simply inverse depth (up to a uniform scale factor), and the homography A is the familiar rotational component of camera motion (orthogonal matrix R) in the case of calibrated cameras, or a product of R with the internal calibration parameters. In other words, when π is at infinity with respect to our camera coordinate frame, then relative affine becomes affine (the plane at infinity is preserved under all representations). Notice that the rays towards the plane at infinity are parallel across the two cameras (see Fig. 3b). Thus, there exists a rotation matrix that aligns the two bundles of rays, and following this line of argument, the same rotation matrix aligns the epipolar lines (scaled appropriately) because orthogonal matrices commute with cross products. We have therefore the algo-

1. Note that the distance can be measured along any fixed direction. We use the perpendicular distance because it is the most natural way of describing the distance between a point and a plane.

rithm of [21] for determining the rotational component of standard calibrated camera motion, given the epipoles. 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 [30], or the camera motion is purely translational [26]. Thus in the general case, we can realize either the relative affine framework or the projective framework.



(a)



(b)

Fig. 3. (a) Relative affine Structure: $k = \frac{z_o}{z}\frac{d}{d_o}$. (b) 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.

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, $\hat{\pi}$, 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 and $k_{\hat{\pi}}$ be the relative affine structures with respect to planes π and $\hat{\pi}$, respectively. From Theorem 2 we have that $k_\pi = \frac{z_o}{z}\frac{d}{d_o}$ and $k_{\hat{\pi}} = \frac{z_o}{z}\frac{\hat{d}}{\hat{d}_o}$. The ratio $k_\pi / k_{\hat{\pi}}$ removes the dependence on the projection center O (z/z_o cancels out) and is therefore a projective invariant (see Fig. 4). This projective invariant is also the ratio of cross-ratios of the rays $\overline{OP}$ and $\overline{OP_o}$ with their intersections with the two planes π and $\hat{\pi}$, which was introduced in [36] as "projective depth." It is also the ratio of two affine structures under parallel projection (recall that d/d_o is the affine structure; see Fig. 1). □

The connection between the relative affine structure and projective structure is shown in the corollary above. Projective invariants are necessarily described with reference to five scene points [9], or equivalently, with reference to two planes and a point laying outside of them both [36]. Corollary 3 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 [36], which is the ratio of the perpendicular distance of a point to two planes (up to a uniform scale factor).

COROLLARY 4. *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 [8], is unique and does not depend on the plane π. Furthermore, $Fv = 0$ and $F^T v' = 0$.*

PROOF. Let $p \in \psi$, $p' \in \psi'$ 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 $p'^T(v' \times Ap) = 0$, or equivalently: $p'^T[v']Ap = 0$, or $p'^T Fp = 0$, 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 (2) and noting that $[v']T = 0$, thus $F = [v']M'RM^{-1}$. Finally, since $Av \cong v'$, $[v']Av \cong [v']v' = 0$, thus $Fv = 0$, and $A^T[v']^T v' = -A^T[v']v' = 0$, thus $F^T v' = 0$. □

Corollary 4 unifies previous results on the nature of what is known by now as the "fundamental matrix" [9], [8], [10]. 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 [21]. 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 an element that vanishes as a product with $[v']$—an observation that was made previously by [15]. Thirdly, the "essential" matrix [22], $E = [v']R$, 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 $M = M' = I$.

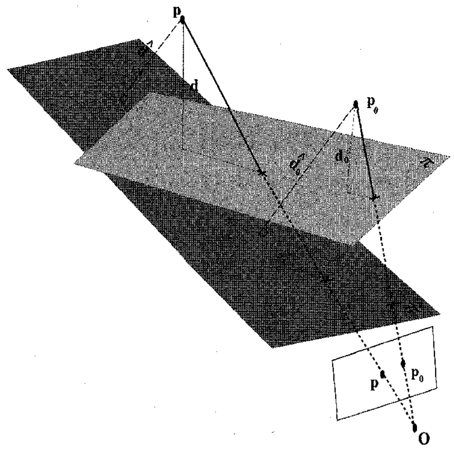


Fig. 4. Projective-depth [36] is the ratio of two relative affine structures, each with respect to a distinct reference plane, which is also the ratio of two affine structures (see Corollary 3 for more details).

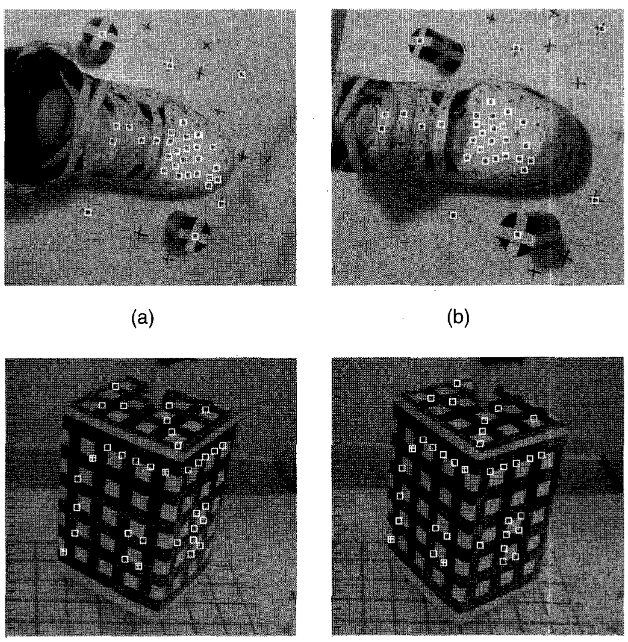


Fig. 5. (a, b): Views, out of a sequence of ten views, of a sneaker. The frames shown here are the first, and tenth of the sequence. The overlayed squares mark the corresponding points that were tracked and subsequently used for our experiments. (c, d): The first and last views of the sequence of six views of a cube. These were the pair of views used for our second set of reconstruction results. Corresponding points are marked by squares, and points on the reference plane are marked by overlaid crosses.

5.1 Application I: Reconstruction

Taken together, the results above demonstrate the ability to compute relative affine structure using many points over two views (more than two views can be accommodated easily in this framework, but we do not address this here). 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_i, and p'_i $i = 0, ..., n$ denote the ith image point on frame ψ and ψ'. Let A denote the homography from frame ψ to frame ψ', v, v' the corresponding epipoles such that $Av \cong v'$, and let k_i denote the relative affine structure of point i. We follow these steps:

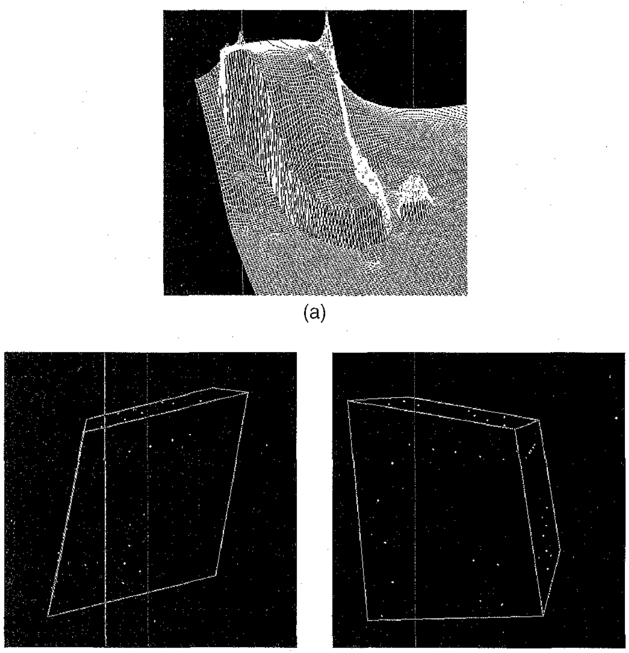
1) Compute epipoles v, v' using the relation $p_i'^{T} F p_i = 0$, over all i. Eight corresponding points are needed for a linear solution, and a least-squares solution is possible if more points are available. In practice the best results were obtained using the non-linear algorithm of [6]. The epipoles follow by $F_j v_j = 0$ and $F^T v'_j = 0$ [8]. The latter readily follows from Corollary 4 as $[v'_j]A_j v_j \cong [v'_j]v'_j = 0$ and $A_j^T [v'_j]^T v'_j = -A_j^T [v'_j] v'_j = 0$.
2) Compute A from the equations $Ap_i \cong p'_i$, $i = 1, 2, 3$, and $Av \cong v'$. This leads to a linear set of eight equations for solving for A up to a scale. A least squares solution is available from the equation $p'_i[v']Ap_i = 0$ for all additional points (Corollary 4). Scale A to satisfy $p'_o \cong Ap_o + v'$.
3) Relative affine structure k_i is given by $p'_i \cong Ap_i + k_i v'$.
4) The projective coordinates of the scene are given by $[x_i, y_i, 1, k_i]$ and the camera transformation matrix is $[A, v']$.

5.2 Application II: Recognition by Alignment

The relative affine invariance relation, captured by Theorem 1, can be used for visual recognition by alignment ([43], [16], and references therein). In other words, the invariance of k can be used to "reproject" the object onto any third view p'', as follows. Given two "model" views in full correspondence $p_i \leftrightarrow p'_i$, $i = 1, ..., n$, 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 $p'' \cong Bp + kv''$, 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 $\hat{p}''_i \cong Bp_i + k_i v''$. 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. [27], [5], [12]), or by using projective structure instead of the relative affine structure [36]. 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 [37].

5.3 Application III: Image Coding

The reprojection paradigm, described in the previous section, can serve as a principle for model-based image compression. 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 reprojection. Alternatively, the sender send the two extreme views and the homographies and epipoles between the first and all other intermediate views. The receiver recovers the correspondence field between the two extreme views, and then synthesizes the remaining views from the received parameters of homographies and epipoles. In case the distance between the two extreme views is "moderate," we found that optical flow techniques can be useful for the stage of obtaining the correspondence field between the views. Experiments can be found later in the text, and more detailed experiments concerning the use of optical flow in full registration of images for purposes of model-based image compression can be found in [4].



(a) (b) (c)

Fig. 6. Results of 3D reconstruction of the collection of sample points. (a) The surface of the shoe reconstructed (relative affine) from around 100 points followed by a membrane interpolation (b), and (c) Two extreme views of the reconstructed (relative affine) cube. Note that the cube appears stretched because the reconstruction is up to a projective transformation from the Euclidean structure.

6 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 pin-hole camera model (radial distortions, decentering, and other effects), and due to errors in obtaining image correspondences.

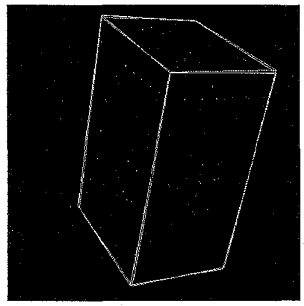


Fig. 7. Graphic comparison between the true and reconstructed structure (see text).

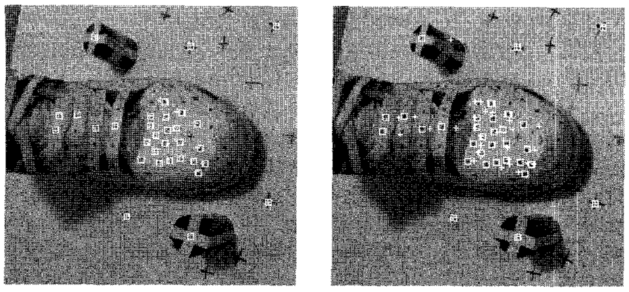
We have selected two objects for experiments. The first object is a sneaker (shown in Fig. 5a, b) with added texture to facilitate the correspondence process. The sneaker is a challenging object because of its complexity, i.e., it has a shape of a natural object and cannot easily be described parametrically (as a collection of planes or algebraic surfaces). However, we do not have ground-truth data of the sneaker, hence the performance of the reprojection and image coding tasks can be appreciated at a quantitative level, whereas the performance of the reconstruction method can be appreciated only at a qualitative level.

The second object is a cube (shown in Fig. 5c, d). The sequence, with correspondences and true 3D data, is available from the computer vision group at the University of Massachusetts at Amherst. Hence, one can measure the performance of reconstruction at a quantitative level. Reconstruction results on this set by other approaches can be found in [7], [29], [20], [45], [1].

On the sneaker images, a set of 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. 5). The correspondence process is based on an implementation of a coarse-to-fine optical-flow algorithm based on [23] and described in [3]. The surface was reconstructed with respect to the ground plane (that served as a reference plane) and transformed to a camera coordinate system by assuming that the ground plane is parallel to the image plane (it is actually not). Fig. 6a shows the reconstructed surface followed by a membrane interpolation. For better surface fitting, around 100 matching points were used. The reconstructed surface appears faithful to the true structure of the sneaker.

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 [6]. The two methods gave rise to very similar results for reconstruction, and slightly different results for reprojection (see later).

For the cube object, a face of the cube was selected as a reference plane (the face marked by overlaid crosses). The homography A was recovered using four coplanar points or from three corresponding points and the epipoles



(a) (b)

Fig. 8. Results of reprojection 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 reprojection. (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.

(Proposition 1). In the latter case, the epipoles were recovered by the implementation described in [6].

Relative affine structure was recovered using the two views displayed in Fig. 5c, d. For purposes of display of the relative affine reconstructed 3D points of the cube, we assumed that $M = M' = I$ and that the reference plane is parallel to the image plane, and then with a simple calculation transformed k into z (depth). Since these assumptions are not completely met, the reconstructed cube is slightly stretched and sheared (the reconstruction cannot be worse than some projective transformation of the cube), as seen in Fig. 6b, c. The reconstructed points are shown from extreme views from which errors in reconstruction would clearly show up in display. As can be seen, the points are situated reasonably well on the three faces of the cube. The lines in the display were added for convenience and were computed by the intersection of the least-squares approximation to the three reconstructed faces of the cube.

For a more quantitative evaluation of accuracy we recovered the 3D projective transformation that maps $[x, y, 1, k]$ to $[X, Y, Z, 1]$. In general this can be done by having at least five ground points (points whose X, Y, Z coordinates are known), or by employing constraints such as coplanarity, perpendicular lines and planes in the scene, and known distances between points in the scene (the latter constraint is non-linear). Since the true 3D data is provided with the sequence we chose the former method. To compare our results with those reported in [7], [29], [20], [45], [1], the level of error was calculated as the ratio of the average mean square error in depth with the overall average depth of the cube. The average depth values of the sample points on the cube is 626.48 units. The average error we found between the the given depth values and the reconstructed values is 1.43 (std 1.99), which provides a 0.2% level of relative error. The range of error levels reported in [7], [29], [20], [45], [1] was 0.2-0.7%. In other words, the accuracy level we obtained with the given data is on average 1/500 of the extent of the object from the camera. It should be noted that these measures of performance are not complete because the level of accuracy on obtaining the corresponding feature points was not provided with the data.

Finally, a graphic comparison between the true and reconstructed structure of the cube is shown in Fig. 7 by displaying a view of the cube in which both the true and reconstructed structure are overlaid.

In the reprojection application (see Section 5.2), 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.

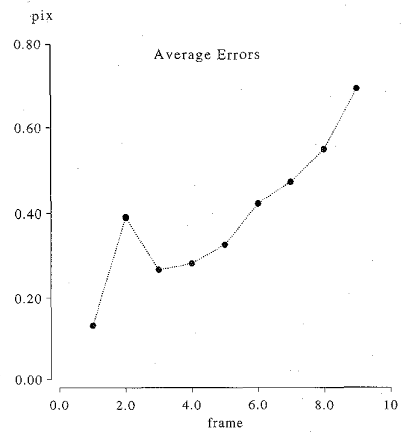
In general, the performance was better when the ground plane was used for recovering the epipoles (i.e., first the homography matrix due to the ground plane was recovered, and then used to recover the epipoles, rather than first recovering the fundamental matrix). When the intermediate view was the fifth in the sequence (Fig. 5c), the average error in reprojection was 1.1 pixels (with standard deviation of 0.98 pixel). When the intermediate view was the second frame in the sequence (Fig. 5b), 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. 8. The re-projected points are represented by crosses overlayed on the last frame (the re-projected view).

When the epipoles were recovered directly from the fundamental matrix, 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). 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 [37].

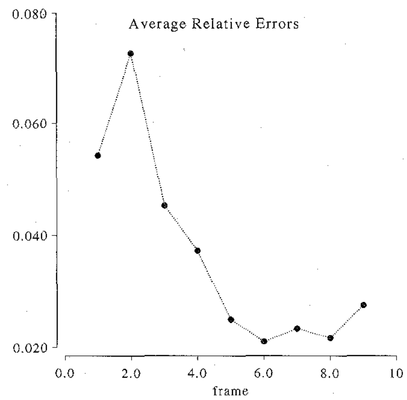
For the image coding paradigm (see Section 5.3), relative affine structure of the 34 sample points were computed between the first and last frame of the 10 frame sequence (Fig. 5a, d). Fig. 5a shows a graph of the average reprojection error for all the intermediate frames (from second to ninth frames). Fig. 5b 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 reprojected frame is farther from the first frame (increase of base-line). In all frames, the average error is less than one pixel, indicating a relatively robust performance in practice.

7 SUMMARY

The framework of "relative affine" was introduced and shown to be general and sharper (in the sense of requiring four rather than five basis points) than the projective results for purposes of 3D reconstruction from two views and for the task of recognition by alignment. One of the key ideas



(a)



(b)

Fig. 9. Error in reprojection 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.

in this work is to define and recover an invariant that stands in the middle ground between affine and projective. The middle ground is achieved by having the camera center of one arbitrary view as part of the projective reference frame (of five points), thus obtaining the first result described in Theorem 1 (originally in [35]). The result simply states that under general uncalibrated camera motion, the sharpest result we can obtain is that all the degrees of freedom are captured by four points (thus the scene may undergo at most 3D affine transformations) and a single unknown projective transformation (from the arbitrary viewer-centered representation $\mathcal{R}_0$ to the camera coordinate frame). The invariants that are obtained in this way are viewer-centered since the camera center is part of the reference frame and are called "relative affine structure." This statement, that all the available degrees of freedom are captured by four points and one projective transformation, was also recently presented in [40] using different notations and tools than those used here and in [35], [38].

This "middle ground" approach has several advantages. First, the results are sharper than a full projective reconstruction approach ([8], [15], [36]) where five scene points are needed. The increased sharpness translates to a remarkably simple framework captured by a single equation (1). Second, the manner in which the results were derived provides the means for unifying a wide range of other previous results, thus obtaining a canonical framework. Following Theorem 2, the corollaries show how this "middle ground" reduces back to full affine structure and extends into full projective structure (Corollaries 1 and 3). The corollaries also show how the "plane at infinity" is easily manipulated in this framework, thereby making further connections among projective affine and Euclidean results in general and less general situations. The corollaries also unify the various results related to the epipolar geometry of two views: the Essential matrix of [22], the Fundamental matrix of [8], and other related results of [15] (Corollary 4). All the above connections and results are often obtained as a single-line proof and follow naturally from the relative affine framework.

Acknowledgments

A. Shashua acknowledges support from McDonnell-Pew postdoctoral fellowship (during 1993/94 at MIT), and recently support from US-IS Binational Science Foundation 94-00120/1 and by the European ACTS project No. AC074 VANGUARD. N. Navab was supported by a Nichidai Foundation fellowship provided by Prof. Aaron Bobick at MIT Media Laboratory.

References

[1] A. Azarbayejani and A. Pentland, "Recursive Estimation of Motion, Structure, and Focal Length," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 6, pp. 562-575, June 1995.

[2] E. Barrett, P.M. Payton, and G. Gheen, "Robust Algebraic Invariant Methods with Applications in Geometry and Imaging," *Proc. SPIE Symp. Remote Sensing*, San Diego, July 1995.

[3] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani, "Hierarchical Model-Based Motion Estimation," *Proc. European Conf. Computer Vision*, Santa Margherita Ligure, Italy, June 1992.

[4] D. Beymer, A. Shashua, and T. Poggio, "Example Based Image Analysis and Synthesis," A.I. Memo no. 1431, Artificial Intelligence Laboratory, Massachusetts Inst. of Technology, Oct. 1993.

[5] S. Demey, A. Zisserman, and P. Beardsley, "Affine and Projective Structure from Motion," *Proc. British Machine Vision Conf.*, Oct. 1992.

[6] R. Deriche, Z. Zhang, Q.T. Luong, and O.D. Faugeras, "Robust Recovery of the Epipolar Geometry for an Uncalibrated Stereo Rig," *Proc. European Conf. Computer Vision*, pp. 567–576, Stockholm, Sweden, May 1994.

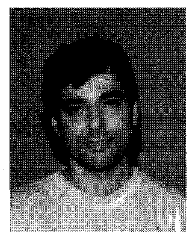
[7] R. Dutta, R. Manmatha, L.R. Williams, and E.M. Riseman, "A Data Set for Quantitative Motion Analysis," *Proc. 1989 IEEE Conf. Computer Vision and Pattern Recognition*, Los Alamitos, Calif.: IEEE CS Press, June 1989.

[8] O.D. Faugeras, "What Can Be Seen in Three Dimensions with an Uncalibrated Stereo Rig?" *Proc. European Conf. Computer Vision*, pp. 563–578, Santa Margherita Ligure, Italy, June 1992.

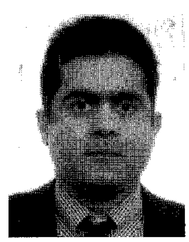
[9] O.D. Faugeras, "Stratification of Three-Dimensional Vision: Projective, Affine and Metric Representations," *J. Optical Soc. of Am.*, vol. 12, no. 3, pp. 465–484, 1995.

[10] O.D. Faugeras, Q.T. Luong, and S.J. Maybank, "Camera Self Calibration: Theory and Experiments," *Proc. European Conf. Computer Vision*, pp. 321–334, Santa Margherita Ligure, Italy, June 1992.

[11] O.D. Faugeras and F. Lustman, "Let Us Suppose that the World Is Piecewise Planar," *Int'l Symp. Robotics Research*, O.D. Faugeras and G. Giralt, eds., pp. 33–40. Cambridge, Mass.: MIT Press, 1986.
[12] O.D. Faugeras and L. Robert, "What Can Two Images Tell Us about a Third One?" *Proc. European Conf. Computer Vision*, pp. 485–492, Stockholm, Sweden, May 1994.
[13] R. Hartley, "Projective Reconstruction and Invariants from Multiple Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 10, pp. 1,036–1,040, 1994.
[14] R. Hartley and R. Gupta, "Computing Matched-Epipolar Projections," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 549–555, New York, 1993.
[15] R. Hartley, R. Gupta, and T. Chang, "Stereo from Uncalibrated Cameras," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 761–764, Champaign, Ill., June 1992.
[16] D.P. Huttenlocher and S. Ullman, "Recognizing Solid Objects by Alignment with an Image," *Int'l J. Computer Vision*, vol. 5, no. 2, pp. 195–212, 1990.
[17] D.W. Jacobs, "Space Efficient 3D Model Indexing," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 439–444, 1992.
[18] J.J. Koenderink and A.J. Van Doorn, "Affine Structure from Motion," *J. Optical Soc. Am.*, vol. 8, pp. 377–385, 1991.
[19] R. Kumar and P. Anandan, "Direct Recovery of Shape from Multiple Views: A Parallax Based Approach," *Proc. Int'l Conf. Pattern Recognition*, Jerusalem, Israel, Oct. 1994.
[20] R. Kumar, H.S. Sawhney, and A.R. Hanson, "3D Model Acquisition from Monocular Image Sequences," *1992 IEEE Conf. Computer Vision and Pattern Recognition*, pp. 209–215, Los Alamitos, Calif.: IEEE CS Press, June 1992.
[21] C.H. Lee, "Structure and Motion from Two Perspective Views via Planar Patch," *Proc. Int'l Conf. Computer Vision*, pp. 158–164, Tampa, Fla., Dec. 1988.
[22] H.C. Longuet-Higgins, "A Computer Algorithm for Reconstructing a Scene from Two Projections," *Nature*, vol. 293, pp. 133–135, 1981.
[23] B.D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Proc. IJCAI*, pp. 674–679, Vancouver, Canada, 1981.
[24] Q.T. Luong and T. Vieville, "Canonic Representations for the Geometries of Multiple Projective Views," *Proc. European Conf. Computer Vision*, pp. 589–599, Stockholm, Sweden, May 1994.
[25] R. Mohr, F. Veillon, and L. Quan, "Relative 3D Reconstruction Using Multiple Uncalibrated Images," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 543–548, New York, June 1993.
[26] T. Moons, L. Van Gool, M. Van Diest, and E. Pauwels, "Affine Reconstruction from Perspective Image Pairs," *Proc. Second European Workshop on Invariants*, Ponta Delagada, Azores, Oct. 1993.
[27] J.L. Mundy, R.P. Welty, M.H. Brill, P.M. Payton, and E.B. Barrett, "3-D Model Alignment Without Computing Pose," *Proc. ARPA Image Understanding Workshop*, pp. 727–735. San Mateo, Calif.: Morgan Kaufmann, Jan. 1992.
[28] N. Navab and A. Shashua, "Algebraic Description of Relative Affine Structure: Connections to Euclidean, Affine and Projective Structure," Technical Report 270, Media Laboratory, Massachusetts Inst. of Technology, 1994.
[29] J. Oliensis and J.I. Thomas, "Incorporating Motion Error in Multi-Frame Structure from Motion," *Proc. IEEE Workshop Visual Motion*, pp. 8–13, Los Alamitos, Calif.: IEEE CS Press, Oct. 1991.
[30] L. Quan, "Affine Stereo Calibration for Relative Affine Shape Reconstruction," *Proc. British Machine Vision Conf.*, pp. 659–668, 1993.
[31] L. Robert and O.D. Faugeras, "Relative 3D Positioning and 3D Convex Hull Computation from a Weakly Calibrated Stereo Pair," *Proc. Int'l Conf. Computer Vision*, pp. 540–544, Berlin, May 1993.
[32] H.S. Sawhney, "3D Geometry from Planar Parallax," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 929–934, Seattle, June 1994.
[33] L.S. Shapiro, A. Zisserman, and M. Brady, "Motion from Point Matches Using Affine Epipolar Geometry," *Proc. European Conf. Computer Vision*, pp. 73–84, Stockholm, Sweden, May 1994.
[34] A. Shashua, "Correspondence and Affine Shape from Two Orthographic Views: Motion and Recognition," A.I. Memo no. 1327, Artificial Intelligence Laboratory, Massachusetts Inst. of Technology, Dec. 1991.
[35] A. Shashua, "On Geometric and Algebraic Aspects of 3D Affine and Projective Structures from Perspective 2D Views," *Applications of Invariance in Computer Vision*, J.L. Mundy, A. Zisserman, and D. Forsyth, eds., *Proc. Second European Workshop Invariants*, Ponta Delagada, Azores, Oct. 1993 (also in MIT AI Memo 1405, July 1993).
[36] A. Shashua, "Projective Structure from Uncalibrated Images: Structure from Motion and Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 8, pp. 778–790, Aug. 1994.
[37] A. Shashua, "Algebraic Functions for Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 779–789, Aug. 1995.
[38] A. Shashua and N. Navab, "Relative Affine Structure: Theory and Application to 3D Reconstruction from Perspective Views," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 483–489, Seattle, 1994.
[39] G. Sparr, "An Algebraic-Analytic Method for Reconstruction from Image Correspondences," *Seventh Scandinavian Conf. Image Analysis*, Aalborg, Denmark, 1991.
[40] G. Sparr, "A Common Framework for Kinetic Depth, Reconstruction and Motion for Deformable Objects," *Proc. European Conf. Computer Vision*, pp. 471–482, Stockholm, Sweden, May 1994.
[41] R. Szeliski and S.B. Kang, "Recovering 3D Shape and Motion from Image Streams Using Non-Linear Least Squares," Technical Report D.E.C., Dec. 1992.
[42] R. Tsai and T.S. Huang, "Estimating Three-Dimensional Motion Parameters of a Rigid Planar Patch, II: Singular Value Decomposition," *IEEE Trans. Acoustic, Speech and Signal Processing*, vol. 30, 1982.
[43] S. Ullman, "Aligning Pictorial Descriptions: An Approach to Object Recognition," *Cognition*, vol. 32, pp. 193–254, 1989. Also in MIT AI Memo 931, Dec. 1986.
[44] D. Weinshall, "Model Based Invariants for 3-D Vision," *Int'l J. Computer Vision*, vol. 10, no. 1, pp. 27–42, 1993.
[45] D. Weinshall and C. Tomasi, "Linear and Incremental Acquisition of Invariant Shape Models from Image Sequences," *Proc. Int'l Conf. Computer Vision*, pp. 675–682, Berlin, Germany, May 1993.



Amnon Shashua received the BSc degree in mathematics and computer science from Tel-Aviv University, Tel-Aviv, Israel, in 1986, the MSC degree in mathematics and computer science from the Weizmann Institute of Science, Rehovot, Israel, in 1989, and the PhD degree in computational neuroscience, working at the Artificial Intelligence Laboratory, from the Massachusetts Institute of Technology in 1993. He did postdoctoral research at the Artificial Intelligence Laboratory and at the Center for Biological Computational Learning at MIT and at the Institute of Computer Science of the Hebrew University of Jerusalem, Israel. In October 1995, he joined the faculty of the Computer Science Department of the Technion, where he is currently a senior lecturer.



Nassir Navab received the PhD degree in computer science from the University of Paris XI (Orsay) in January 1993, the Computer Engineering Diploma and the DEA in system control, both in 1988, from the University of Technology of Compiegne, and the "Diplome d'Etudes Generales in Mathematiques and Physiques" in 1985 from the University of Nice-Sophia Antipolis. Dr. Navab has been with Siemens Corporate Research, Princeton, New Jersey, since November 1994. From February 1993 to November 1994, he was a postdoctoral fellow at the Media Laboratory of the Massachusetts Institute of Technology. He was the 1985 valedictorian (Major de Promotion) at the University of Nice-Sophia Antipolis. From 1988 to 1993, he was a research associate in the Computer Vision and Robotics Group of the French National Institute for Research in Computer Science and Control (INRIA), France. He was a research visitor at the Center for Automation Research, the University of Maryland, College Park, in 1988. His current research interests are in computer vision and medical image processing.