MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 1363 September, 1992

Projective Structure from two Uncalibrated Images: Structure from Motion and Recognition

Amnon Shashua

Abstract

This paper addresses the problem of recovering relative structure, in the form of an invariant, from two views of a 3D scene. The invariant structure is computed without any prior knowledge of camera geometry, or internal calibration, and with the property that perspective and orthographic projections are treated alike, namely, the system makes no assumption regarding the existence of perspective distortions in the input images.

We show that, given the location of epipoles, the projective structure invariant can be constructed from only four corresponding points projected from four non-coplanar points in space (like in the case of parallel projection). This result leads to two algorithms for computing projective structure. The first algorithm requires six corresponding points, four of which are assumed to be projected from four coplanar points in space. Alternatively, the second algorithm requires eight corresponding points, without assumptions of coplanarity of object points.

Our study of projective structure is applicable to both structure from motion and visual recognition. We use projective structure to re-project the 3D scene from two model images and six or eight corresponding points with a novel view of the scene. The re-projection process is well-defined under all cases of central projection, including the case of parallel projection.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-85-K-0124. A. Shashua was also supported by NSF-IRI8900267.

1 Introduction

The problem we address in this paper is that of recovering relative, non-metric, structure of a three-dimensional scene from two images, taken from different viewing positions. The relative structure information is in the form of an invariant that can be computed without any prior knowledge of camera geometry, and under all central projections — including the case of parallel projection. The non-metric nature of the invariant allows the cameras to be internally uncalibrated (intrinsic parameters of camera are unknown). The unique nature of the invariant allows the system to make no assumptions about existence of perspective distortions in the input images. Therefore, any degree of perspective distortions is allowed, i.e., orthographic and perspective projections are treated alike, or in other words, no assumptions are made on the size of field of view.

We envision this study as having applications both in the area of structure from motion and in the area of visual recognition. In structure from motion our contribution is an addition to the recent studies of non-metric structure from motion pioneered by Koenderink and Van Doorn (1991) in parallel projection, followed by Faugeras (1992) and Mohr, Quan, Veillon & Boufama (1992) for reconstructing the projective coordinates of a scene up to an unknown projective transformation of 3D projective space. Our approach is similar to Koenderink and Van Doorn's in the sense that we derive an invariant, based on a geometric construction, that records the 3D structure of the scene as a variation from two fixed reference planes measured along the line of sight. Unlike Faugeras and Mohr *et al.* we do not recover the projective coordinates of the scene, and, as a result, we use a smaller number of corresponding points: in addition to the location of epipoles we need only four corresponding points, coming from four non-coplanar points in the scene, whereas Faugeras and Mohr *et al.* require correspondences coming from five points in general position.

The second contribution of our study is to visual recognition of 3D objects from 2D images. We show that our projective invariant can be used to predict novel views of the object, given two model views in full correspondence and a small number of corresponding points with the novel view. The predicted view is then matched against the novel input view, and if the two match, then the novel view is considered to be an instance of the same object that gave rise to the two model views stored in memory. This paradigm of recognition is within the general framework of *alignment* (Fischler and Bolles 1981, Lowe 1985, Ullman 1989, Huttenlocher and Ullman 1987) and, more specifically, of the paradigm proposed by Ullman and Basri (1989) that recognition can proceed using only 2D images, both for representing the model, and when matching the model to the input image. We refer to the problem of predicting a novel view from a set of model views using a limited number of corresponding points, as the problem of *re-projection*.

The problem of re-projection has been dealt with in the past primarily assuming parallel projection (Ullman and Basri 1989, Koenderink and Van Doorn 1991). For the more general case of central projection, Barret, Brill, Haag & Pyton (1991) have recently introduced a quadratic invariant based on the fundamental matrix of Longuet-Higgins (1981), which is computed from eight corresponding points. In Appendix E we show that their result is equivalent to intersecting epipolar lines, and therefore, is singular for certain viewing transformations depending on the viewing geometry between the two model views. Our projective invariant is not based on an epipolar intersection, but is based directly on the relative structure of the object, and does not suffer from any singularities, a finding that implies greater stability in the presence of errors.

The projective structure invariant, and the reprojection method that follows, is based on an extension of Koenderink and Van-Doorn's representation of affine structure as an invariant defined with respect to a reference plane and a reference point. We start by introducing an alternative affine invariant, using two reference planes (section 5), and it can easily be extended to projective space. As a result we obtain a projective structure invariant (section 6).

We show that the difference between the affine and projective case lie entirely in the location of the epipoles, i.e., given the location of epipoles both the affine and projective structures are constructed by linear methods using the information captured from four corresponding points projected from four non-coplanar points in space. In the projective case we need additional corresponding points — solely for the purpose of recovering the location of the epipoles (Theorem 1, section 6).

We show that the projective structure invariant can be recovered from two views — produced by parallel or central projection — and six corresponding points, four of which are assumed to be projected from four coplanar points in space (section 7.1). Alternatively, the projective structure can be recovered from eight corresponding points, without assuming coplanarity of object points (section 8.1). The 8-point method uses the fundamental matrix approach (Longuett-Higgins, 1981) for recovering the location of epipoles (as suggested by Faugeras, 1992).

Finally, we show that, for both schemes, it is possible to limit the viewing transformations to the group of rigid motions, i.e., it is possible to work with perspective projection assuming the cameras are calibrated. The result, however, does not include orthographic projection.

Experiments were conducted with both algorithms, and the results show that the 6-point algorithm is stable under noise and under conditions that violate the assumption that four object points are coplanar. The 8-point algorithm, although theoretically superior because of lack of the coplanarity assumption, is considerably more sensitive to noise.

2 Why not Classical SFM?

The work of Koenderink and Van Doorn (1991) on affine structure from two orthographic views, and the work of Ullman and Basri (1989) on re-projection from two orthographic views, have a clear practical aspect: it is known that at least three orthographic views are required to recover metric structure, i.e., relative depth

(Ullman 1979, Huang & Lee 1989, Aloimonos & Brown 1989). Therefore, the suggestion to use affine structure instead of metric structure allows a recognition system to perform re-projection from two-model views (Ullman & Basri), and to generate novel views of the object produced by affine transformations in space, rather than by rigid transformations (Koenderink & Van Doorn).

This advantage, of working with two rather than three views, is not present under perspective projection, however. It is known that two perspective views are sufficient for recovering metric structure (Roach & Aggarwal 1979, Longuett-Higgins 1981, Tsai & Huang 1984, Faugeras & Maybank 1990). The question, therefore, is why look for alternative representations of structure, and new methods for performing re-projection?

There are three major problems in structure from motion methods: (i) critical dependence on an orthographic or perspective model of projection, (ii) internal camera calibration, and (iii) the problem of stereo-triangulation.

The first problem is the strict division between methods that assume orthographic projection and methods that assume perspective projection. These two classes of methods do not overlap in their domain of application. The perspective model operates under conditions of significant perspective distortions, such as driving on a stretch of highway, requires a relatively large field of view and relatively large depth variations between scene points (Adiv 1989, Dutta & Synder 1990, Tomasi 1991, Broida *et al.* 1990). The orthographic model, on the other hand, provides a reasonable approximation when the imaging situation is at the other extreme, i.e., small field of view and small depth variation between object points (a situation for which perspective schemes often break down). Typical imaging situations are at neither end of these extremes and, therefore, would be vulnerable to errors in both models. From the standpoint of performing recognition, this problem implies that the viewer has control over his field of view — a property that may be reasonable to assume at the time of model acquisition, but less reasonable to assume occurring at recognition time.

The second problem is related to internal camera calibration. The assumption of perspective projection includes a distinguishable point, known as the principal point, which is at the intersection of the optical axis and the image plane. The location of the principal point is an internal parameter of the camera, which may deviate somewhat from the geometric center of the image plane, and therefore, may require calibration. Perspective projection also assumes that the image plane is perpendicular to the optical axis and the possibility of imperfections in the camera requires, therefore, the recovery of the two axes describing the image frame, and of the focal length. Although the calibration process is somewhat tedious, it is sometimes necessary for many of the available commercial cameras (Brown 1971, Faig 1975, Lenz and Tsai 1987, Faugeras, Luong and Maybank 1992). The problem of calibration is lesser under orthographic projection because the projection does not have a distinguishable ray; therefore any point can serve as an origin, however must still be considered because of the assumption that the image plane is perpendicular to the projecting rays.

The third problem is related to the way shape is typically represented under the perspective projection model. Because the center of projection is also the origin of the coordinate system for describing shape, the shape difference (e.g., difference in depth, between two object points), is orders of magnitude smaller than the distance to the scene, and this makes the computations very sensitive to noise. The sensitivity to noise is reduced if images are taken from distant viewpoints (large base-line in stereo triangulation), but that makes the process of establishing correspondence between points in both views more of a problem, and hence, may make the situation even worse. This problem does not occur under the assumption of orthographic projection because translation in depth is lost under orthographic projection, and therefore, the origin of the coordinate system for describing shape (metric and non-metric) is object-centered, rather than viewer-centered (Tomasi, 1991).

These problems, in isolation or put together, make much of the reason for the sensitivity of structure from motion methods to errors. The recent work of Faugeras (1992) and Mohr *et al.* (1992) addresses the problem of internal calibration by assuming central projection instead of perspective projection. Faugeras and Mohr *et al.* then proceed to reconstruct the projective coordinates of the scene. Since projective coordinates are measured relative to the center of projection, this approach does not address the problem of stereo-triangulation or the problem of uniformity under both orthographic and perspective projection models.

3 Camera Model and Notations

We assume that objects in the world are rigid and are viewed under *central projection*. In central projection the center of projection is the origin of the camera coordinate frame and can be located anywhere in projective space. In other words, the center of projection can be a point in Euclidean space or an *ideal* point (such as happens in parallel projection). The image plane is assumed to be arbitrarily positioned with respect to the camera coordinate frame (unlike perspective projection where it is parallel to the xy plane). We refer to this as a *non-rigid camera* configuration. The motion of the camera, therefore, consists of the translation of the center of projection, rotation of the coordinate frame around the new location of the center of projection, and followed by tilt, pan, and focal length scale of the image plane with respect to the new optical axis. This model of projection will also be referred to as perspective projection with an uncalibrated camera.

We also include in our derivations the possibility of having a *rigid camera* configuration. A rigid camera is simply the familiar model of *perspective projection* in which the center of projection is a point in Euclidean space and the image plane is fixed with respect to the camera coordinate frame. A rigid camera motion, therefore, consists of translation of the center of projection followed by rotation of the coordinate frame and focal length scaling. Note that a rigid camera implicitly assumes internal calibration, i.e., the optical axis pierces

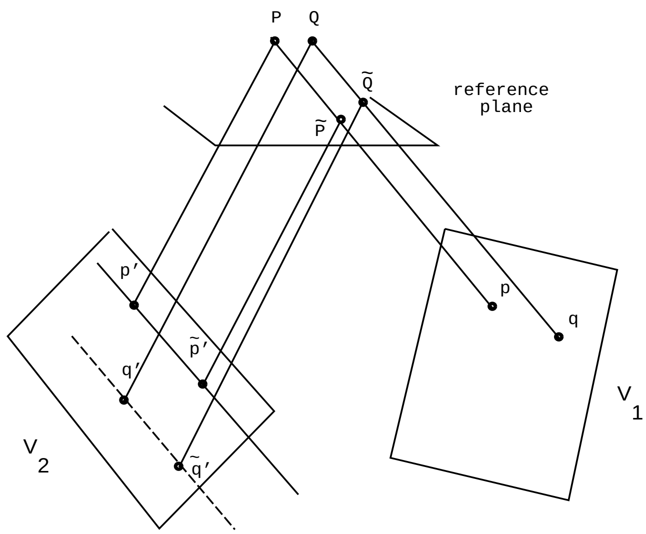

Figure 1: Koenderink and Van Doorn's Affine Structure.

through a fixed point in the image and the image plane is perpendicular to the optical axis.

We denote object points in capital letters and image points in small letters. If P denotes an object point in 3D space, p, p', p'' denote its projections onto the first, second and novel projections, respectively. We treat image points as rays (homogeneous coordinates) in 3D space, and refer to the notation $p = (x, y, 1)$ as the *standard representation* of the image plane. We note that the true coordinates of the image plane are related to the standard representation by means of a projective transformation of the plane. In case we deal with central projection, all representations of image coordinates are allowed, and therefore, without loss of generality we work with the standard representation (more on that in Appendix A).

4 Affine Structure: Koenderink and Van Doorn's Version

The affine structure invariant described by Koenderink and Van Doorn (1991) is based on a geometric construction using a single reference plane, and a reference point not coplanar with the reference plane. In affine geometry (induced by parallel projection), it is known from the fundamental theorem of plane projectivity, that three (non-collinear) corresponding points are sufficient to uniquely determine all other correspondences (see Appendix A for more details on plane projectivity under affine and projective geometry). Using three corresponding points between two views provides us, therefore, with a transformation (affine transformation) for determining the location of all points of the plane passing through the three reference points in the second image plane.

Let P be an arbitrary point in the scene projecting onto p, p' on the two image planes. Let $\tilde{P}$ be the projection of P onto the reference plane along the ray towards the first image plane, and let $\tilde{p}'$ be the projection of $\tilde{P}$ onto the second image plane (p' and $\tilde{p}'$ coincide if P is on the reference plane). Note that the location of $\tilde{p}'$ is known via the affine transformation determined by the projections of the three reference points. Finally, let Q be the fourth reference point (not on the reference plane). Using a simple geometric drawing, the affine structure invariant is derived as follows.

Consider Figure 1. The projections of the reference point Q and an arbitrary point of interest P form two similar trapezoids: $P\tilde{P}p'\tilde{p}'$ and $Q\tilde{Q}q'\tilde{q}'$. From similarity of trapezoids we have,

$$\gamma_p = \frac{|P - \tilde{P}|}{|Q - \tilde{Q}|} = \frac{|p' - \tilde{p}'|}{|q' - \tilde{q}'|}.$$

By assuming that q, q' is a given corresponding point, we obtain a shape invariant that is invariant under parallel projection (the object points are fixed while the camera changes the location and position of the image plane towards the projecting rays).

Before we extend this result to central projection by using projective geometry, we first describe a different affine invariant using two reference planes, rather than one reference plane and a reference point. The new affine invariant is the one that will be applied later to central projection.

5 Affine Structure Using Two Reference Planes

We make use of the same information — the projections of four non-coplanar points — to set up two reference planes. Let P_j, $j = 1, ..., 4$, be the four non-coplanar reference points in space, and let $p_j \longleftrightarrow p'_j$ be their observed projections in both views. The points P_1, P_2, P_3 and P_2, P_3, P_4 lie on two different planes, therefore, we can account for the motion of all points coplanar with each of these two planes. Let P be a point of interest, not coplanar with either of the reference planes, and let $\tilde{P}$ and $\hat{P}$ be its projections onto the two reference planes along the ray towards the first view.

Consider Figure 2. The projection of $P, \tilde{P}$ and $\hat{P}$ onto $p', \tilde{p}'$ and $\hat{p}'$ respectively, gives rise to two similar trapezoids from which we derive the following relation:

$$\alpha_p = \frac{|P - \tilde{P}|}{|P - \hat{P}|} = \frac{|p' - \tilde{p}'|}{|p' - \hat{p}'|}.$$

The ratio α_p is invariant under parallel projection. There is no particular advantage for preferring α_p over γ_p as a measure of affine structure, but as will be described below, this new construction forms the basis for extending affine structure to projective structure, whereas the single reference plane construction does not (see Appendix D for proof).

In the projective plane, we need four coplanar points to determine the motion of a reference plane. We show that, given the epipoles, only three corresponding points for each reference plane are sufficient for recovering the associated projective transformations induced by those planes. Altogether, the construction provides us with four points along each epipolar line. The similarity of trapezoids in the affine case turns, therefore, into a cross-ratio in the projective case.

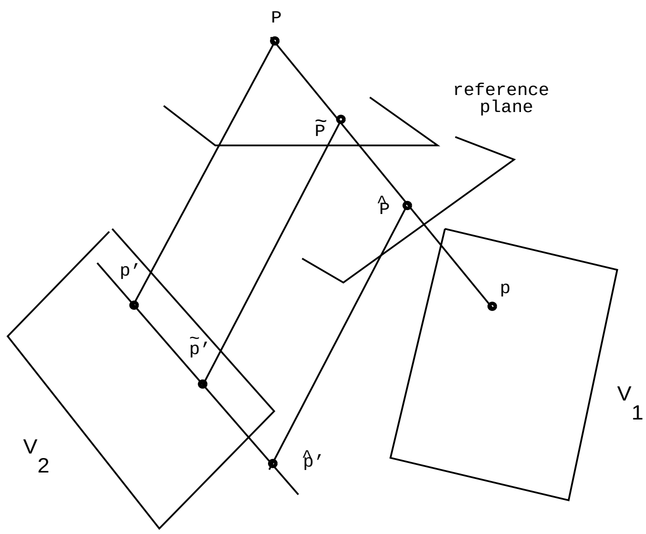

Figure 2: Affine structure using two reference planes.

This leads to the result (Theorem 1) that, in addition to the epipoles, only four corresponding points, projected from four non-coplanar points in the scene, are sufficient for recovering the projective structure invariant for all other points. The epipoles can be recovered by either extending the Koenderink and Van Doorn (1991) construction to projective space using six points (four of which are assumed to be coplanar), or by using other methods, notably those based on the Longuet-Higgins fundamental matrix. This leads to projective structure from eight points in general position.

6 Projective Structure

We assume for now that the location of both epipoles is known, and we will address the problem of finding the epipoles later. The epipoles, also known as the foci of expansion, are the intersections of the line in space connecting the two centers of projection and the image planes. There are two epipoles, one on each image plane — the epipole on the second image we call the left epipole, and the epipole on the first image we call the right epipole. The image lines emanating from the epipoles are known as the *epipolar lines.*

Consider Figure 3 which illustrates the two reference plane construction, defined earlier for parallel projection, now displayed in the case of central projection. The left epipole is denoted by V_l, and because it is on the line V_1V_2 (connecting the two centers of projection), the line PV_1 projects onto the epipolar line $p'V_l$. Therefore, the points $\tilde{P}$ and $\hat{P}$ project onto the points $\tilde{p}'$ and $\hat{p}'$, which are both on the epipolar line $p'V_l$. The points $p', \tilde{p}', \hat{p}'$ and V_l are collinear and projectively related to $P, \tilde{P}, \hat{P}, V_1$, and therefore have the same cross-ratio:

$$\alpha_p = \frac{|P-\tilde{P}|}{|P-\hat{p}|}\cdot\frac{|V_1-\hat{p}|}{|V_1-\tilde{P}|} = \frac{|p'-\tilde{p}'|}{|p'-\hat{p}'|}\cdot\frac{|V_l-\hat{p}'|}{|V_l-\tilde{p}'|}.$$

Note that when the epipole V_l becomes an ideal point (vanishes along the epipolar line), then α_p is the same

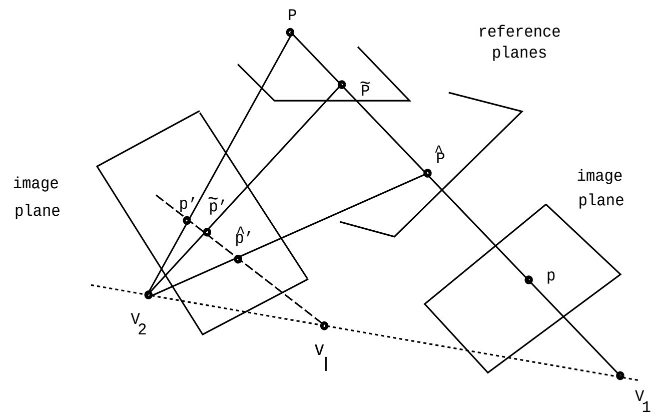

Figure 3: Definition of projective shape as the cross ratio of $p', \tilde{p}', \hat{p}', V_l$.

as the affine invariant defined in section 5 for parallel projection.

The cross-ratio α_p is a direct extension of the affine structure invariant defined in section 5 and is referred to as *projective structure.* We can use this invariant to reconstruct any novel view of the object (taken by a non-rigid camera) without ever recovering depth or even projective coordinates of the object.

Having defined the projective shape invariant, and assuming we still are given the locations of the epipoles, we show next how to recover the projections of the two reference planes onto the second image plane, i.e., we describe the computations leading to $\tilde{p}'$ and $\hat{p}'$.

Since we are working under central projection, we need to identify four coplanar points on each reference plane. In other words, in the projective geometry of the plane, four corresponding points, no three of which are collinear, are sufficient to determine uniquely all other correspondences (see Appendix A, for more details). We must, therefore, identify four corresponding points that are projected from four coplanar points in space, and then recover the projective transformation that accounts for all other correspondences induced from that plane. The following proposition states that the corresponding epipoles can be used as a fourth corresponding point for any three corresponding points selected from the pair of images.

Proposition 1 *A projective transformation, A, that is determined from three arbitrary, non-collinear, 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 transformation A is an induced epipolar transformation, i.e., the ray Ap intersects the epipolar line $p'V_l$ for any arbitrary image point p and its corresponding point p'.*

Comment: An epipolar transformation F is a mapping between corresponding epipolar lines and is determined (not uniquely) from three corresponding epipolar lines and the epipoles. The induced point transformation is $E = (F^{-1})^t$ (induced from the point/line duality of projective geometry, see Appendix C for more details on epipolar transformations).

Proof: Let $p_j \longleftrightarrow p'_j$, $j = 1, 2, 3$, be three arbitrary corresponding points, and let V_l and V_r denote the left and right epipoles. First note that the four points p_j and V_r are projected from four coplanar points in the scene. The reason is that the plane defined by the three object points P_j intersects the line V_1V_2 connecting the two centers of projection, at a point — regular or ideal. That point projects onto both epipoles. The transformation A, therefore, is a projective transformation of the plane passing through the three object points P_1, P_2, P_3. Note that A is uniquely determined provided that no three of the four points are collinear.

Let $\mu\tilde{p}' = Ap$ for some arbitrary point p. Because lines are projective invariants, any point along the epipolar line pV_r must project onto the epipolar line $p'V_l$. Hence, A is an induced epipolar transformation. □

Given the epipoles, therefore, we need just three points to determine the correspondences of all other points coplanar with the reference plane passing through the three corresponding object points. The transformation (collineation) A is determined from the following equations:

$$Ap_j = \rho_j p'_j, \qquad j = 1, 2, 3$$
$$AV_r = \rho V_l,$$

where ρ, ρ_j are unknown scalars, and $A_{3,3} = 1$. One can eliminate ρ, ρ_j from the equations and solve for the matrix A from the three corresponding points and the corresponding epipoles. That leads to a linear system of eight equations, and is described in more detail in Appendix A.

If P_1, P_2, P_3 define the first reference plane, the transformation A determines the location of $\tilde{p}'$ for all other points p ($\tilde{p}'$ and p' coincide if P is coplanar with the first reference plane). In other words, we have that $\tilde{p}' = Ap$. Note that $\tilde{p}'$ is not necessarily a point on the second image plane, but it is on the line $V_2\tilde{P}$. We can determine its location on the second plane by normalizing Ap such that its third component is set to 1.

Similarly, let P_2, P_3, P_4 define the second reference plane (assuming the four object points P_j, $j = 1, ..., 4$, are non-coplanar). The transformation E is uniquely determined by the equations

$$Ep_j = \rho_j p'_j, \qquad j = 2, 3, 4$$
$$EV_r = \rho V_l,$$

and determines all other correspondences induced by the second reference plane (we assume that no three of the four points used to determine E are collinear). In other words, Ep determines the location of $\hat{p}'$ up to a scale factor along the ray $V_2\hat{P}$.

Instead of normalizing Ap and Ep we compute α_p from the cross-ratio of the points represented in homogeneous coordinates, i.e., the cross-ratio of the four rays $V_2p', V_2\tilde{p}', V_2\hat{p}', V_2V_l$, as follows: Let the rays p', V_l be represented as a linear combination of the rays $\tilde{p}' = Ap$ and $\hat{p}' = Ep$, i.e.,

$$p' = \tilde{p}' + k\hat{p}'$$
$$V_l = \tilde{p}' + k'\hat{p}',$$

then $\alpha_p = \frac{k}{k'}$ (see Appendix B for more details). This way of computing the cross-ratio is preferred over the more familiar cross-ratio of four collinear points, because it enables us to work with all elements of the projective plane, including ideal points (a situation that arises, for instance, when epipolar lines are parallel, and in general under parallel projection).

We have therefore shown the following result:

Theorem 1 *In the case where the location of epipoles are known, then four corresponding points, coming from four non-coplanar points in space, are sufficient for computing the projective structure invariant α_p for all other points in space projecting onto corresponding points in both views, for all central projections, including parallel projection.*

This result shows that the difference between parallel and central projection lies entirely on the epipoles. In both cases four non-coplanar points are sufficient for obtaining the invariant, but in the parallel projection case we have prior knowledge that both epipoles are ideal, therefore they are not required for determining the transformations A and E (in other words, A and E are affine transformations, more on that in Section 7.2).

Another point to note with this result is that the minimal number of corresponding points needed for re-projection is smaller than the previously reported number (Faugeras 1992, Mohr *et al.* 1992) for recovering the projective coordinates of object points. Faugeras shows that five corresponding points coming from five points in general position (i.e., no four of them are coplanar) can be used, together with the epipoles, to recover the projective coordinates of all other points in space. Because the projective structure invariant requires only four points, this implies that re-projection is done more directly than through full reconstruction of projective coordinates, and therefore is likely to be more stable.

We next discuss algorithms for recovering the location of epipoles. The problem of recovering the epipoles is well known and several approaches have been suggested in the past (Longuet-Higgins and Prazdny 1980, Rieger-Lawton 1985, Faugeras and Maybank 1990, Hildreth 1991, Faugeras 1992, Faugeras, Luong and Maybank 1992). We start with a method that requires six corresponding points (two additional points to the four we already have). The method is a direct extension of the Koenderink and Van Doorn (1991) construction in parallel projection, and was described earlier by Lee (1988) for the purpose of recovering the translational component of camera motion.

The second algorithm for locating the epipoles is adopted from Faugeras (1992) and is based on the fundamental matrix of Longuet-Higgins (1981).

7 Epipoles from Six Points

We can recover the correspondences induced from the first reference plane by selecting four corresponding points, assuming they are projected from four coplanar object points. Let $p_j = (x_j, y_j, 1)$ and $p'_j = (x'_j, y'_j, 1)$ and $j = 1, ..., 4$ represent the standard image coordinates of the four corresponding points, no three of which are

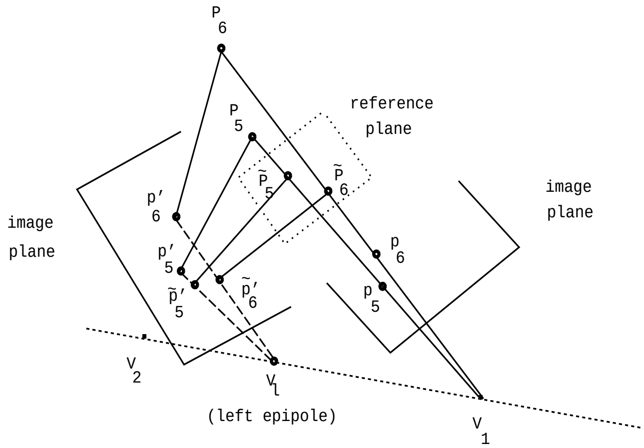

Figure 4: The geometry of locating the left epipole using two points out of the reference plane.

collinear, in both projections. Therefore, the transformation A is uniquely determined by the following equations,

$$\rho_j p'_j = A p_j .$$

Let $\tilde{p}' = Ap$ be the homogeneous coordinate representation of the ray $V_2\tilde{P}$, and let $\tilde{p}^{-1} = A^{-1}p'$.

Having accounted for the motion of the reference plane, we can easily find the location of the *epipoles* (in standard coordinates). Given two object points P_5, P_6 that are *not* on the reference plane, we can find both epipoles by observing that $\tilde{p}'$ is on the left epipolar line, and similarly that $\tilde{p}^{-1}$ is on the right epipolar line. Stated formally, we have the following proposition:

Proposition 2 *The left epipole, denoted by V_l, is at the intersection of the line $p'_5\tilde{p}'_5$ and the line $p'_6\tilde{p}'_6$. Similarly, the right epipole, denoted by V_r, is at the intersection of $p_5\tilde{p}_5^{-1}$ and $p_6\tilde{p}_6^{-1}$.*

Proof: It is sufficient to prove the claim for one of the epipoles, say the left epipole. Consider Figure 4 which describes the construction geometrically. By construction, the line $P_5\tilde{P}_5V_1$ projects to the line $p'_5\tilde{p}'_5$ via V_2 (points and lines are projective invariants) and therefore they are coplanar. In particular, V_1 projects to V_l which is located at the intersection of $p'_5\tilde{p}'_5$ and V_1V_2. Similarly, the line $p'_6\tilde{p}'_6$ intersects V_1V_2 at $\hat{V}_l$. Finally, V_l and $\hat{V}_l$ must coincide because the two lines $p'_5\tilde{p}'_5$ and $p'_6\tilde{p}'_6$ are coplanar (both are on the image plane). □

Algebraically, we can recover the ray V_1V_2, or V_l up to a scale factor, using the following formula:

$$V_l = (p'_5 \times \tilde{p}'_5) \times (p'_6 \times \tilde{p}'_6).$$

Note that V_l is defined with respect to the standard coordinate frame of the second camera. We treat the epipole V_l as the ray V_1V_2 with respect to V_2, and the epipole V_r as the same ray but with respect to V_1. Note also that the third component of V_l is zero if epipolar lines are parallel, i.e., V_l is an ideal point in projective terms (happening under parallel projection, or when the non-rigid camera motion brings the image plane to a position where it is parallel to the line V_1V_2).

In the case where more than two epipolar lines are available (such as when more than six corresponding points are available), one can find a least-squares solution for the epipole by using a principle component analysis, as follows. Let B be a $k \times 3$ matrix, where each row represents an epipolar line. The least squares solution to V_l is the unit eigenvector associated with the smallest eigenumber of the 3×3 matrix B^tB. Note that this can be done analytically because the characteristic equation is a cubic polynomial.

Altogether, we have a six point algorithm for recovering both the epipoles, and the projective structure α_p, and for performing re-projection onto any novel view. We summarize in the following section the 6-point algorithm.

7.1 Re-projection Using Projective Structure: 6-point Algorithm

We assume we are given two model views of a 3D object, and that all points of interest are in correspondence. We assume these correspondences can be based on measures of correlation, as used in optical-flow methods (see also Shashua 1991, Bachelder & Ullman 1992 for methods for extracting correspondences using combination of optical flow and affine geometry).

Given a novel view we extract six corresponding points (with one of the model views): $p_j \longleftrightarrow p'_j \longleftrightarrow p''_j$, $j = 1, ..., 6$. We assume the first four points are projected from four coplanar points, and the other corresponding points are projected from points that are not on the reference plane. Without loss of generality, we assume the standard coordinate representation of the image planes, i.e., the image coordinates are embedded in a 3D vector whose third component is set to 1 (see Appendix A). The computations for recovering projective shape and performing re-projection are described below.

1: Recover the transformation A that satisfies $\rho_j p'_j = Ap_j$, $j = 1, ..., 4$. This requires setting up a linear system of eight equations (see Appendix A). Apply the transformation to all points p, denoting $\tilde{p}' = Ap$. Also recover the epipoles $V_l = (p'_5 \times \tilde{p}'_5) \times (p'_6 \times \tilde{p}'_6)$ and $V_r = (p_5 \times A^{-1}p'_5) \times (p_6 \times A^{-1}p'_6)$.

2: Recover the transformation E that satisfies $\rho V_l = EV_r$ and $\rho_j p'_j = Ep_j$, $j = 4, 5, 6$.

3: Compute the cross-ratio of the points p', Ap, Ep, V_l, for all points p and denote that by α_p (see Appendix B for details on computing the cross-ratio of four rays).

4: Perform step 1 between the first and novel view: recover $\tilde{A}$ that satisfies $\rho_j p''_j = \tilde{A}p_j$, $j = 1, ..., 4$, apply $\tilde{A}$ to all points p and denote that by $\tilde{p}'' = \tilde{A}p$, recover the epipoles $V_{ln} = (p''_5 \times \tilde{p}''_5) \times (p''_6 \times \tilde{p}''_6)$ and $V_{rn} = (p_5 \times \tilde{A}^{-1}p''_5) \times (p_6 \times \tilde{A}^{-1}p''_6)$.

5: Perform step 2 between the first and novel view: Recover the transformation $\tilde{E}$ that satisfies $\rho V_{ln} = \tilde{E}V_{rn}$ and $\rho_j p''_j = Ep_j$, $j = 4, 5, 6$.

6: For every point p, recover p'' from the cross-ratio α_p and the three rays $\tilde{A}p, \tilde{E}p, V_{ln}$. Normalize p'' such that its third coordinate is set to 1.

The entire procedure requires setting up a linear system of eight equations four times (Step 1,2,4,5) and computing cross-ratios (linear operations as well).

We discuss below an important property of this procedure which is the transparency with respect to projection model: central and parallel projection are treated alike — a property which has implications on stability of re-projection no matter what degree of perspective distortions are present in the images.

7.2 The Case of Parallel Projection

The construction for obtaining projective structure is well defined for all central projections, including the case where the center of projection is an ideal point, i.e., such as happening with parallel projection. The construction has two components: the first component has to do with recovering the epipolar geometry via reference planes, and the second component is the projective invariant α_p.

From Proposition 1 the projective transformations A and E can be uniquely determined from three corresponding points and the corresponding epipoles. If both epipoles are ideal, the transformations become affine transformations of the plane (an affine transformation separates ideal points from Euclidean points). All other possibilities (both epipoles are Euclidean, one epipole Euclidean and the other epipole ideal) lead to projective transformations. Because a projectivity of the projective plane is uniquely determined from any four points on the projective plane (provided no three are collinear), the transformations A and E are uniquely determined under all situations of central projection — including parallel projection.

The projective invariant α_p is the same as the one defined under parallel projection (Section 5) — affine structure is a particular instance of projective structure in which the epipole V_l is an ideal point. By using the same invariant for both parallel and central projection, and because all other elements of the geometric construction hold for both projection models, the overall system is transparent to the projection model being used.

The first implication of this property has to do with stability. Projective structure does not require any perspective distortions, therefore all imaging situations can be handled — wide or narrow field of views. The second implication is that 3D visual recognition from 2D images can be achieved in a uniform manner with regard to the projection model. For instance, we can recognize (via re-projection) a perspective image of an object from only two orthographic model images, and in general any combination of perspective and orthographic images serving as model or novel views is allowed.

The results so far required prior knowledge (or assumption) that four of the corresponding points are coming from coplanar points in space. This requirement can be avoided, using two more corresponding points (making eight points overall), and is described in the next section.

8 Epipoles from Eight Points

We adopt a recent algorithm suggested by Faugeras (1992) which is based on Longuet-Higgins' (1981) fundamental matrix. The method is very simple and requires eight corresponding points for recovering the epipoles.

Let F be an epipolar transformation, i.e., $Fl = \mu l'$, where $l = V_r \times p$ and $l' = V_l \times p'$ are corresponding epipolar lines. We can rewrite the projective relation of epipolar lines using the matrix form of cross-products:

$$F(V_r \times p) = F[V_r]p = \rho l',$$

where $[V_r]$ is a skew symmetric matrix (and hence has rank 2). From the point/line incidence property we have that $p' \cdot l' = 0$ and therefore, ${p'}^t F[V_r]p = 0$, or ${p'}^t Hp = 0$ where $H = F[V_r]$. The matrix H is known as the fundamental matrix introduced by Longuet-Higgins (1981), and is of rank 2. One can recover H (up to a scale factor) directly from eight corresponding points, or by using a principle components approach if more than eight points are available. Finally, it is easy to see that

$$HV_r = 0,$$

and therefore the epipole V_r can be uniquely recovered (up to a scale factor). Note that the determinant of the first principle minor of H vanishes in the case where V_r is an ideal point, i.e., $h_{11}h_{22} - h_{12}h_{21} = 0$. In that case, the x, y components of V_r can be recovered (up to a scale factor) from the third row of H. The epipoles, therefore, can be uniquely recovered under both central and parallel projection. We have arrived at the following theorem:

Theorem 2 *In the case where we have eight corresponding points of two views taken under central projection (including parallel projection), four of these points, coming from four non-coplanar points in space, are sufficient for computing the projective structure invariant α_p for the remaining four points and for all other points in space projecting onto corresponding points in both views.*

We summarize in the following section the 8-point scheme for reconstructing projective structure and performing re-projection onto a novel view.

8.1 8-point Re-projection Algorithm

We assume we have eight corresponding points between two model views and the novel view, $p_j \longleftrightarrow p'_j \longleftrightarrow p''_j$, $j = 1, ..., 8$, and that the first four points are coming from four non-coplanar points in space. The computations for recovering projective structure and performing re-projection are described below.

1: Recover the fundamental matrix H (up to a scale factor) that satisfies ${p'_j}^t Hp_j$, $j = 1, ..., 8$. The right epipole V_r then satisfies $HV_r = 0$. Similarly, the left epipole is recovered from the relation $p^t \tilde{H} p'$ and $\tilde{H} V_l = 0$.

2: Recover the transformation A that satisfies $\rho V_l = AV_r$ and $\rho_j p'_j = Ap_j$, $j = 1, 2, 3$. Similarly, recover the transformation E that satisfies $\rho V_l = EV_r$ and $\rho_j p'_j = Ep_j$, $j = 2, 3, 4$.

3: Compute α_p as the cross-ratio of p', Ap, Ep, V_l, for all points p.

4: Perform step 1 and 2 between the first and novel view: recover the epipoles V_{rn}, V_{ln}, and the transformations $\tilde{A}$ and $\tilde{E}$.

5: For every point p, recover p'' from the cross-ratio α_p and the three rays $\tilde{A}p, \tilde{E}p, V_{ln}$. Normalize p'' such that its third coordinate is set to 1.

We discuss next the possibility of working with a rigid camera (i.e., perspective projection and calibrated camera).

9 The Rigid Camera Case

The advantage of the non-rigid camera model (or the central projection model) used so far is that images can be obtained from uncalibrated cameras. The price paid for this property is that the images that produce the same projective structure invariant (equivalence class of images of the object) can be produced by applying non-rigid transformations of the object, in addition to rigid transformations.

In this section we show that it is possible to verify whether the images were produced by rigid transformations, which is equivalent to working with perspective projection assuming the cameras are internally calibrated. This can be done for both schemes presented above, i.e., the 6-point and 8-point algorithms. In both cases we exclude orthographic projection and assume only perspective projection.

In the perspective case, the second reference plane is the image plane of the first model view, and the transformation for projecting the second reference plane onto any other view is the rotational component of camera motion (rigid transformation). We recover the rotational component of camera motion by adopting a result derived by Lee (1988), who shows that the rotational component of motion can be uniquely determined from two corresponding points and the corresponding epipoles. We then show that projective structure can be uniquely determined, up to a uniform scale factor, from two calibrated perspective images.

Proposition 3 (Lee, 1988) *In the case of perspective projection, the rotational component of camera motion can be uniquely recovered, up to a reflection, from two corresponding points and the corresponding epipoles. The reflection component can also be uniquely determined by using a third corresponding point.*

Proof: Let $l'_j = p'_j \times V_l$ and $l_j = p_j \times V_r$, $j = 1, 2$ be two corresponding epipolar lines. Because R is an orthogonal matrix, it leaves vector magnitudes unchanged, and we can normalize the length of l'_1, l'_2, V_l to be of the same length of l_1, l_2, V_r, respectively. We have therefore, $l'_j = Rl_j$, $j = 1, 2$, and $V_l = RV_r$, which is sufficient for determining R up to a reflection. Note that because R is a rigid transformation, it is both an epipolar and an induced epipolar transformation (the induced transformation E is determined by $E = (R^{-1})^t$, therefore $E = R$ because R is an orthogonal matrix).

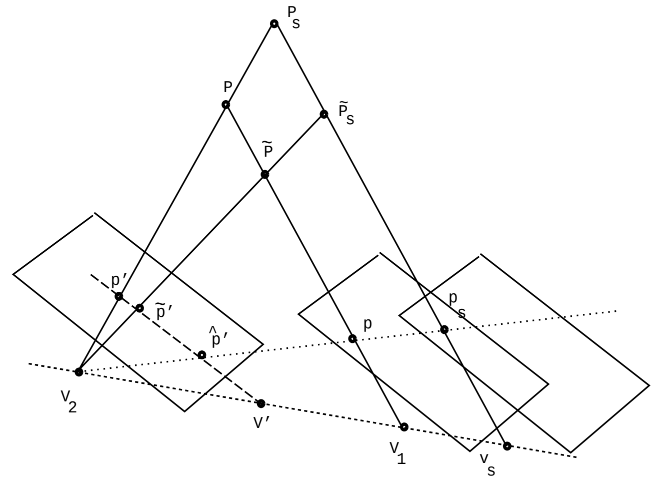

Figure 5: Illustration that projective shape can be recovered only up to a uniform scale (see text).

To determine the reflection component, it is sufficient to observe a third corresponding point $p_3 \longleftrightarrow p'_3$. The object point P_3 is along the ray $V_1 p_3$ and therefore has the coordinates $\alpha_3 p_3$ (w.r.t. the first camera coordinate frame), and is also along the ray $V_2 p'_3$ and therefore has the coordinates $\alpha'_3 p'_3$ (w.r.t. the second camera coordinate frame). We note that the ratio between α_3 and α'_3 is a positive number. The change of coordinates is represented by:

$$\beta V_r + \alpha_3 R p_3 = \alpha'_3 p'_3,$$

where β is an unknown constant. If we multiply both sides of the equation by l'_j, $j = 1, 2, 3$, the term βV_r drops out, because V_r is incident to all left epipolar lines, and after substituting l_j^t with ${l'_j}^t R$, we are left with,

$$\alpha_3 l_j^t \cdot p_3 = \alpha'_3 {l'_j}^t \cdot p'_3,$$

which is sufficient for determining the sign of l'_j. □

The rotation matrix R can be uniquely recovered from any three corresponding points and the corresponding epipoles. Projective structure can be reconstructed by replacing the transformation E of the second reference plane, with the rigid transformation R (which is equivalent to treating the first image plane as a reference plane). We show next that this can lead to projective structure up to an unknown uniform scale factor (unlike the non-rigid camera case).

Proposition 4 *In the perspective case, the projective shape constant α_p can be determined, from two views, at most up to a uniform scale factor.*

Proof: Consider Figure 5, and let the effective translation be $V_2 - V_s = k(V_2 - V_1)$, which is the true translation scaled by an unknown factor k. Projective shape, α_p, remains fixed if the scene and the focal length of the first view are scaled by k: from similarity of triangles we have,

$$k = \frac{V_s - V_2}{V_1 - V_2} = \frac{p_s - V_s}{p - V_1} = \frac{f_s}{1} = \frac{P_s - V_s}{P - V_1} = \frac{P_s - V_2}{P - V_2}$$

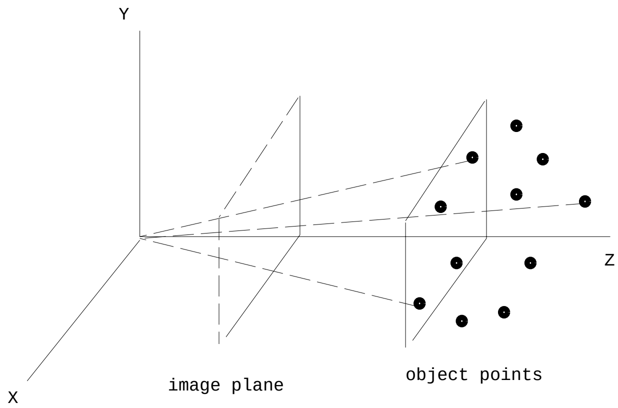

Figure 6: The basic object configuration for the experimental set-up.

where f_s is the scaled focal length of the first view. Since the magnitude of the translation along the line V_1V_2 is irrecoverable, we can assume it is null, and compute α_p as the cross-ratio of p', Ap, Rp, V_l which determines projective structure up to a uniform scale. □

Because α_p is determined up to a uniform scale, we need an additional point in order to establish a common scale during the process of re-projection (we can use one of the existing six or eight points we already have). We obtain, therefore, the following result:

Theorem 3 *In the perspective case, a rigid re-projection from two model views onto a novel view is possible, using four corresponding points coming from four non-coplanar points, and the corresponding epipoles. The projective structure computed from two perspective images, is invariant up to an overall scale factor.*

Orthographic projection is excluded from this result because it is well known that the rotational component cannot be uniquely determined from two orthographic views (Ullman 1979, Huang and Lee 1989, Aloimonos and Brown 1989). To see what happens in the case of parallel projection note that the epipoles are vectors on the xy plane of their coordinate systems (ideal points), and the epipolar lines are two vectors perpendicular to the epipole vectors. The equation $RV_r = V_l$ takes care of the rotation in plane (around the optical axis). The other two equations $Rl_j = l'_j$, $j = 1, 2$, take care only of rotation around the epipolar direction — rotation around an axis perpendicular to the epipolar direction is not accounted for. The equations for solving for R provide a non-singular system of equations but do produce a rotation matrix with no rotational components around an axis perpendicular to the epipolar direction.

10 Simulation Results Using Synthetic Objects

We ran simulations using synthetic objects to illustrate the re-projection process using the 6-point scheme under various imaging situations. We also tested the robustness of the re-projection method under various types of noise. Because the 6-point scheme requires that four of the corresponding points be projected from four coplanar points in space, it is of special interest to see how the method behaves under conditions that violate this assumption, and under noise conditions in general. The stability of the 8-point algorithm largely depends on the method for recovering the epipoles. The method adopted from Faugeras (1992), described in Section 8, based on the fundamental matrix, tends to be very sensitive to noise if the minimal number of points (eight points) are used. We have, therefore, focused the experimental error analysis on the 6-point scheme.

Figure 6 illustrates the experimental set-up. The object consists of 26 points in space arranged in the following manner: 14 points are on a plane (reference plane) ortho-parallel to the image plane, and 12 points are out of the reference plane. The reference plane is located two focal lengths away from the center of projection (focal length is set to 50 units). The depth of out-of-plane points varies randomly between 10 to 25 units away from the reference plane. The x, y coordinates of all points, except the points $P_1, ..., P_6$, vary randomly between 0 — 240. The 'privileged' points $P_1, ..., P_6$ have x, y coordinates that place these points all around the object (clustering privileged points together will inevitably contribute to instability).

The first view is simply a perspective projection of the object. The second view is a result of rotating the object around the point $(128, 128, 100)$ with an axis of rotation described by the unit vector $(0.14, 0.7, 0.7)$ by an angle of 29 degrees, followed by a perspective projection (note that rotation about a point in space is equivalent to rotation about the center of projection followed by translation). The third (novel) view is constructed in a similar manner with a rotation around the unit vector $(0.7, 0.7, 0.14)$ by an angle of 17 degrees. Figure 7 (first row) displays the three views. Also in Figure 7 (second row) we show the result of applying the transformation due to the four coplanar points $p_1, ..., p_4$ (Step 1, see Section 7.1) to all points in the first view. We see that all the coplanar points are aligned with their corresponding points in the second view, and all other points are situated along epipolar lines. The display on the right in the second row shows the final re-projection result (8-point and 6-point methods produce the same result). All points re-projected from the two model views are accurately (noise-free experiment) aligned with their corresponding points in the novel view.

The third row of Figure 7 illustrates a more challenging imaging situation (still noise-free). The second view is orthographically projected (and scaled by 0.5) following the same rotation and translation as before, and the novel view is a result of a central projection onto a tilted image plane (rotated by 12 degrees around a coplanar axis parallel to the x-axis). We have therefore the situation of recognizing a non-rigid perspective projection from a novel viewing position, given a rigid perspective projection and a rigid orthographic projection from two model viewing positions. The 6-point re-projection scheme was applied with the result that all re-projected points are in accurate alignment with their corresponding points in the novel view. Identical results were ob-

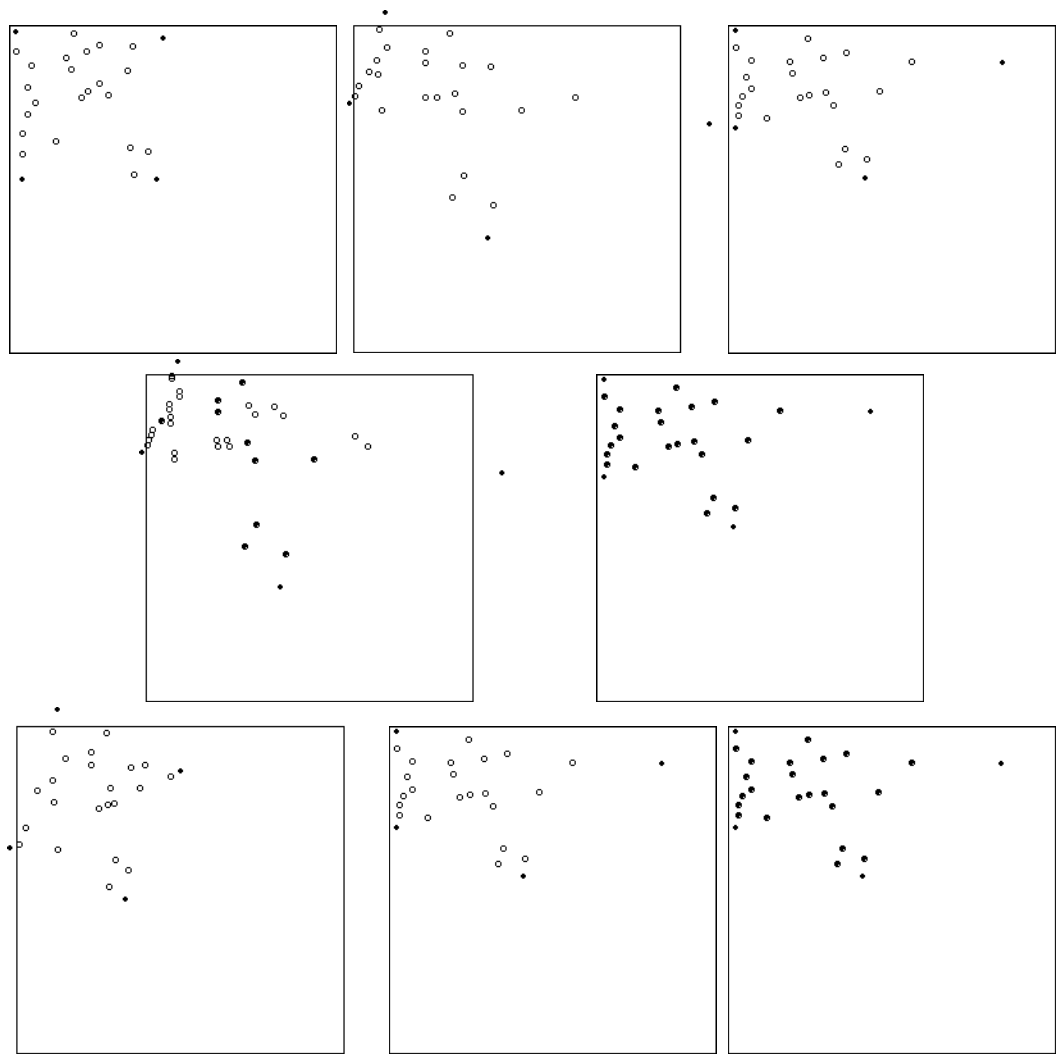

Figure 7: Illustration of Re-projection. *Row 1 (left to right):* Three views of the object, two model views and a novel view, constructed by rigid motion following perspective projection. The filled dots represent $p_1, ..., p_4$ (coplanar points). *Row 2:* Overlay of the second view and the first view following the transformation due to the reference plane (Step 1, Section 7.1). All coplanar points are aligned with their corresponding points, the remaining points are situated along epipolar lines. The righthand display is the result of re-projection — the re-projected image perfectly matches the novel image (noise-free situation). *Row 3:* The lefthand display shows the second view which is now orthographic. The middle display shows the third view which is now a perspective projection onto a tilted image plane. The righthand display is the result of re-projection which perfectly matches the novel view.

served with the 8-point algorithms.

The remaining experiments, discussed in the following sections, were done under various noise conditions. We conducted three types of experiments. The first experiment tested the stability under the situation where $P_1, ..., P_4$ are non-coplanar object points. The second experiment tested stability under random noise added to all image points in all views, and the third experiment tested stability under the situation that less noise is added to the privileged six points, than to other points.

10.1 Testing Deviation from Coplanarity

In this experiment we investigated the effect of translating P_1 along the optical axis (of the first camera position) from its initial position on the reference plane ($z = 100$) to the farthest depth position ($z = 125$), in increments of one unit at a time. The experiment was conducted using several objects of the type described above (the six privileged points were fixed, the remaining points were assigned random positions in space in different trials), undergoing the same motion described above (as in Figure 7, first row). The effect of depth translation to the level $z = 125$ on the location of p_1 is a shift of 0.93 pixels, on p_1' is 1.58 pixels, and on the location of p_1'' is 3.26 pixels. Depth translation is therefore equivalent to perturbing the location of the projections of P_1 by various degrees (depending on the 3D motion parameters).

Figure 8 shows the average pixel error in re-projection over the entire range of depth translation. The average pixel error was measured as the average of deviations from the re-projected point to the actual location of the corresponding point in the novel view, taken over all points. Figure 8 also displays the result of re-projection for the case where P_1 is at $z = 125$. The average error is 1.31, and the maximal error (the point with the most deviation) is 7.1 pixels. The alignment between the re-projected image and the novel image is, for the most part, fairly accurate.

10.2 Situation of Random Noise to all Image Locations

We next add random noise to all image points in all three views (P_1 is set back to the reference plane). This experiment was done repeatedly over various degrees of noise and over several objects. The results shown here have noise between 0–1 pixels randomly added to the x and y coordinates separately. The maximal perturbation is therefore $\sqrt{2}$, and because the direction of perturbation is random, the maximal error in relative location is double, i.e., 2.8 pixels. Figure 9 shows the average pixel errors over 10 trials (one particular object, the same motion as before). The average error fluctuates around 1.6 pixels. Also shown is the result of re-projection on a typical trial with average error of 1.05 pixels, and maximal error of 5.41 pixels. The match between the re-projected image and the novel image is relatively good considering the amount of noise added.

10.3 Random Noise Case 2

A more realistic situation occurs when the magnitude of noise associated with the privileged six points is much lower than the noise associated with other points, for the reason that we are interested in tracking points of interest that are often associated with distinct intensity structure (such as the tip of the eye in a picture of a face). Correlation methods, for instance, are known to perform much better on such locations, than on areas having smooth intensity change, or areas where the change in intensity is one-dimensional. We therefore applied a level of 0–0.3 perturbation to the x and y coordinates of the six points, and a level of 0–1 to all other points (as before). The results are shown in Figure 10. The average pixel error over 10 trials fluctuates around 0.5 pixels, and the re-projection shown for a typical trial (average error 0.52, maximal error 1.61) is in relatively good correspondence with the novel view. With larger perturbations at a range of 0–2, the algorithm behaves proportionally well, i.e., the average error over 10 trials is 1.37.

11 Summary

In this paper we focused on the problem of recovering relative, non-metric, structure from two views of a 3D object. Specifically, the invariant structure we recover does not require internal camera calibration, does not involve full reconstruction of shape (Euclidean or projective coordinates), and treats parallel and central projection as an integral part of one unified system. We have also shown that the invariant can be used for the purposes of visual recognition, within the framework of the alignment approach to recognition.

The study is based on an extension of Koenderink and Van Doorn's representation of affine structure as an invariant defined with respect to a reference plane and a reference point. We first showed that the KV affine invariant cannot be extended directly to a projective invariant (Appendix D), but there exists another affine invariant, described with respect to two reference planes, that can easily be extended to projective space. As a result we obtained the projective structure invariant.

We have shown that the difference between the affine and projective case lie entirely in the location of epipoles, i.e., given the location of epipoles both the affine and projective structure are constructed from the same information captured by four corresponding points projected from four non-coplanar points in space. Therefore, the additional corresponding points in the projective case are used solely for recovering the location of epipoles.

We have shown that the location of epipoles can be recovered under both parallel and central projection using six corresponding points, with the assumption that four of those points are projected from four coplanar points in space, or alternatively by having eight corresponding points without assumptions on coplanarity. The overall method for reconstructing projective structure and achieving re-projection was referred to as the 6-point and the 8-point algorithms. These algorithms have the unique property that projective structure can be recovered from both orthographic and perspective images from uncalibrated cameras. This property implies, for instance, that we can perform recognition of a perspective image of an object given two orthographic images as

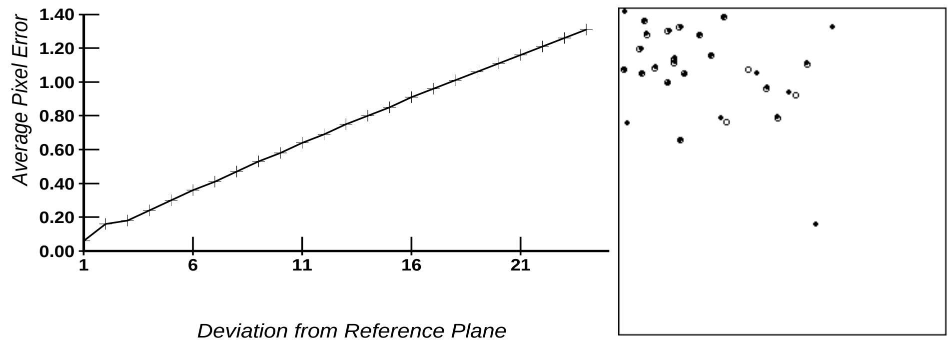

Figure 8: Deviation from coplanarity: average pixel error due to translation of P_1 along the optical axis from $z = 100$ to $z = 125$, by increments of one unit. The result of re-projection (overlay of re-projected image and novel image) for the case $z = 125$. The average error is 1.31 and the maximal error is 7.1.

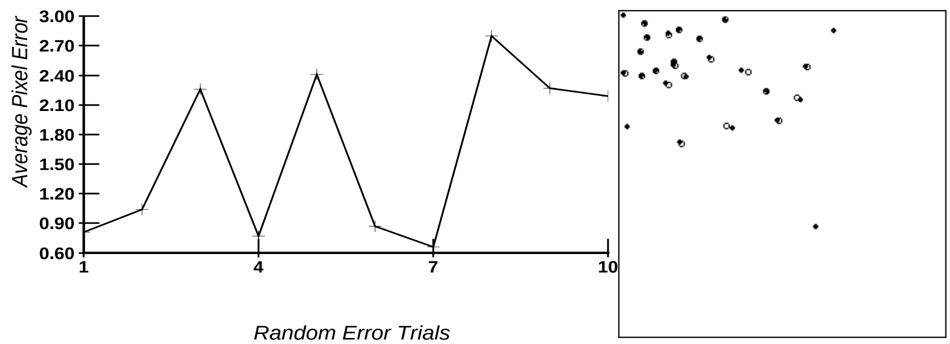

Figure 9: Random noise added to all image points, over all views, for 10 trials. Average pixel error fluctuates around 1.6 pixels. The result of re-projection on a typical trial with average error of 1.05 pixels, and maximal error of 5.41 pixels.

a model. It also implies greater stability because the size of the field of view is no longer an issue in the process of reconstructing shape or performing re-projection.

Acknowledgments

I want to thank David Jacobs and Shimon Ullman for discussions and comments on this work.

A Fundamental Theorem of Plane Projectivity

The fundamental theorem of plane projectivity states that a projective transformation of the plane is completely determined by four corresponding points. We prove the theorem by first using a geometric drawing, and then algebraically by introducing the concept of rays (homogeneous coordinates). The appendix ends with the system of linear equations for determining the correspondence of all points in the plane, given four corresponding points (used repeatedly throughout this paper).

Definitions: A *perspectivity* between two planes is defined as a central projection from one plane onto the other. A *projectivity* is defined as made out of a finite sequence of perspectivities. A projectivity, when represented in an algebraic form, is called a *projective transformation*. The fundamental theorem states that a projectivity is completely determined by four corresponding points.

Geometric Illustration

Consider the geometric drawing in Figure 11. Let A, B, C, U be four coplanar points in the scene, and let A', B', C', U' be their projection in the first view, and A'', B'', C'', U'' be their projection in the second view. By construction, the two views are projectively related to each other. We further assume that no three of the points are collinear (four points form a quadrangle), and without loss of generality let U be located within the triangle ABC. Let BC be the x-axis and BA be the y-axis. The projection of U onto the x-axis, denoted by U_x, is the intersection of the line AU with the x-axis. Similarly U_y is the intersection of the line CU with the y-axis. because straight lines project onto straight lines, we have that U_x, U_y correspond to U'_x, U'_y if and only if U corresponds to U'. For any other point P, coplanar with $ABCU$ in space, its coordinates P_x, P_y are constructed in a similar manner. We therefore have that B, U_x, P_x, C are collinear and therefore the cross ratio must be equal to the cross ratio of B', U'_x, P'_x, C', i.e.

$$\frac{BC \cdot U_x P_x}{BP_x \cdot U_x C} = \frac{B'C' \cdot U'_x P'_x}{B'P'_x \cdot U'_x C'}.$$

This form of cross ratio is known as the *canonical cross*

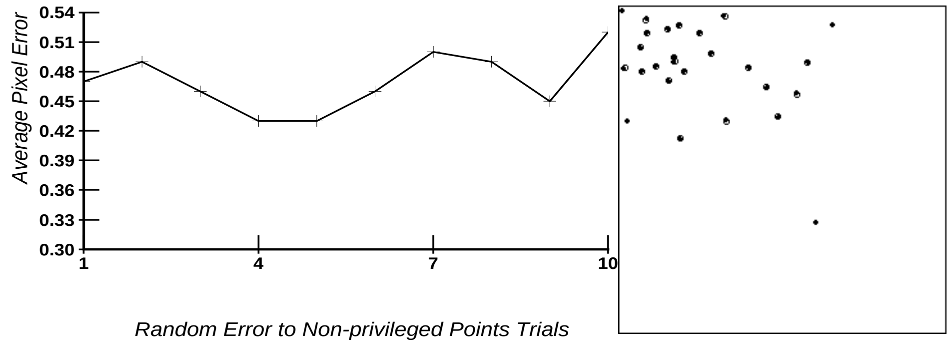

Figure 10: Random noise added to non-privileged image points, over all views, for 10 trials. Average pixel error fluctuates around 0.5 pixels. The result of re-projection on a typical trial with average error of 0.52 pixels, and maximal error of 1.61 pixels.

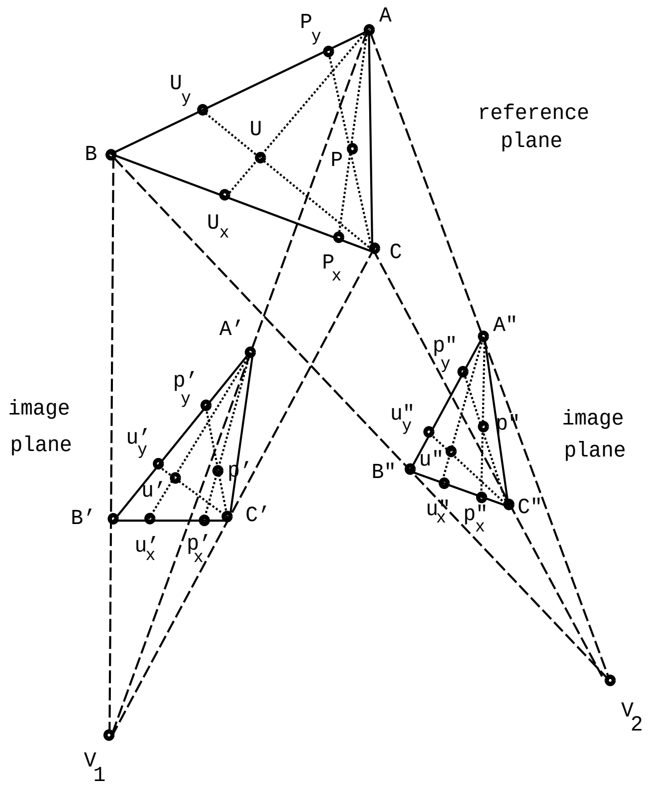

Figure 11: The geometry underlying plane projectivity from four points.

ratio. In general there are 24 cross ratios, six of which are numerically different (see Appendix B for more details on cross-ratios). Similarly, the cross ratio along the y-axis of the reference frame is equal to the cross ratio of the corresponding points in both views.

Therefore, for any point p' in the first view, we construct its x and y locations, p'_x, p'_y, along $B'C'$ and $B'A'$, respectively. From the equality of cross ratios we find the locations of p''_x, p''_y, and that leads to p''. Because we have used only projective constructions, i.e. straight lines project to straight lines, we are guaranteed that p' and p'' are corresponding points.

Algebraic Derivation

From an algebraic point of view it is convenient to view points as laying on rays emanating from the center of projection. A ray representation is also called the *homogeneous coordinates* representation of the plane, and is achieved by adding a third coordinate. Two vectors represent the same point $X = (x, y, z)$ if they differ at most by a scale factor (different locations along the same ray). A key result, which makes this representation amenable to application of linear algebra to geometry, is described in the following proposition:

Proposition 5 *A projectivity of the plane is equivalent to a linear transformation of the homogeneous representation.*

The proof is omitted here, and can be found in Tuller (1967, Theorems 5.22, 5.24). A projectivity is equivalent, therefore, to a linear transformation applied to the rays. Because the correspondence between points and coordinates is not one-to-one, we have to take scalar factors of proportionality into account when representing a projective transformation. An arbitrary projective transformation of the plane can be represented as a non-singular linear transformation (also called *collineation*) $\rho X' = TX$, where ρ is an arbitrary scale factor.

Given four corresponding rays $p_j = (x_j, y_j, 1) \longleftrightarrow p'_j = (x'_j, y'_j, 1)$, we would like to find a linear transformation T and the scalars ρ_j such that $\rho_j p'_j = T p_j$. Note that because only ratios are involved, we can set $\rho_4 = 1$. The following are a basic lemma and theorem adapted from Semple and Kneebone (1952).

Lemma 1 *If $p_1, ..., p_4$ are four vectors in R^3, no three of which are linearly dependent, and if $e_1, ..., e_4$ are respectively the vectors $(1,0,0), (0,1,0), (0,0,1), (1,1,1)$, there exists a non-singular linear transformation A such that $Ae_j = \lambda_j p_j$, where the λ_j are non-zero scalars; and the matrices of any two transformations with this property differ at most by a scalar factor.*

Proof: Let p_j have the components $(x_j, y_j, 1)$, and without loss of generality let $\lambda_4 = 1$. The matrix A satisfies three conditions $Ae_j = \lambda_j p_j$, $j = 1, 2, 3$ if and only if $\lambda_j p_j$ is the j'th column of A. Because of the fourth condition, the values $\lambda_1, \lambda_2, \lambda_3$ satisfy

$$[p_1, p_2, p_3] \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = p_4$$

and since, by hypothesis of linear independence of p_1, p_2, p_3, the matrix $[p_1, p_2, p_3]$ is non-singular, the λ_j are uniquely determined and non-zero. The matrix A is therefore determined up to a scalar factor. $\square$

Theorem 4 *If $p_1, ..., p_4$ and $p'_1, ..., p'_4$ are two sets of four vectors in R^3, no three vectors in either set being linearly dependent, there exists a non-singular linear transformation T such that $Tp_j = \rho_j p'_j$ $(j = 1, ..., 4)$, where the ρ_j are scalars; and the matrix T is uniquely determined apart from a scalar factor.*

Proof: By the lemma, we can solve for A and λ_j that satisfy $Ae_j = \lambda_j p_j$ $(j = 1, ..., 4)$, and similarly we can choose B and μ_j to satisfy $Be_j = \mu_j p'_j$; and without loss of generality assume that $\lambda_4 = \mu_4 = 1$. We then have, $T = BA^{-1}$ and $\rho_j = \frac{\mu_j}{\lambda_j}$. If, further, $Tp_j = \rho_j p'_j$ and $Up_j = \sigma_j p'_j$, then $TAe_j = \rho_j \lambda_j p'_j$ and $UAe_j = \sigma_j \lambda_j p'_j$; and therefore, by the lemma, $TA = \tau UA$, i.e., $T = \tau U$ for some scalar τ. $\square$

The immediate implication of the theorem is that one can solve directly for T and ρ_j $(\rho_4 = 1)$. Four points provide twelve equations and we have twelve unknowns (nine for T and three for ρ_j). Furthermore, because the system is linear, one can look for a least squares solution by using more than four corresponding points (they all have to be coplanar): each additional point provides three more equations and one more unknown (the ρ associated with it).

Alternatively, one can eliminate ρ_j from the equations, set $T_{3,3} = 1$ and set up directly a system of eight linear equations as follows. In general we have four corresponding rays $p_j = (x_j, y_j, z_j) \longleftrightarrow p'_j = (x'_j, y'_j, z'_j)$, $j = 1, ..., 4$, and the linear transformation T satisfies $\rho_j p'_j = Tp_j$. By eliminating ρ_j, each pair of corresponding rays contributes the following two linear equations:

$$x_j t_{1,1} + y_j t_{1,2} + z_j t_{1,3} - \frac{x_j x'_j}{z'_j} t_{3,1} - \frac{y_j x'_j}{z'_j} t_{3,2} = \frac{z_j x'_j}{z'_j}$$

$$x_j t_{2,1} + y_j t_{2,2} + z_j t_{2,3} - \frac{x_j y'_j}{z'_j} t_{3,1} - \frac{y_j y'_j}{z'_j} t_{3,2} = \frac{z_j y'_j}{z'_j}$$

A similar pair of equations can be derived in the case $z'_j = 0$ (ideal points) by using either x'_j or y'_j (all three cannot be zero).

Projectivity Between Two image Planes of an Uncalibrated Camera

We can use the fundamental theorem of plane projectivity to recover the projective transformation that was illustrated geometrically in Figure 11. Given four corresponding points $(x_j, y_j) \longleftrightarrow (x'_j, y'_j)$ that are projected from four coplanar points in space we would like to find the projective transformation A that accounts for all other correspondences $(x, y) \longleftrightarrow (x', y')$ that are projected from coplanar points in space.

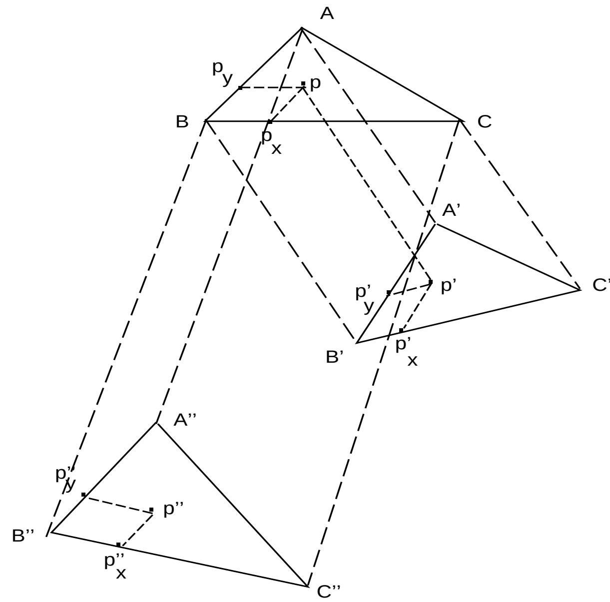

Figure 12: Setting a projectivity under parallel projection.

The standard way to proceed is to assume that both image planes are parallel to their xy plane with a focal length of one unit, or in other words to embed the image coordinates in a 3D vector whose third component is 1. Let $p_j = (x_j, y_j, 1)$ and $p'_j = (x'_j, y'_j, 1)$ be the the chosen representation of image points. The true coordinates of those image points may be different (if the image plane are in different positions than assumed), but the main point is that all such representations are projectively equivalent to each other. Therefore, $\rho_j p_j = B\hat{p}_j$ and $\mu_j p'_j = C\hat{p}'_j$, where $\hat{p}_j$ and $\hat{p}'_j$ are the true image coordinates of these points. If T is the projective transformation determined by the four corresponding points $\hat{p}_j \longleftrightarrow \hat{p}'_j$, then $A = CTB^{-1}$ is the projective transformation between the assumed representations $p_j \longleftrightarrow p'_j$.

Therefore, the matrix A can be solved for directly from the correspondences $p_j \longleftrightarrow p'_j$ (the system of eight equations detailed in the previous section). For any given point $p = (x, y, 1)$, the corresponding point $p' = (x', y', 1)$ is determined by Ap followed by normalization to set the third component back to 1.

A.1 Plane Projectivity in Affine Geometry

In parallel projection we can take advantage of the fact that parallel lines project to parallel lines. This allows to define coordinates on the plane by subtending lines parallel to the axes (see Figure 12). Note also that the two trapezoids $BB'p'_x p_x$ and $BB'C'C$ are similar trapezoids, therefore,

$$\frac{BC}{p_x C} = \frac{B'C'}{p'_x C'}.$$

This provides a geometric derivation of the result that three points are sufficient to set up a projectivity between any two planes under parallel projection.

Algebraically, a projectivity of the plane can be

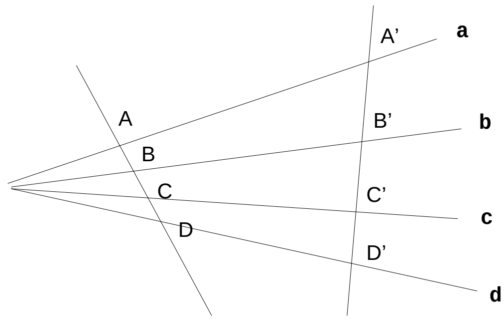

Figure 13: The cross-ratio of four distinct concurrent rays is equal to the cross-ratio of the four distinct points that result from intersecting the rays by a transversal.

uniquely represented as a 2D affine transformation of the non-homogeneous coordinates of the points. Namely, if $p = (x, y)$ and $p' = (x', y')$ are two corresponding points, then

$$p' = Ap + w$$

where A is a non-singular matrix and w is a vector. The six parameters of the transformation can be recovered from two non-collinear sets of three points, p_o, p_1, p_2 and p'_o, p'_1, p'_2. Let

$$A = \begin{bmatrix} x'_1 - x'_o, x'_2 - x'_o \\ y'_1 - y'_o, y'_2 - y'_o \end{bmatrix} \begin{bmatrix} x_1 - x_o, x_2 - x_o \\ y_1 - y_o, y_2 - y_o \end{bmatrix}^{-1}$$

and $w = p'_o - Ap_o$, which together satisfy $p'_j - p'_o = A(p_j - p_o)$ for $j = 1, 2$. For any arbitrary point p on the plane, we have that p is spanned by the two vectors $p_1 - p_o$ and $p_2 - p_o$, i.e., $p = \alpha_1(p_1 - p_o) + \alpha_2(p_2 - p_o)$; and because translation in depth is lost in parallel projection, we have that $p' = \alpha_1(p'_1 - p'_o) + \alpha_2(p'_2 - p'_o)$, and therefore $p' - p'_o = A(p - p_o)$.

B Cross-Ratio and the Linear Combination of Rays

The cross-ratio of four collinear points A, B, C, D is preserved under central projection and is defined as:

$$\alpha = \frac{AB}{AC} \div \frac{DB}{DC} = \frac{A'B'}{A'C'} \div \frac{D'B'}{D'C'},$$

(see Figure 13). All permutations of the four points are allowed, and in general there are six distinct cross-ratios that can be computed from four collinear points. Because the cross-ratio is invariant to projection, any transversal meeting four distinct concurrent rays in four distinct points will have the same cross ratio — therefore one can speak of the cross-ratio of rays (concurrent or parallel) $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}$.

The cross-ratio result in terms of rays, rather than points, is appealing for the reasons that it enables the application of linear algebra (rays are represented as points in homogeneous coordinates), and more important, enables us to treat ideal points as any other point (critical for having an algebraic system that is well defined under both central and parallel projection).

The cross-ratio of rays is computed algebraically through linear combination of points in homogeneous coordinates (see Gans 1969, pp. 291–295), as follows. Let the the rays $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}$ be represented by vectors $(a_1, a_2, a_3), ..., (d_1, d_2, d_3)$, respectively. We can represent the rays $\boldsymbol{a}, \boldsymbol{d}$ as a linear combination of the rays $\boldsymbol{b}, \boldsymbol{c}$, by

$$\boldsymbol{a} = \boldsymbol{b} + k\boldsymbol{c}$$

$$\boldsymbol{d} = \boldsymbol{b} + k'\boldsymbol{c}$$

For example, k can be found by solving the linear system of three equation $\rho\boldsymbol{a} = \boldsymbol{b} + k\boldsymbol{c}$ with two unknowns ρ, k (one can solve using any two of the three equations, or find a least squares solution using all three equations). We shall assume, first, that the points are Euclidean. The ratio in which A divides the line BC can be derived by:

$$\frac{AB}{AC} = \frac{\frac{a_1}{a_3} - \frac{b_1}{b_3}}{\frac{a_1}{a_3} - \frac{c_1}{c_3}} = \frac{\frac{b_1 + kc_1}{b_3 + kc_3} - \frac{b_1}{b_3}}{\frac{b_1 + kc_1}{b_3 + kc_3} - \frac{c_1}{c_3}} = -k\frac{c_3}{b_3}$$

Similarly, we have $\frac{DB}{DC} = -k'\frac{c_3}{b_3}$ and, therefore, the cross-ratio of the four rays is $\alpha = \frac{k}{k'}$. The same result holds under more general conditions, i.e., points can be ideal as well:

Proposition 6 *If A, B, C, D are distinct collinear points, with homogeneous coordinates $\boldsymbol{b} + k\boldsymbol{c}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{b} + k'\boldsymbol{c}$, then the canonical cross-ratio is $\frac{k}{k'}$.*

(for a complete proof, see Gans 1969, pp. 294–295). For our purposes it is sufficient to consider the case when one of the points, say the vector $\boldsymbol{d}$, is ideal (i.e. $d_3 = 0$). From the vector equation $\rho\boldsymbol{d} = \boldsymbol{b} + k'\boldsymbol{c}$, we have that $k' = -\frac{b_3}{c_3}$ and, therefore, the ratio $\frac{DB}{DC} = 1$. As a result, the cross-ratio is determined only by the first term, i.e., $\alpha = \frac{AB}{AC} = k$ — which is what we would expect if we represented points in the Euclidean plane and allowed the point D to extend to infinity along the line A, B, C, D (see Figure 13).

The derivation so far can be translated directly to our purposes of computing the projective shape constant by replacing $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}$ with $p', \tilde{p}', \hat{p}', V_l$, respectively.

C On Epipolar Transformations

Proposition 7 *The epipolar lines pV_r and $p'V_l$ are perspectively related.*

Proof: Consider Figure 14. We have already established that p projects onto the left epipolar line $p'V_l$. By definition, the right epipole V_r projects onto the left epipole V_l, therefore, because lines are projective invariants the line pV_r projects onto the line $p'V_l$. $\square$

The result that epipolar lines in one image are perspectively related to the epipolar lines in the other image, implies that there exists a projective transformation F that maps epipolar lines l_j onto epipolar lines l'_j, that is $Fl_j = \rho_j l'_j$, where $l_j = p_j \times V_r$ and $l'_j = p'_j \times V_l$. From the property of point/line duality of projective geometry (Semple and Kneebone, 1952), the transformation E that maps points on left epipolar lines onto points on the corresponding right epipolar lines is induced from F, i.e., $E = (F^{-1})^t$.

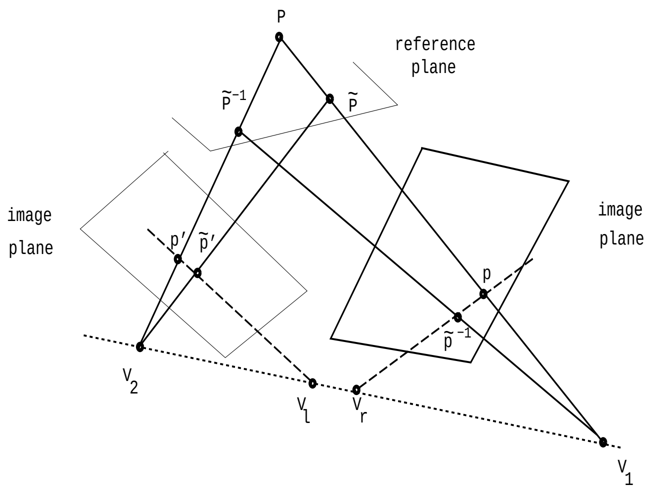

Figure 14: Epipolar lines are perspectively related.

Proposition 8 (point/line duality) *The transformation for projecting p onto the left epipolar line $p'V_l$, is $E = (F^{-1})^t$.*

Proof: Let l, l' be corresponding epipolar lines, related by the equation $\rho l' = Fl$. Let p, p' be any two points, one on each epipolar line (not necessarily corresponding points). From the point/line incidence axiom we have that $l^t \cdot p = 0$. By substituting l we have

$$\left[\rho F^{-1} l'\right]^t \cdot p = 0 \qquad \Longrightarrow \qquad \rho {l'}^t \cdot \left[F^{-t} p\right] = 0.$$

Therefore, the collineation $E = (F^{-1})^t$ maps points p onto the corresponding left epipolar line. □

It is intuitively clear that the epipolar line transformation F is not unique, and therefore the induced transformation E is not unique either. The correspondence between the epipolar lines is not disturbed under translation along the line V_1V_2, or under non-rigid camera motion that results from tilting the image plane with respect to the optical axis such that the epipole remains on the line V_1V_2.

Proposition 9 *The epipolar transformation F is not unique.*

Proof: A projective transformation is determined by four corresponding pencils. The transformation is unique (up to a scale factor) if no three of the pencils are linearly dependent, i.e., if the pencils are lines, then no three of the four lines should be coplanar. The epipolar line transformation F can be determined by the corresponding epipoles, $V_r \longleftrightarrow V_l$, and three corresponding epipolar lines $l_j \longleftrightarrow l'_j$. We show next that the epipolar lines are coplanar, and therefore, F cannot be determined uniquely.

Let p_j and p'_j, $j = 1, 2, 3$, be three corresponding points and let $l_j = p_j \times V_r$ and $l'_j = p'_j \times V_l$. Let $\bar{p}_3 = \alpha p_1 + \beta p_2$, $\alpha + \beta = 1$, be a point on the epipolar line $p_3 V_r$ collinear with p_1, p_2. We have,

$$l_3 = p_3 \times V_r = (a\bar{p}_3 + bV_r) \times V_r = a\bar{p}_3 \times V_r = a\alpha l_1 + a\beta l_2,$$

and similarly $l'_3 = \alpha' l'_1 + \beta' l'_2$. □

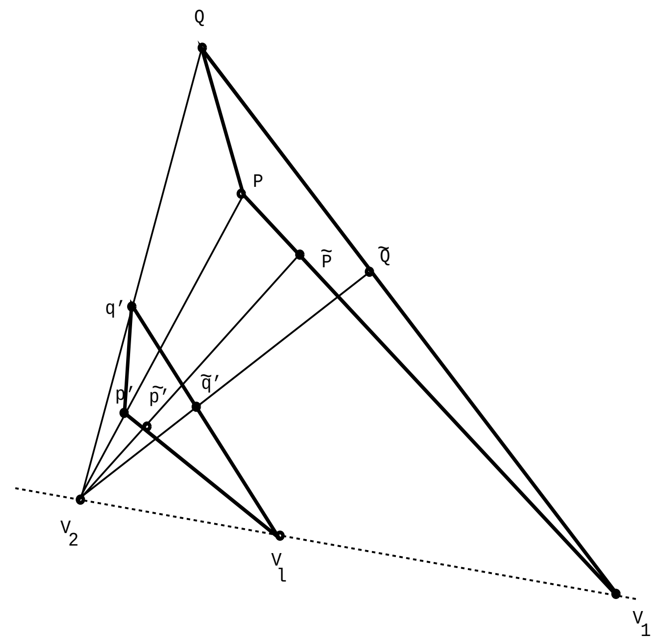

Figure 15: See text.

The epipolar transformation, therefore, has three free parameters (one for scale, the other two because the equation $Fl_3 = \rho_3 l_3$ has dropped out).

D Affine Structure in Projective Space

Proposition 10 *The affine structure invariant, based on a single reference plane and a reference point, cannot be directly extended to central projection.*

Proof: Consider the drawing in Figure 15. Let Q be the reference point, P be an arbitrary point of interest in space, and $\tilde{Q}, \tilde{P}$ be the projection of Q and P onto the reference plane (see section 4 for definition of affine structure under parallel projection).

The relationship between the points $P, Q, \tilde{P}, \tilde{Q}$ and the points $p', \tilde{p}', q', \tilde{q}'$ can be described as a perspectivity between two triangles. However, in order to establish an invariant between the two triangles one must have a coplanar point outside each of the triangles, therefore the five corresponding points are not sufficient for determining an invariant relation (this is known as the 'five point invariant' which requires that no three of the points be collinear). □

E On the Intersection of Epipolar Lines

Barret *et al.* (1991) derive a quadratic invariant based on Longuet-Higgins' fundamental matrix. We describe briefly their invariant and show that it is equivalent to performing re-projection using intersection of epipolar lines.

In section 8 we derived Longuet-Higgins' fundamental matrix relation ${p'}^t Hp = 0$. Barret *et al.* note that the equation can be written in vector form $\boldsymbol{h}^t \cdot \boldsymbol{q} = 0$, where $\boldsymbol{h}$ contains the elements of H and

$$\boldsymbol{q} = (x'x, x'y, x', y'x, y'y, y', x, y, 1).$$

Therefore, the matrix

$$B = \begin{bmatrix} \boldsymbol{q}_1 \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{q}_9 \end{bmatrix} \qquad (1)$$

must have a vanishing determinant. Given eight corresponding points, the condition $|B| = 0$ leads to a constraint line in terms of the coordinates of any ninth point, i.e., $\alpha x + \beta y + \gamma = 0$. The location of the ninth point in any third view can, therefore, be determined by intersecting the constraint lines derived from views 1 and 3, and views 2 and 3.

Another way of deriving this re-projection method is by first noticing that H is a correlation that maps p onto the corresponding epipolar line $l' = V_l \times p'$ (see section 8). Therefore, from views 1 and 3 we have the relation

$$p''^t \tilde{H} p = 0,$$

and from views 2 and 3 we have the relation

$$p''^t \hat{H} p' = 0,$$

where $\tilde{H}p$ and $\hat{H}p'$ are two intersecting epipolar lines. Given eight corresponding points, we can recover $\tilde{H}$ and $\hat{H}$. The location of any ninth point p'' can be recovered by intersecting the lines $\tilde{H}p$ and $\hat{H}p'$.

This way of deriving the re-projection method has an advantage over using the condition $|B| = 0$ directly, because one can use more than eight points in a least squares solution (via SVD) for the matrices $\tilde{H}$ and $\hat{H}$.

Approaching the re-projection problem using intersection of epipolar lines is problematic for novel views that have a similar epipolar geometry to that of the two model views (these are situations where the two lines $\tilde{H}p$ and $\hat{H}p'$ are nearly parallel, such as when the object rotates around nearly the same axis for all views). We therefore expect sensitivity to errors also under conditions of small separation between views. The method becomes more practical if one uses multiple model views instead of only two, because each model views adds one epipolar line and all lines should intersect at the location of the point of interest in the novel view.

References

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

[2] J. Aloimonos and C.M. Brown. On the kinetic depth effect. *Biological Cybernetics*, 60:445–455, 1989.

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

[4] E.B. Barrett, M.H. Brill, N.N. Haag, and P.M. Pyton. General methods for determining projective invariants in imagery. *Computer Vision, Graphics, and Image Processing*, 53:46–65, 1991.

[5] T. Broida, S. Chandrashekhar, and R. Chellapa. recursive 3-d motion estimation from a monocular image sequence. *IEEE Transactions on Aerospace and Electronic Systems*, 26:639–656, 1990.

[6] D.C. Brown. Close-range camera calibration. *Photogrammetric Engineering*, 37:855–866, 1971.

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

[8] W. Faig. Calibration of close-range photogrammetry systems: Mathematical formulation. *Photogrammetric Engineering and Remote Sensing*, 41:1479–1486, 1975.

[9] O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proceedings of the European Conference on Computer Vision*, pages 563–578, Santa Margherita Ligure, Italy, June 1992.

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

[11] O.D. Faugeras and S. Maybank. Motion from point matches: Multiplicity of solutions. *International Journal of Computer Vision*, 4:225–246, 1990.

[12] M.A. Fischler and R.C. Bolles. Random sample concensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24:381–395, 1981.

[13] D. Gans. *Transformations and Geometries*. Appleton-Century-Crofts, New York, 1969.

[14] E.C. Hildreth. Recovering heading for visually-guided navigation. A.I. Memo No. 1297, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, June 1991.

[15] T.S. Huang and C.H. Lee. Motion and structure from orthographic projections. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-11:536–540, 1989.

[16] D.P. Huttenlocher and S. Ullman. Object recognition using alignment. In *Proceedings of the International Conference on Computer Vision*, pages 102–111, London, December 1987.

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

[18] R.K. Lenz and R.Y. Tsai. Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology. In *fourth int. Symp. on Robotics Research*, pages 68–75, Santa Cruz, CA, Aug 1987.

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

[20] H.C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proceedings of the Royal Society of London B*, 208:385–397, 1980.

[21] D.G. Lowe. *Perceptual organization and visual recognition*. Kluwer Academic Publishing, Hingham, MA, 1985.

[22] R. Mohr, L. Quan, F. Veillon, and B. Boufama. Relative 3D reconstruction using multiple uncalibrated images. Technical Report RT 84-IMAG, LIFIA — IRIMAG, France, June 1992.

[23] J.H. Rieger and D.T. Lawton. Processing differential image motion. *Journal of the Optical Society of America*, 2:354–360, 1985.

[24] J.W. Roach and J.K. Aggarwal. Computer tracking of objects moving in space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1:127–135, 1979.

[25] J.G. Semple and G.T. Kneebone. *Algebraic Projective Geometry*. Clarendon Press, Oxford, 1952.

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

[27] C. Tomasi. *shape and motion from image streams: a factorization method*. PhD thesis, School of Computer Science, Carnagie Mellon University, Pittsburgh, PA, 1991.

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

[29] A. Tuller. *A modern introduction to geometries*. D. Van Nostrand Co., Princeton, NJ, 1967.

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

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

[32] S. Ullman and R. Basri. Recognition by linear combination of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-13:992—1006, 1991. Also in M.I.T AI Memo 1052, 1989.