Projective Reconstruction of Approximately Planar Scenes

Robert T. Collins

A version of this paper was presented at the 21st AIPR Workshop in Washington DC, October 14–16, 1992. A few minor editing changes have been made since then, and one reference to a paper that did not yet exist at that time has been added. The work described here was funded by DARPA/TACOM contract number DAAE07-91-C-R035, by NSF grant number CDA-8922572, and by the RADIUS project under DARPA/Army contract TEC DACA76-92-R-0028.

The full citation for the original paper is

> Robert T. Collins, "Projective Reconstruction of Approximately Planar Scenes," in *Interdisciplinary Computer Vision: An Exploration of Diverse Applications,* (21st AIPR Workshop), Jane Harmon, Editor, Proc. SPIE 1839, pp. 174–185.

Projective Reconstruction of Approximately Planar Scenes*

Robert T. Collins

Department of Computer Science
University of Massachusetts at Amherst
Amherst, MA. 01003-4610

Abstract

The fundamental theorem of projective geometry states that the transformation mapping coplanar points from an object onto an image plane can be determined given the correspondence between four or more object points and their projections in the image. This theorem can be used to predict where other features in the object plane will appear in the image, and conversely, to project new image features back onto the object plane. In this paper we examine what happens when these mathematical results are applied to real-world data containing random sensor noise and deviations from coplanarity.

1 Inferring Planar Structure

One of the main goals of computer vision is to infer three-dimensional scene structure from one or more two-dimensional images. Modeling the world in all its complexity is a daunting task, however. Even in the field of computer graphics, where complete knowledge of the scene to be displayed is presumed available, there are still unresolved issues in representing general curved surfaces and volumes, and the variety of textures found in natural scenes. Add to this the problem of efficiently recovering such representations from 2D images, and the task seems nearly intractable. One simplification is to focus on man-made domains where planar surfaces and linear surface markings predominate. The limitations of these domains are mitigated by the fact that useful applications exist where the planarity assumption does hold, such as indoor mobile robot navigation. Even in unrestricted environments, the world is sometimes flat enough locally to approximate by piecewise planar patches. Such is the case for city and campus navigation, and aerial photo-interpretation.

Picture-taking induces a mapping from the 3D world down to a 2D image. When a picture is taken of a planar surface, the mapping is from 2D to 2D. Part of the relevance of projective geometry for computer vision is that the mapping from planar surface coordinates to image coordinates can be described quite accurately by an invertible projective transformation. The fundamental theorem of projective geometry shows how to estimate this transformation given a set of just four known correspondences between object points and image points. The resulting transformation can be used to predict where other features in the object plane will appear in the image. More importantly, the inverse transformation can be used to project previously unseen image features back onto the object, thereby extending the object model.

The theorems of projective geometry were developed with mathematically precise objects in mind. In contrast, a practical vision system must deal with errorful measurements extracted from real image sensors. In this paper, we consider two sources of discrepancy between the idealized projective transformation model and its real-world applications. The first source is random observation error in positions of the image features due to sensor noise, which leads to errors in the planar reconstruction. We present a novel approach to modeling random errors in the projective plane. The second source of uncertainty considered is due to deviations of the object features from true coplanarity. This situation leads to a systematic error in predicted feature locations. We show that this systematic error can be harnessed to quantify the deviation of a set of features from coplanarity, allowing partial 3D reconstruction of scene features outside of the original reference plane. Examples from the domain of aerial image interpretation will be presented.

*This paper was presented at the 21st AIPR Workshop, Washington DC, October 14–16, 1992. This work was funded by DARPA/TACOM contract number DAAE07-91-C-R035, by NSF grant number CDA-8922572, and by the RADIUS project under DARPA/Army contract TEC DACA76-92-R-0028.

2 Projective Transformations

The importance of projective geometry for describing the image formation process is due to the pinhole camera model. Consider a left-handed camera coordinate system with focal point at the origin and focal axis pointing out along the positive Z axis, passing perpendicular through the image plane $Z = f$. The pinhole image of a scene point (x, y, z) is the point $(fx/z, fy/z, f)$ where a line through both the scene point and the origin intersects the image plane. The pinhole projection of a coplanar set of scene points onto an image is just one example of a much larger class of projective mappings. This section briefly summarizes properties of projective mappings between planes and their representation via homogeneous coordinates. Some of the relevant material can be found in [Mohr91, Faug88, Tsai82]. For a more comprehensive discussion of projective transformations, the reader is invited to consult a projective geometry text [Spri64].

2.1 Homographies

A general projective transformation between planes can be written algebraically as

$$X' = \frac{aX + bY + c}{gX + hY + i} \quad , \quad Y' = \frac{dX + eY + f}{gX + hY + i} \qquad (1)$$

where (X, Y) and (X', Y') are points represented in the 2D local coordinate systems of each plane. Unfortunately, this is a nonlinear transformation that is undefined when the denominator is zero, corresponding to a point mapping to infinity.

In order to make a projective transformation bijective, a line of points at infinity is explicitly added to each plane, to correspond to the cases where the denominator in (1) goes to zero. A plane that has been augmented in this way is a new geometric entity called the *projective plane*. The projective plane has a different global topology than the affine or Euclidean plane, and this has implications for the representation of observed points and their uncertainty. This topic is explored in Section 3.

Using homogeneous coordinates, infinite points can be manipulated the same as finite ones, and the transformation of equation (1) become linear

$$k \begin{bmatrix} X' \\ Y' \\ S' \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} X \\ Y \\ S \end{bmatrix} \qquad (2)$$

where k is a nonzero scalar, S and S' are 1 for finite points in the plane, and 0 for infinite points. Since homogeneous coordinates are equivalent up to scalar multiples, the transformation matrix can be multiplied by any nonzero constant and still represent the same mapping, and therefore has only 8 independent parameters. A nonsingular projective mapping that is linear in homogeneous coordinates is called a *homography*. Matrices representing homographies form a group under matrix multiplication and matrix inverse.

Because they are linear, invertible and closed under composition, homographies greatly simplify the analysis of projective mappings. Figure 1 shows a familiar computer vision scenario where pictures are taken of a single planar surface by two cameras from different viewpoints. Under the pinhole camera model, each

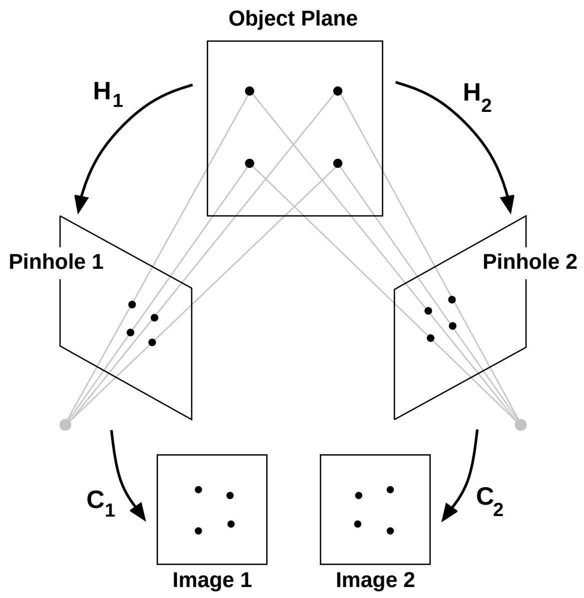

Figure 1: An object plane viewed by two cameras whose deviations from the pinhole camera model can be described by linear camera parameters. Corresponding points in any two planes in this diagram are related by a homography. Refer to the text.

camera effects a homography (in fact a *perspectivity*) from the object plane to a pinhole image plane, with all point correspondences lying on lines passing through the focal point. The perspectivities are labeled $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ in Figure 1. In actuality, the pinhole camera model does not adequately characterize images produced by real cameras. If the deviation of a particular camera from the pinhole model is governed by linear camera calibration parameters, the resulting image is an affine transformation of the pure pinhole projection [Horn86]. These transformations are labeled $\boldsymbol{C}_1$ and

C_2 in the figure. Affine transformations are a subgroup of the homography group, and the composition of a pinhole perspectivity followed by an affine deformation yields yet another homography. It is therefore easy to derive the transformation between any two planes in the diagram; for instance the transformation mapping points in image 1 into corresponding points in image 2 is $\boldsymbol{C}_2\boldsymbol{H}_2\boldsymbol{H}_1^{-1}\boldsymbol{C}_1^{-1}$.

Early research in the field studied the homography $\boldsymbol{H}_1\boldsymbol{H}_2^{-1}$ relating pinhole images of coplanar points. In [Faug88], Faugeras and Lustman show that the interframe point homography is proportional to

$$(d\boldsymbol{R} + \boldsymbol{T} n^t) \tag{3}$$

where n and d are the unit normal and perpendicular distance to the object plane, and $\boldsymbol{R}$ and $\boldsymbol{T}$ are the rotation matrix and translation vector bringing the two camera coordinate systems into alignment. They also show how to decompose the homography matrix to recover $\boldsymbol{R}$, n, and $\boldsymbol{T}/d$, up to a possible twofold ambiguity (see also [Tsai82]). More recent research has focused on the object plane to image plane mapping $\boldsymbol{C}_i\boldsymbol{H}_i$ and its inverse $\boldsymbol{H}_i^{-1}\boldsymbol{C}_i^{-1}$ which backprojects image plane points to their appropriate object plane positions regardless of camera location or linear distortion parameters. This backprojection lies at the core of model extension work by Mohr [Mohr91, Mohr90] and Collins [Coll92].

Unfortunately, some camera images are dominated by nonlinear lens distortions, and any analysis based on projective transformations becomes invalid. Under radial distortion, for instance, the images of colinear points may no longer be colinear, and the corresponding plane to plane mapping is no longer a homography. In the remainder of this paper nonlinear camera parameters are neglected; when such lens distortions are nonnegligable a preprocessing step must be performed to remove their effects [Gros90].

2.2 Estimation of a homography

The fundamental theorem of projective geometry states that a plane to plane homography is completely determined by the correspondences of four points, no three of which are colinear. In practice it is better to use as many point and line correspondences as possible to reduce errors in the estimated transformation caused by noise in the observed image data. Faugeras and Lustman present a least squares approach to homography estimation [Faug88]. Each finite point to point correspondence supplies two constraints on the eight independent parameters to be estimated. Using the notation of equation (1) these constraints are

$$\begin{aligned} aX + bY + c - X'(gX + hY + i) &= 0 \\ dX + eY + f - Y'(gX + hY + i) &= 0. \end{aligned}$$

Since possible solutions for the set of parameters are equivalent up to scalar multiples, a further constraint like $i = 1$ is imposed to provide a unique solution. One problem with the above constraints is that they are only valid when all the points are finite (S and S' in Equation 2 must not be zero). When one or both of the points in a correspondence are infinite, a modified pair of constraints should be used. Making sure that the constraint equations properly handle points at infinity is important, since a standard coordinatization of the projective plane with respect to four basis points involves computing a homography mapping finite points to infinite points [Spri64]. Points at infinity also arise in practice.

Points and lines are duals in the projective plane. Intuitively this means geometric constructions that are valid for points are also valid for lines. The homogeneous coordinate representation of a line in the projective plane is formed as the vector cross-product $\boldsymbol{p} \times \boldsymbol{q}$ of the homogeneous coordinates $\boldsymbol{p}$ and $\boldsymbol{q}$ of any two distinct points on the line. The result is another 3-place vector that can be scaled arbitrarily. A given homogeneous coordinate vector can therefore be interpreted either as a point or as a line. The principle of *duality* states that points and lines are indeed interchangeable, as long as the exchange is carried out systematically. For example, the dual of the fundamental theorem of projective geometry states that a plane-to-plane homography can also be determined by the correspondences of four corresponding lines, no three of which meet at a point.

3 Projective Data Fusion

In the last section it was shown that many useful plane to plane mappings can be represented by linear, invertible projective transformations called homographies, and that these transformations can be estimated given a small number of point or line correspondences between the two planes. When a transformation is estimated between two images of the same planar object, points in one image can be readily mapped to their corresponding locations in the other image, to aid the search for further correspondences. When an estimated transformation goes from object to image, the resulting homography can not only be used to predict where features in the object plane will appear in the image, but the inverse transformation that maps from image coordinates back into object coordinates allows an object model to be extended by adding new observed points and lines. These results provide powerful methods for inferring planar scene structure without first solving for camera motion, pose, or calibration parameters.

Figures 2 and 3 presents a concrete example of this type of approach. Figure 2a shows a near-nadir aerial photograph of a cultural site, taken at a high enough elevation that the whole scene can be considered planar. Extracting line segment features from the image (Figure 2c) is a useful first step towards producing a map of the site. After an initial map has been built, it can be extended to include larger areas of the scene by adding information from other images, as in a mosaic. Figures 2b and 2d show a second, oblique view of the same area, and its associated line segment features. The strong perspective distortion induced by the oblique angle seems to make this a poor candidate for extending the initial scene map. However, if at least four point or line correspondences between the two views can be found, a projective transformation mapping the second image into the coordinate frame of the first can be computed, allowing the oblique view to be effectively "unwarped" into registration with the initial nadir view (for an alternative approach, see [Coll93]). Figure 3 shows the final registration, as a mosaic of the extracted image line segments.

One real-world issue becomes immediately apparent from this example. If ideal features were being transformed between the two images, corresponding features would overlap exactly in the final registered mosaic. This does not occur in practice, of course, due to errors in the positions of extracted image features. What is needed is a method for representing uncertainty in the homogeneous coordinates of extracted points and lines, and for fusing multiple uncertain estimates when they are available. Each point or line feature in homogeneous coordinates represents a point in the projective plane; multiple noisy estimates of the same object feature form a sample of points in the projective plane, clustered around the point in the projective plane representing the homogeneous coordinates of the true object feature location. This section introduces a class of probability distributions for describing sample point clusters in the projective plane.

3.1 Probabilities in the Projective Plane

There are many ways to visualize the projective plane. In Section 2 the projective plane is described as the Euclidean plane augmented with a line of points at infinity. This is not the best way to visualize the projective plane, however, since the Euclidean plane is topologically open, while the projective plane is topologically closed. To understand the difference, consider a hypothetical traveler in the plane following a ray starting at the origin, going through point (x, y), and continuing out infinitely far. After finally "arriving" at infinity, the traveler is located at point $(x, y, 0)$ in homogeneous coordinates. But in homogenous coordinates this is the same point as $(-x, -y, 0)$, so the intrepid explorer can keep traveling "past infinity" in the same direction, eventually passing through point $(-x, -y)$ and finally returning to the origin.

Because of this wraparound effect, if the topology of the projective plane is ignored and it is treated as a Euclidean plane, a single cluster of points centered around a point at infinity will appear as two clusters infinitely far apart. Any estimation technique based on "averaging" these points in the Euclidean plane will produce bad results in this case. Proper handling of points at infinity is not just of theoretical interest. Such points **do arise** in practice. For instance, parallel lines in the world project in the image as lines that converge to a *vanishing point*. When the lines project to parallel lines in the image, the vanishing point is said to be at infinity. Since parallel image lines will normally be corrupted by errors, some line intersections will appear to be close to infinity in one direction, while some will appear to be infinitely far away in the opposite direction. A general-purpose algorithm for vanishing point analysis therefore needs to be able to handle clusters of vanishing point estimates centered around points at infinity [Coll90a].

Since the projective plane is topologically closed, it is better to think of it as a closed 2D space like the surface of a sphere. More formally, define an equivalence relation $\sim$ on $R^3 - \{(0,0,0)\}$ such that $(x_1, x_2, x_3) \sim (y_1, y_2, y_3)$ iff there exists some nonzero k such that $x_i = ky_i, i = 1, 2, 3$. The projective plane P^2 is defined as the the quotient space

$$(R^3 - \{(0,0,0)\}) \,/ \sim .$$

Viewing R^3 geometrically as Euclidean 3-space, each member of the quotient space is an equivalence class of points along an infinite line through the origin (excluding the origin itself which would otherwise need to be a member of all the equivalence classes). Consider now the surface of the unit sphere $S^2 = \{(x_1, x_2, x_3) | x_1^2 + x_2^2 + x_3^2 = 1\}$, and form the quotient space $S^2 / \sim$. Each equivalence class now contains one pair of diametrically opposite points. Equating equivalence classes in the obvious way shows the surface of the unit sphere with antipodal points equated is isomorphic to the projective plane.

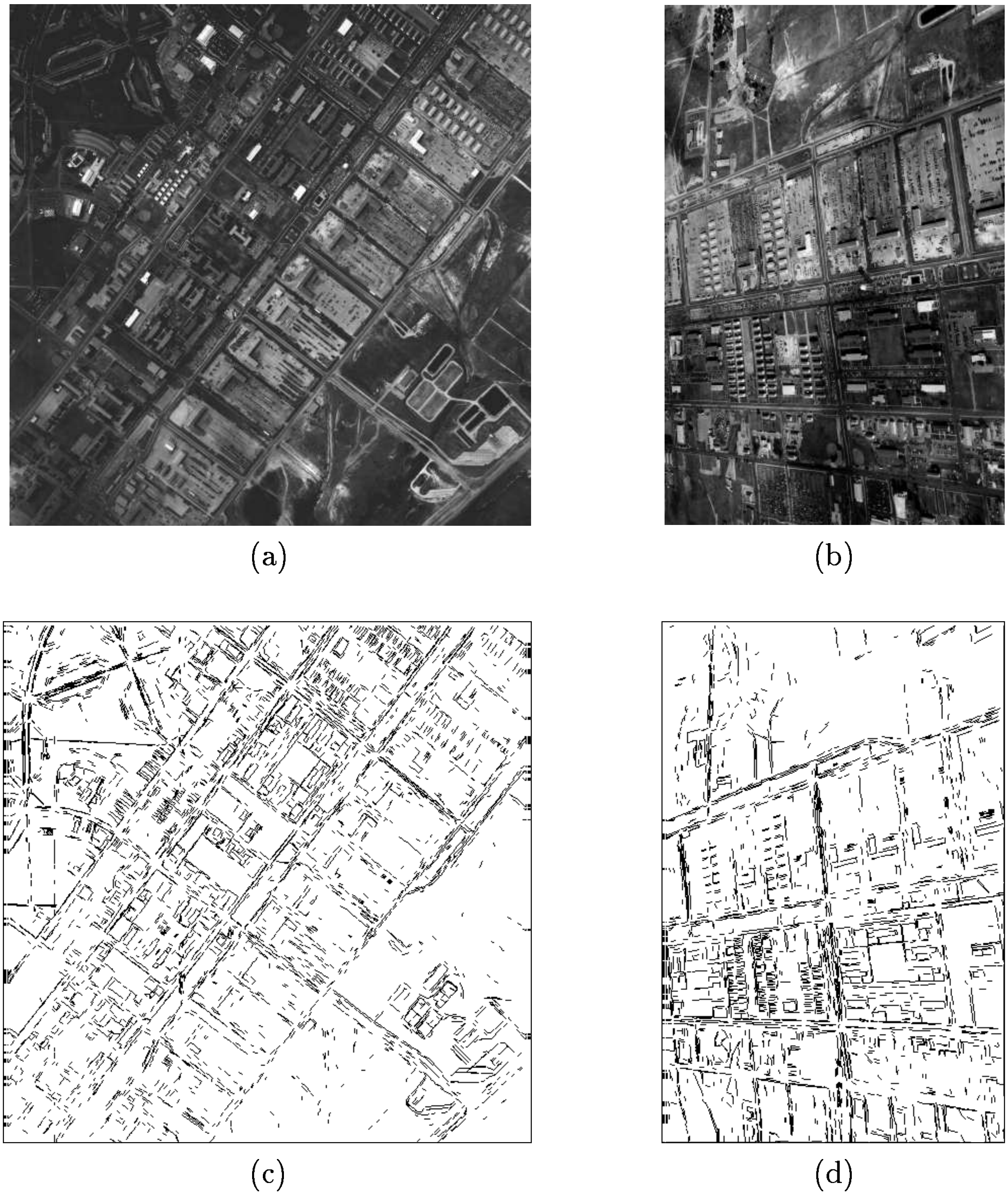

Figure 2: Data for estimating the projective transformation between two views of a scene. (a) A nadir view, suitable for site mapping. (b) an oblique view with strong perspective distortion. (c) Line segment features extracted for the nadir view. (d) Line segments for the oblique view.

Figure 3: A final registered mosaic for the two images in Figure 2. Line segments from the oblique view have been projectively unwarped into the coordinate system of the nadir view.

3.2 Distributions on the Sphere

The most important benefit to come from this isomorphism is that it allows probability distributions on the sphere to be reinterpreted as distributions in the projective plane. Since diametrically opposite points on the sphere must be treated as equivalent in order to represent the projective plane, an appropriate distribution must possess the property of antipodal symmetry, i.e. the probability density at any point on the sphere must be the same at the diametrically opposite point.

A useful characterization of distributions on the sphere is presented in Beran [Bera79]. Beran considers exponential distributions on the sphere, that is, distributions of the form $exp\{P\}$ where P is a polynomial evaluated over the surface of the sphere. This assumption is not as restrictive as it seems, since any strictly positive function F on the sphere can be represented as $exp\{ln\{F\}\}$. One good reason for considering exponential distributions is their ease of use in maximum likelihood estimation [Mend87].

Assuming a distribution of the form $exp\{P\}$, Beran notes that the polynomial P can be decomposed using spherical harmonics, analogous to the way polynomials in Euclidean space are decomposed using Fourier analysis. If the distribution is required to have antipodal symmetry, all odd order harmonics are identically zero. This leaves an expression $exp\{Y_0 + Y_2 + Y_4 + \ldots\}$. The zeroth harmonic is a constant, so the $exp\{Y_0\}$ term can be factored out and absorbed into the distribution's normalization constant. Therefore, the low order approximation to any antipodally symmetric exponential distribution on the sphere is of the form $exp\{Y_2\}$. A distribution having this form has already been studied in the statistical literature, where it is called Bingham's distribution [Bing74].

Bingham's distribution can be described as a trivariate Gaussian vector with zero mean and arbitrary covariance matrix, conditioned on the length of the vector being unity. Bingham's distribution thus represents the portion of a trivariate Gaussian distribution that intersects the surface of the unit sphere, with varying ellipsoidal shapes of the underlying Gaussian contours producing a variety of distributional forms on the sphere (see Figure 4). Bingham's distribution has been used previously in a computer vision setting to represent uncertainties in line and plane orientations estimated from vanishing point analysis and stereo line correspondences [Coll90b, Coll90a]. An application of Bingham's distribution to projective model extension is presented in [Coll92], where maximum likelihood estimation is used to fuse clusters of homogeneous coordinate estimates.

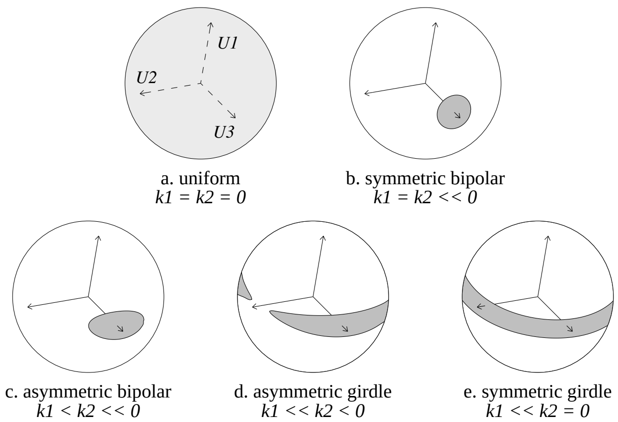

Figure 4: Bingham's Distribution – representative contours for varying shape parameter magnitudes.

4 Towards Nonplanar Reconstructions

Up to this point in the discussion, it has been assumed that the scene is either planar, or is viewed from far enough away that it appears planar. In this section we ease away from the planarity assumption by considering objects that are only approximately planar, that is, the deviation from planarity is enough to be noticeable, yet small enough that the perpendicular distance from any point to the plane is small compared with the perpendicular distance from the plane to the camera.

When deviations from planarity occur, the mapping from world to image is no longer completely described by a homography. We assume in this section that two views of a scene are available, that *some* coplanar set of scene features (at least four) are available, and that an image to image homography has been computed that maps the coplanar features from one image to their corresponding locations in the other. What can be said about the transformed images of scene features outside of the plane? Their positions as predicted by the homography will not in general coincide with their actual locations in the second image, due to their deviations from the plane. However, as is shown in this section, the difference between their predicted and actual positions is highly correlated with their 3D position with respect to the scene plane, and that in some cases, distances from scene features to the plane can be computed up to a single unknown scale factor.

As a concrete example, two aerial photographs are shown in Figures 5a and 5b. The camera viewpoint is close enough for noticeable departures from global scene planarity to appear. Figures 5c and 5d show 37 corresponding pairs of points that were chosen by hand. Several sets of points delimit the tops of buildings, and are therefore coplanar in the scene. The four pairs of coplanar points marked with a cross bound a rooftop, and were used to estimate a homography from the first image into the second. All points from the first image were then mapped into the second image using this homography, and their positions noted. Figure 6 shows difference vectors between predicted locations (in black) of points from image one, and actual point locations (in white) where they were found in image two. The four pairs of coplanar reference points line up exactly, as they must by definition of the homography. There is a remarkable structure to the remaining difference vectors. First, they all seem to be roughly parallel. We shall show that in fact all difference vectors must lie on infinite lines that intersect in a single point, which in this particular case is far off the image. Further examining Figure 6, we note that the difference vectors for structures taller than the rooftop used to compute the homography are oriented in one direction, while difference vectors for structures shorter than the rooftop are oriented in the opposite direction. This too is a property that holds in general, and allows us to qualitatively partition scene points into three categories depending on the orientation of their difference vectors: those lying closer to the viewer than the reference plane (positive vector sense), those lying on the plane (difference is zero), and those lying further away (negative vector sense). Finally, the length of a difference vector in Figure 6 seems to be highly correlated with the distance of the scene point from the rooftop reference plane. Figure 7 presents a graph showing signed lengths for each difference vector, overlaid with ground truth signed distances from the corresponding scene points to the reference plane. Both sets of numbers are scaled so that their maximum magnitude is 1. The correlation is striking. We show in this section that in the special case of nearly parallel difference vectors, the signed length of each vector is approximately proportional to the signed distance of the scene point from the object plane used to estimate the homography.

4.1 Difference Vector Pencils

When a planar object is viewed from two different positions, the image coordinates of an object point in frame 1 can be transformed into corresponding image coordinates in frame 2 by a homography. When the image coordinates of a scene point lying off of the object plane are mapped via the homography from frame 1 to frame 2, the coordinates predicted by the homography will differ from the actual corresponding point coordinates in frame 2. It is the purpose of this section to show that the difference vector between the predicted and actual image coordinates point either towards or away from a single *focus of expansion* (FOE) point in the image, depending on whether the scene point is closer to the viewer than the object plane, or farther away. If the difference vectors are extended out infinitely in both directions, they will form a *pencil* of lines, all intersecting at the FOE.

First consider the case of a calibrated camera pair. Given a scene point $\boldsymbol{P}$ in the coordinate system of camera 1, the position of $\boldsymbol{P}$ represented in the coordinate system of camera 2 is $\boldsymbol{RP}+\boldsymbol{T}$, where $\boldsymbol{R}$ is some 3×3 rotation matrix and $\boldsymbol{T}$ is a 3×1 translation vector. The image of point P in frame 2 is therefore the point where vector $\boldsymbol{RP}+\boldsymbol{T}$ intersects the image plane of camera 2. When the image of point $\boldsymbol{P}$ in frame 1 is transformed into frame 2 using a homography, it is assumed that

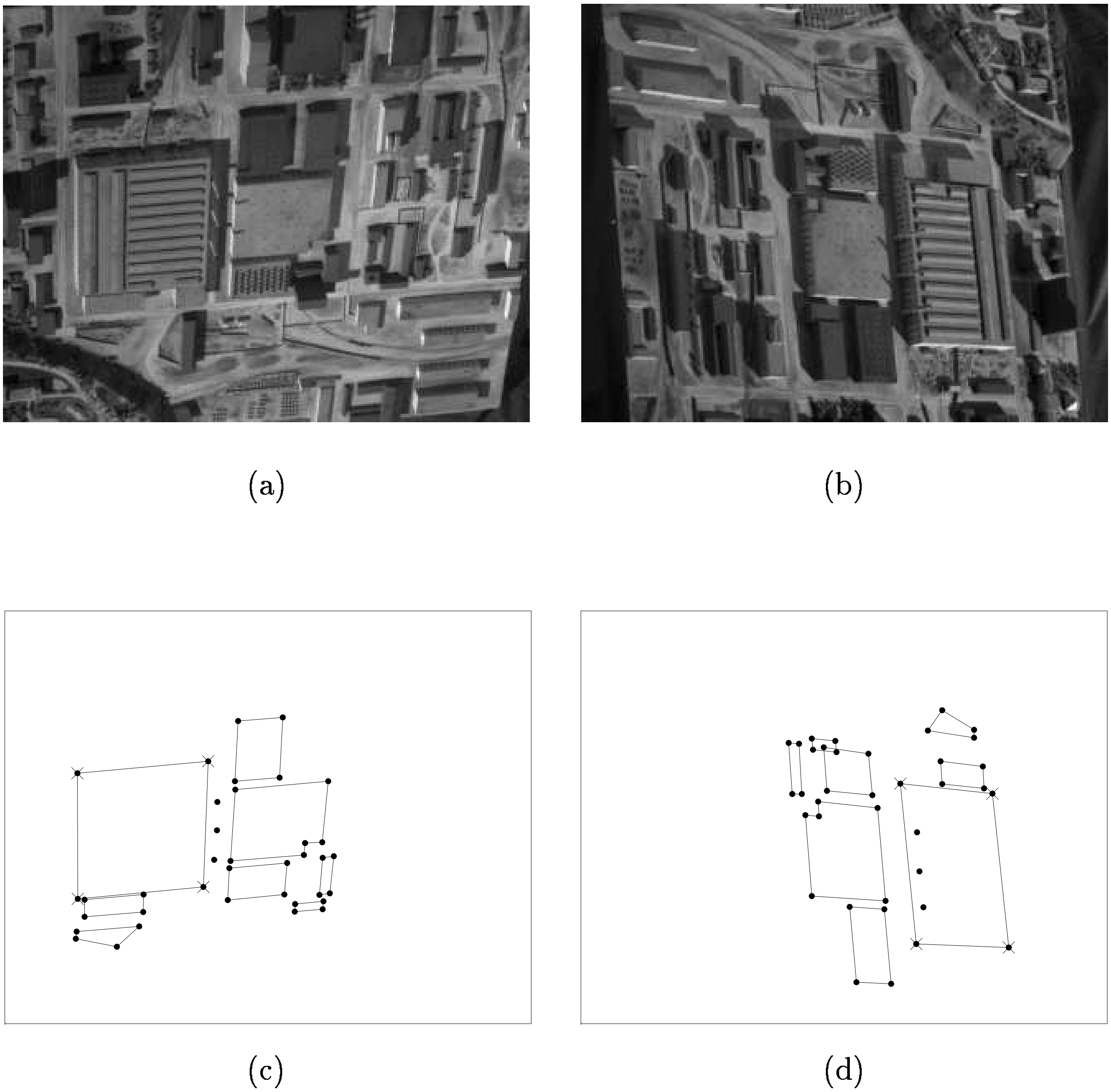

Figure 5: Two aerial photographs that are not well-approximated by a single plane. (a) Radius Model Board 1, Image J8. (b) Radius Model Board 1, Image J2. (c) Interesting points extracted by hand from Image J8. (d) Corresponding points extracted by hand from Image J2. Some building boundaries have been added for clarity. Crosses mark points that will be used to estimate a homography between the two images.

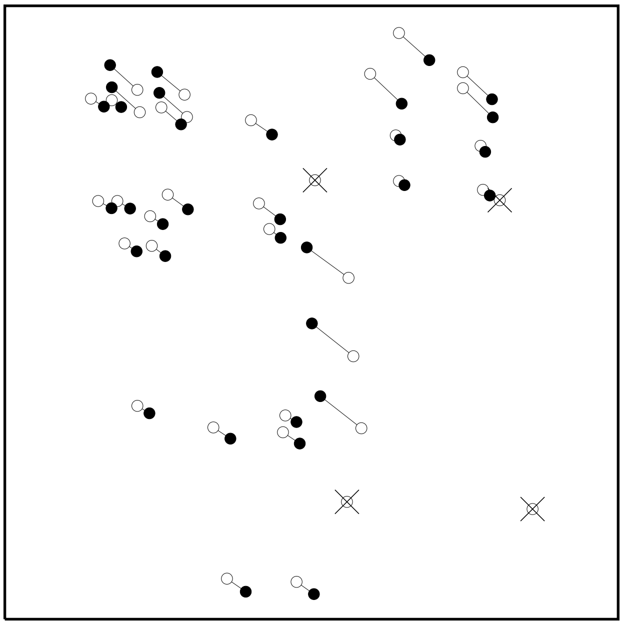

Figure 6: Difference vectors between points from image one transformed by a planar homography into image two (black dots), and their corresponding actual positions in image two (white dots). Points marked with crosses were used to define the homography.

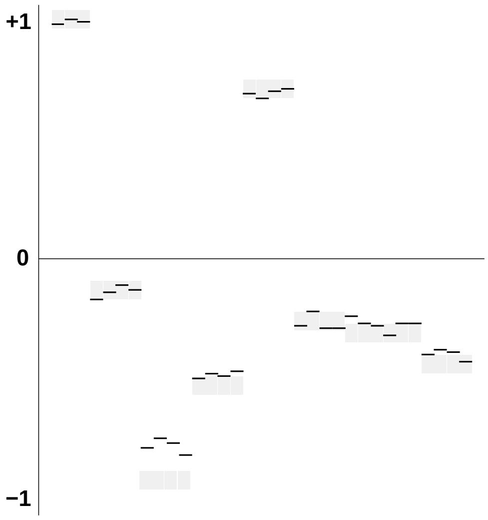

Figure 7: Graph of signed difference vector lengths (thin, black bars) overlaid with ground truth signed distances from the scene plane (wide, light grey bars). Both sets of value were scaled independently so that their maximum magnitude is 1. The poor vertical overlap between the lowest cluster of vector lengths and their ground truth distances is due to an incorrectly measured ground truth building height.

point $\boldsymbol{P}$ lies on the object plane the homography was estimated from. In other words, the homography maps scene point $\boldsymbol{P}$ not into its corresponding image location in scene frame 2 , but into the image location of the unique point on the object plane that lies on the same line of sight as point $\boldsymbol{P}$. This point can be represented in the coordinate system of frame 1 as $k\boldsymbol{P}$, where $k > 1$ if the scene point is closer to the viewer than the object plane, and $k < 1$ if the scene point lies farther away. The point $k\boldsymbol{P}$ can be represented in the coordinate system of camera 2 as $\boldsymbol{R}(k\boldsymbol{P}) + \boldsymbol{T}$, and therefore the homography predicts that scene point $\boldsymbol{P}$ will appear along vector $\boldsymbol{R}(k\boldsymbol{P}) + \boldsymbol{T}$, not along the vector it actually appears on, namely $\boldsymbol{RP} + \boldsymbol{T}$.

The difference vector in frame 2 between the actual location of scene point $\boldsymbol{P}$, and the position predicted for it by the homography must lie on the infinite line representing the intersection of image plane 2 with the plane through vectors $\boldsymbol{RP} + \boldsymbol{T}$ and $\boldsymbol{R}(k\boldsymbol{P}) + \boldsymbol{T}$. The normal vector of this plane is

$$(\boldsymbol{RP} + \boldsymbol{T}) \times (\boldsymbol{R}(k\boldsymbol{P}) + \boldsymbol{T}) \;=\; (1-k)(\boldsymbol{RP} \times \boldsymbol{T}). \quad (4)$$

This formula shows that the translation vector $\boldsymbol{T}$ also lies in the plane defined by predicted and actual image location vectors, thus the position in frame 2 where the direction of translation pierces the image, called the FOE, is colinear with the predicted and actual image locations of scene point $\boldsymbol{P}$ in frame 2. All difference vectors therefore lie on infinite lines passing through the FOE.

The normal vector in equation 4 changes sign depending on whether scale variable k is greater than or less than 1, which in turn depends on whether the scene point $\boldsymbol{P}$ is closer to the viewer than the object plane, or farther away. The distance of the scene point relative to the object plane therefore determines whether the difference vector between predicted and actual locations in image 2 is directed towards the FOE or away from it. If point $\boldsymbol{P}$ is closer to the viewer, the difference vector will point away from the FOE; if $\boldsymbol{P}$ is farther away, the difference vector is directed towards the FOE. It should be noted that these directions must be reversed if the translation vector $\boldsymbol{T}$ is oriented away from the scene, in which case what we have been calling the FOE is really a *focus of contraction* (FOC).

Recall from Section 2.1 that an uncalibrated camera produces an image that is some affine deformation of the pure pinhole image produced by a perfectly calibrated camera. To modify our discussion to consider uncalibrated cameras, we must therefore subject the difference vectors between predicted and actual locations to some unknown affine transformation. This transformation has no effect on the main results of this section, however. The pencil of difference vectors will be transformed into some other pencil of vectors, but they all still lie on infinite lines passing through a unique image point (the transformed FOE). Likewise, the sense of the vectors, whether they point away from or towards the transformed FOE, depends only on whether the scene points lie in front of or behind the object plane. The analysis of this section therefore allows image points to be partitioned into scene points lying in front of, on, or behind some object plane, even when the two images are produced by unknown, uncalibrated cameras.

4.2 Scene Distances from Parallel Differences

The difference between where a planar homography predicts a scene point will be, and where it is actually found in an image, forms a vector lying along a line through the (possibly affine warped) focus of expansion. This means that the difference between predicted and actual locations is an artifact of the translational motion between the two cameras only, and is not affected at all by their relative orientations. This simple fact explains why the length of a difference vector is so highly correlated with distance of the scene point from the object plane.

For calibrated cameras, the FOE (or FOC) found as the common intersection point of all the difference vectors determines the direction of translation between the cameras. Determining scene structure then reduces to analyzing what is in essence a pure translational flow field. If we knew the magnitude of the translation (the distance between the two cameras), we could theoretically deduce completely the structure of the scene from the lengths of the difference vectors. If the magnitude of translation is not known, the depths can be recovered up to a single unknown scale factor. Consideration of uncalibrated cameras complicates the matter, however. When a difference vector pencil is deformed by an affine transformation, the lengths of the vectors change unequally depending on their orientations. These unknown changes in length invalidate the computation of scene depths.

There is a special case, however, where an affine transformation scales all difference vectors by the same amount. This is the case when all difference vectors are parallel, corresponding to the FOE being far off the image (the direction of translation being nearly parallel to the image plane). In this case, the difference vectors are still scaled by an unknown amount, but they are all scaled by the same unknown amount, so that reconstructions of the scene up to a scale factor will still remain valid. The goal of this section is to show that the lengths of parallel difference vectors are proportional to the distances of their corresponding scene points from the object plane. We assume a calibrated camera system, since the result remains valid for uncalibrated cameras as well.

As before, let $\boldsymbol{P}$ be a scene point in the coordinate system of image 1, and let $k\boldsymbol{P}$ be the point on the object plane that projects to the same image location. These points are represented in the coordinate system of camera 2 as $\boldsymbol{RP}+\boldsymbol{T}$ and $\boldsymbol{R}(k\boldsymbol{P})+\boldsymbol{T}$ respectively. From now on, we concentrate on the coordinate system of camera 2. Let $\boldsymbol{Q}=(X,Y,Z)=\boldsymbol{RP}$ and $\boldsymbol{q}=(f/Z)\boldsymbol{Q}$ be the image of $\boldsymbol{Q}$, as shown in Figure 8. Since the

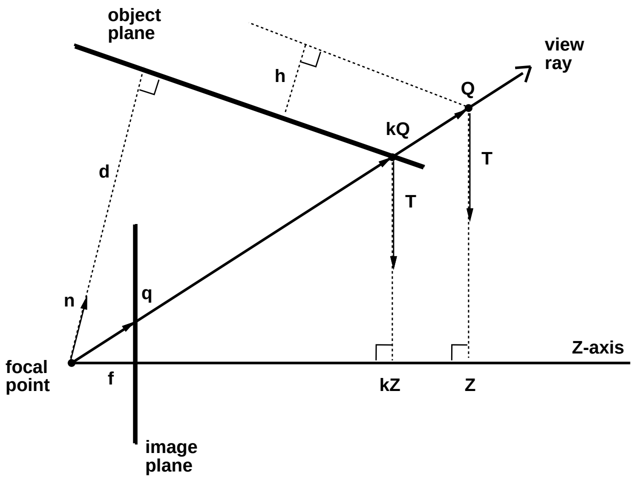

Figure 8: Sketch in the coordinate system for camera 2 of the derivation of difference vector length with respect to scene depth.

translation vector $\boldsymbol{T}$ is parallel to the image plane, the difference vector between predicted and actual point locations in frame 2 will be

$$\frac{f}{kZ}(k\boldsymbol{Q}+\boldsymbol{T})-\frac{f}{Z}(\boldsymbol{Q}+\boldsymbol{T}) \;=\; f(\frac{1}{kZ}-\frac{1}{Z})\boldsymbol{T}$$

by similar triangles, as shown in the figure. The signed length of the difference vector is thus

$$ft(\frac{1}{kZ}-\frac{1}{Z}) \tag{5}$$

where t is the unknown magnitude of the translation between camera 1 and camera 2.

We now wish to relate scene depths to distances from the object plane. Let the object plane be defined by normal vector $\boldsymbol{n}$ and distance d in the coordinate system of camera 2, and let the scene point under discussion be at a distance h from the object plane. Since the point $k\boldsymbol{Q}$ lies on the object plane, it satisfies the plane equation $\boldsymbol{n}^t(k\boldsymbol{Q}) = d$. Rewriting $\boldsymbol{Q} = (Z/f)\boldsymbol{q}$, we find that

$$kZ = \frac{d\,f}{\boldsymbol{n}^t\boldsymbol{q}}. \tag{6}$$

Likewise, point $\boldsymbol{Q}$ satisfies the plane equation $\boldsymbol{n}^t\boldsymbol{Q} = d + h$ so that

$$Z = \frac{(d+h)\,f}{\boldsymbol{n}^t\boldsymbol{q}}. \tag{7}$$

Substituting equations 6 and 7 into 5 yields

$$ft\left(\frac{\boldsymbol{n}^t\boldsymbol{q}}{df} - \frac{\boldsymbol{n}^t\boldsymbol{q}}{(d+h)f}\right) = \frac{t(\boldsymbol{n}^t\boldsymbol{q})}{d(d+h)}h \tag{8}$$

the desired equation for difference vector length in terms of distance h of a scene point from the object.

It is apparent from equation 8 that even in the case of parallel difference vectors, the length of a vector is not exactly proportional to the distance of the scene point from the object plane. However, for approximately planar structures where the perpendicular distance h of a scene point to the plane is small compared with the perpendicular distance d of the object plane from the camera, we are justified in replacing $(d + h)$ by d. The only other factor in equation 8 that is not a constant scale factor is $\boldsymbol{n}^t\boldsymbol{q}$, which varies according to the vector $\boldsymbol{q}$ representing points in the image plane. Locally, these vectors do not vary much, and $\boldsymbol{n}^t\boldsymbol{q}$ will remain nearly constant. Other situations in which $\boldsymbol{n}^t\boldsymbol{q}$ is approximately constant include cameras with a small field of view, and viewing angles where the line of sight is roughly perpendicular to the object plane.

5 Conclusion

Many coordinate system mappings relevant to the visual reconstruction of piecewise planar scenes can be represented by linear, invertible projective transformations called homographies. Once this is recognized, powerful methods from the field of projective geometry become available. The fundamental theorem of projective geometry shows how projective transformations can be estimated from a small number of point correspondences. When a transformation is estimated between two images of the same planar object, points in one image can be readily mapped to their corresponding locations in the other image, to aid the search for further correspondences. A homography estimated between an object plane and its projected image allows extension of the object model with new points and lines from the image, without first solving for camera motion, pose, or intrinsic calibration parameters.

In order to apply the theorems of projective geometry in practical settings, deviations from mathematical perfection must be considered. In this paper we have examined two sources of discrepancy between the idealized projective transformation model and its real-world applications. The first source is random observation error in positions of the image features due to sensor noise, which leads to errors in the planar reconstruction. It was shown that the projective plane has a different global topology than the Euclidean plane, and therefore standard data fusion techniques are not necessarily appropriate. Using homogeneous coordinates, a topologically correct model of the projective plane was developed and shown to be isormorphic to the surface of a unit sphere having antipodal points equated. Consideration of antipodally symmetric probability distributions on the sphere identified Bingham's distribution as a first-order approximation to any probability distribution in the projective plane. In [Coll92], data fusion in the projective plane is performed using maximum likelihood estimation on samples from Bingham density clusters.

The second deviation from a pure projective geometry setting that was considered in this paper was an analysis of how noncoplanar points are transformed under a homography. When points outside of an object plane are transformed by an image to image homography, there is a systematic difference between their image locations predicted by the homography and their actual perceived locations. All difference vectors lie along infinite lines that converge at an FOE. Furthermore, the direction of a vector either towards or away from the FOE determines whether the scene point is in front of the object plane, or behind it, as seen by the camera. Finally, lengths of the difference vectors are highly correlated with distances of the corresponding scene points from the object plane. This phenomena was explained for the case where all difference vectors are parallel in the image. Future work will try to extend these results to achieve greater accuracy and generality.

References

[Bera79] R. Beran, "Exponential Models for Directional Data," *The Annals of Statistics,* Vol. 7, No. 6, 1979, pp. 1162–1178.

[Bing74] C. Bingham, "An Antipodally Symmetric Distribution on the Sphere," *The Annals of Statistics,* Vol. 2, 1974, pp. 1201–1225.

[Coll90a] R.T. Collins and R.S. Weiss, "Vanishing Point Calculation as a Statistical Inference on the Unit Sphere," *Proc. Third International Conference on Computer Vision,* Osaka, Japan, December 1990, pp. 400–403.

[Coll90b] R.T. Collins and R.S. Weiss, "Deriving Line and Surface Orientation by Statistical Methods," *Proc. Darpa I.U. Workshop,* Pittsburgh, PA., Sept. 1990, pp. 433–438.

[Coll92] R.T. Collins, "Single Plane Model Extension using Projective Transformations," *Proc. Darpa I.U. Workshop,* San Diego, CA, Jan. 1992, pp. 917–923.

[Coll93] R.T. Collins and J.R. Beveridge, "Matching Perspective Views of Coplanar Structures using Projective Unwarping and Similarity Matching," *Proc. IEEE Computer Vision and Pattern Recognition,* New York City, June 1993, pp. 240–245. Also appeared in *Proc. Darpa I.U. Workshop,* Washington, DC, April 1993, pp. 459–463.

[Faug88] O.D. Faugeras and F. Lustman, "Motion and Structure from Motion in a Piecewise Planar Environment," *International Journal of Pattern Recognition and Artificial Intelligence,* Vol. 2, 1988, pp. 485–508.

[Gros90] W.I. Grosky and L.A. Tamburino, "A Unified Approach to the Linear Camera Calibration Problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 12, 1990, pp. 663–671.

[Horn86] B.K.P. Horn, *Robot Vision,* MIT Press, Cambridge, MA., 1986.

[Mend87] J.M. Mendel, *Lessons in Digital Estimation Theory,* Prentice-Hall Signal Processing Series, Prentice-Hall, Inc., NJ. 1987.

[Mohr90] R. Mohr and E. Arbogast, "It Can be Done Without Camera Calibration," *Pattern Recog. Letters,* V. 12, 1990, pp. 39–43.

[Mohr91] R. Mohr and L. Morin, "Relative Positioning from Geometric Invariants," *Computer Vision and Pattern Recognition,* Maui, Hawaii, June 1991, pp. 139–144.

[Spri64] C.E. Springer, *Geometry and Analysis of Projective Spaces,* W.H. Freeman and Company, San Francisco, 1964.

[Tsai82] R.Y. Tsai, T.S. Huang, and W. Zhu, "Estimating Three-Dimensional Motion Parameters of a Rigid Planar Patch, II: Singular Value Decomposition," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* Vol. 30, No. 4, 1982, pp. 525–534.