MODEL ACQUISITION USING STOCHASTIC PROJECTIVE GEOMETRY

A Dissertation Presented

by

ROBERT T. COLLINS

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 1993

Department of Computer Science

MODEL ACQUISITION USING STOCHASTIC PROJECTIVE GEOMETRY

A Dissertation Presented

by

Robert T. Collins

Approved as to style and content by:

Edward M. Riseman, Chair of Committee

Allen R. Hanson, Member

Edwina L. Rissland, Member

Donald Geman, Member

Richard S. Weiss, Member

W. Richards Adrion, Department Chair
Computer and Information Science

In memory of my grandfather,

Donald E. Markert

March 10, 1908
July 24, 1993

ACKNOWLEDGEMENTS

First, I would like to thank my immediate family for their support and understanding during the final stages of this dissertation. My wife, Yanxi, always seems to have an infinite supply of energy, and has been a great inspiration to me. I thank her for the dinners out, teaching me tennis, and skating, trying to teach me dancing, showing me how to make dumplings, and for giving me a reason to travel all over the world. I would also like to thank my mother-in-law, Wei-Hua, for cooking us gourmet Chinese meals every night, and Charlie, the cat, for being a little “tao-yen”.

I thank my parents for providing a supportive environment while I was growing up, and for watching with an uncritical eye as I educated myself out of the job market. If home is the place that, when you go there, they have to let you in, then my parent’s house truly is “home”. I also want to thank my grandmother for keeping things in perspective by asking me, now and then, whether I had finished my “paper” yet. And of course, my brother, for providing an existence proof that there is a life out there, if only one would get out of school.

The VISIONS group has really been a wonderful place to be, these past ten years. This is primarily due to the guidance of Ed Riseman and Al Hanson, who shelter us all so well from the day to day chaos of supporting a research lab. They have managed to bring together a special group of people. I would like to thank Bob Heller for knowing everything about computers, and Val Conti for having just the coolest party house imaginable. And how often have I gone running to LaurieW and Janet Turnbull with a problem, and found out just who *really* runs this place?

To all my friends and colleagues, I wish you warm fuzzy feelings and a great deal of success.

ABSTRACT

MODEL ACQUISITION USING STOCHASTIC PROJECTIVE GEOMETRY

SEPTEMBER 1993

ROBERT T. COLLINS

B.S., UNIVERSITY OF HARTFORD

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

PH.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Edward M. Riseman

This thesis presents a methodology for scene reconstruction that is based on the principles of projective geometry, while dealing with uncertainty at a fundamental level. Uncertainty in geometric features is represented and manipulated using probability density functions on projective space, allowing geometric constructions to be carried out via statistical inference.

The main contribution of this thesis is the development of *stochastic projective geometry*, a formalism for performing uncertain geometric reasoning during the scene reconstruction process. The homogeneous coordinates of points and lines in the projective plane are represented by antipodal pairs of points on the unit sphere, and geometric uncertainty in their location is represented using Bingham's distribution. Geometric reasoning about homogeneous coordinate vectors reduces to well-defined manipulations on probability density functions. For example, a Bayesian approach to evidence combination on the sphere is presented for fusing noisy homogeneous coordinate observations constrained by known projective incidence relations. The result is an uncertainty calculus in projective space analogous to the Gaussian uncertainty calculus in affine space. The main strength of the Gaussian calculus is maintained, namely its uniform treatment of uncertainty in all stages of the geometric

reasoning process. At the same time, the limitations of the Gaussian density function as a representation of uncertainty in projective space are removed.

The effectiveness of stochastic projective geometry for dealing with noisy projective relationships is demonstrated on three geometric vision problems: deriving line and plane orientations using vanishing point analysis, partitioning scene features into planar patches using line correspondence stereo, and extending a partial model of planar surface structure using projective invariants.

TABLE OF CONTENTS

LIST OF TABLES

Table | Page

LIST OF FIGURES

CHAPTER 1

Introduction

One of the main goals of computer vision is to infer three-dimensional scene structure from one or more two-dimensional images. Object boundaries and surface markings appear as intensity discontinuities in the image that can be reliably detected under a wide range of viewing conditions. Algorithms for extracting these discontinuities produce a sparse representation of the image composed of purely geometric features like vertices, line segments, and curves. The task of inferring scene structure can therefore be viewed abstractly as a process of geometric construction, where three-dimensional scene geometry is inferred from a set of two-dimensional geometric primitives and a description of the imaging process.

Geometric modeling of complex environments can be difficult. 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. Add to this the problem of efficiently recovering such representations from 2D images and the task seems computationally intractable. One simplification is to focus on planar patch models. Applications exist where significant portions of the scene are indeed planar, such as indoor mobile robotics, city or campus navigation, and aerial photo-interpretation of cultural sites.

Even in unrestricted environments, the world is almost always flat enough locally to approximate by piecewise planar patches. The question is not *whether* we can approximate the world this way, but *how well.*

Adopting the world-planarity assumption is attractive because the relevant geometric entities, namely points, lines and planes, are linear subspaces and a rich set of mathematical tools from the field of projective geometry can be applied. The relevance of projective geometry to the visual interpretation of man-made scenes cannot be overstressed: Projective geometry provides a mathematical foundation for characterizing and representing the relationships between linear subspaces that remain invariant under the imaging process.

Part of the relevance of projective geometry for computer vision is that the process of picture formation can be accurately modeled as a perspective projection from the 3D world to a 2D image. Colinear line segments in the world appear as colinear segments in the image plane because colinearity is a property that is invariant under perspective projection. When a picture is taken of a planar surface the perspective mapping is 2D to 2D, and the projection from local surface coordinates to image coordinates is invertible. The fundamental theorem of projective geometry shows how to estimate a 2D projective transformation given 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 new 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. A more robust form of projective geometry is needed, one that allows for possible imprecision in its geometric primitives. In this thesis, uncertainty in geometric object models is represented and manipulated using probability density functions on projective space, allowing valid geometric constructions to be carried out via statistical inference. The result is a methodology for scene reconstruction based on the principles of projective geometry, yet also dealing with uncertainty at a fundamental level.

It will be shown that projective n-space can be visualized using the surface of a unit sphere in $n+1$-dimensional Euclidean space. Each point in projective space is represented as a pair of opposing or *antipodal* points on the sphere. By this map from projective space to the unit sphere, antipodally symmetric probability distributions on the sphere may be interpreted as probability distributions over the points of projective space. Standard constructions of projective geometry can then be augmented by statistical inferences on the sphere, resulting in a Bayesian inference engine for performing geometric scene reconstruction. The effectiveness of this framework is demonstrated on several geometric problems, including the derivation of 3D line and plane orientations from a single image using vanishing point analysis, the extraction of a planar patch scene model using stereo line correspondences, and the reconstruction of planar surface structure using multiple images taken from unknown viewpoints by uncalibrated cameras.

The remainder of this chapter develops these topics in more detail. Some familiarity with the fundamental concepts of real projective geometry is assumed. For convenience, basic background material is presented in Appendix A. The connection between projective geometry and computer vision is made in Section 1.1 through the pinhole camera model of image formation. In Section 1.2 the analytic formulation of projective geometry via homogeneous coordinates is used to provide a Euclidean model of projective space, opening the door to the introduction of metrics and probability theory. This chapter closes with an overview of the rest of the thesis.

1.1 Camera Geometry

The importance of projective geometry for describing the image formation process is due to the pinhole camera model. Figure 1.1 shows a standard, left-handed pinhole camera coordinate system. The origin is located at the camera focal point. The positive Z axis is aligned with the focal axis, which is perpendicular to the image plane that is located some unit distance away. Points in the image are measured in a local, U-V coordinate system whose origin is at the principle point (intersection of the focal axis with the image plane), and with image axes U and V parallel to the camera coordinate system X and Y axes.

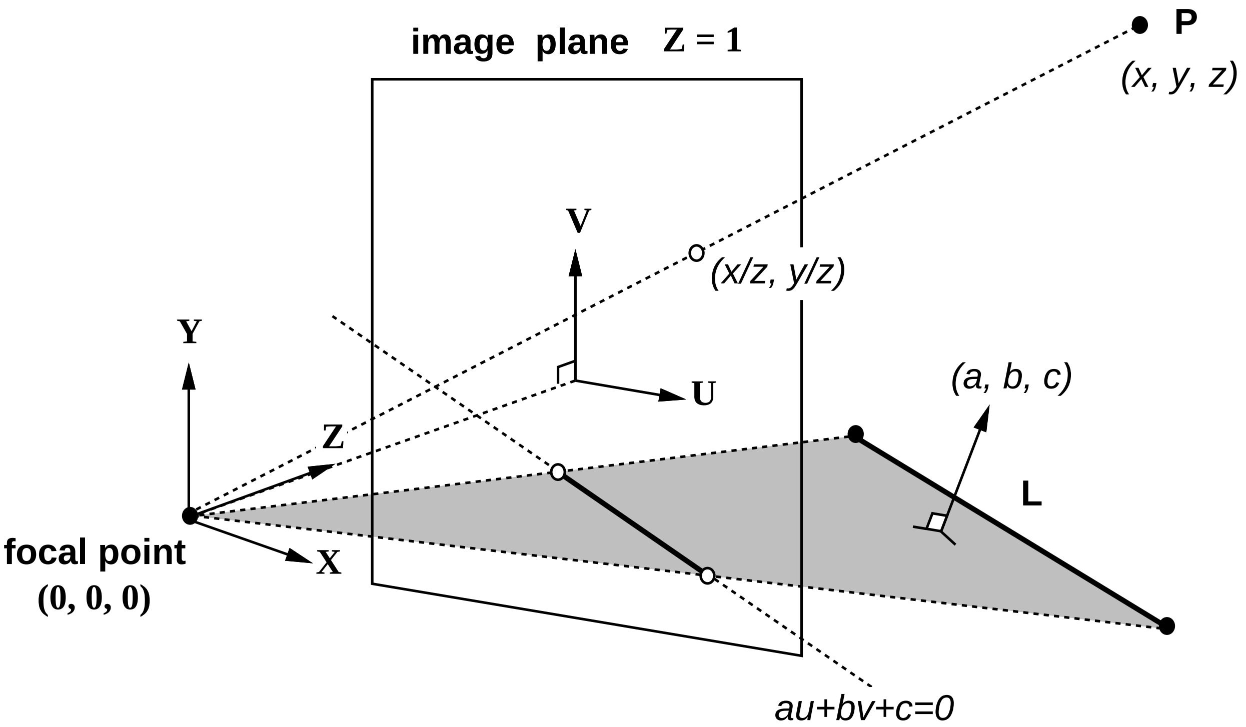


Figure 1.1: The pinhole camera model of image formation. The mapping from scene coordinates into image coordinates is a 3D-to-2D perspective projection.

The pinhole image of a scene point $P = (x, y, z)$ is the point $(u, v) = (x/z, y/z)$ where the *projection line* passing through both the scene point and the focal point intersects the image plane $Z = 1$. Likewise, a line segment in the scene projects to a line segment in the image. All projection lines for points on line segment L lie in a *projection plane* $aX + bY + cZ = 0$, whose normal vector (a, b, c) is the cross product of any two distinct points on line L. Intersecting this projection plane with the image plane $Z = 1$ forms the infinite image line $aU + bV + c = 0$, along which the projected line segment must lie.

Since the pinhole camera model is clearly a perspective projection from the 3D scene onto the 2D image plane $Z = 1$, geometric properties that are invariant under perspective projection are of fundamental importance to a vision system. For instance, barring pathological views, colinear 3D lines in the world will always project to colinear 2D line segments in an image, whereas the projection of parallel 3D lines onto an image will not always produce parallel 2D lines. This is because colinearity is invariant under nonsingular perspective projections, while parallelism is not.

1.1.1 Perspective Projection

It was well-known even in the early days of computer vision that the perspective projection that maps 3D object points to 2D image points can be represented as a linear transformation in homogeneous coordinates [Roberts65]. The derivation involves applying a rigid rotation and translation to transform a point in local object coordinates L-M-N (refer to the coordinate system labels in Figure 1.2) into the camera coordinate system X-Y-Z, then applying pinhole projection to yield a point in the 2D image coordinate system U-V. This process is written in matrix form as

$$k\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} l \\ m \\ n \\ 1 \end{bmatrix} \tag{1.1}$$

$$= \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}\begin{bmatrix} l \\ m \\ n \\ 1 \end{bmatrix} \tag{1.2}$$

with k an arbitrary homogeneous coordinate scale factor, $R = \{r_{i,j}\}$ an orthonormal rotation matrix, and $T = \{t_i\}$ a translation vector. The result is a 3×4 *projection matrix* of parameters p_{ij} that are functions of the location and orientation of camera with respect to the external scene – i.e. the *extrinsic* camera parameters.

For a subset of object points lying on a single plane, the projection matrix that maps object points to image points simplifies to a 3×3 matrix. One planar face of the object in Figure 1.2 has a 2D coordinate system P-Q imprinted on it. Let that local plane coordinate system be positioned with respect to the full 3D object coordinate system L-M-N by known rotation and translation parameters r'_{ij} and t'_i. That is

$$\begin{bmatrix} l \\ m \\ n \\ 1 \end{bmatrix} = \begin{bmatrix} r'_{11} & r'_{12} & r'_{13} & t'_1 \\ r'_{21} & r'_{22} & r'_{23} & t'_2 \\ r'_{31} & r'_{32} & r'_{33} & t'_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} p \\ q \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} r'_{11} & r'_{12} & t'_1 \\ r'_{21} & r'_{22} & t'_2 \\ r'_{31} & r'_{32} & t'_3 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} p \\ q \\ 1 \end{bmatrix}. \tag{1.3}$$

By combining equations (1.2) and (1.3) the transformation from local 2D object plane

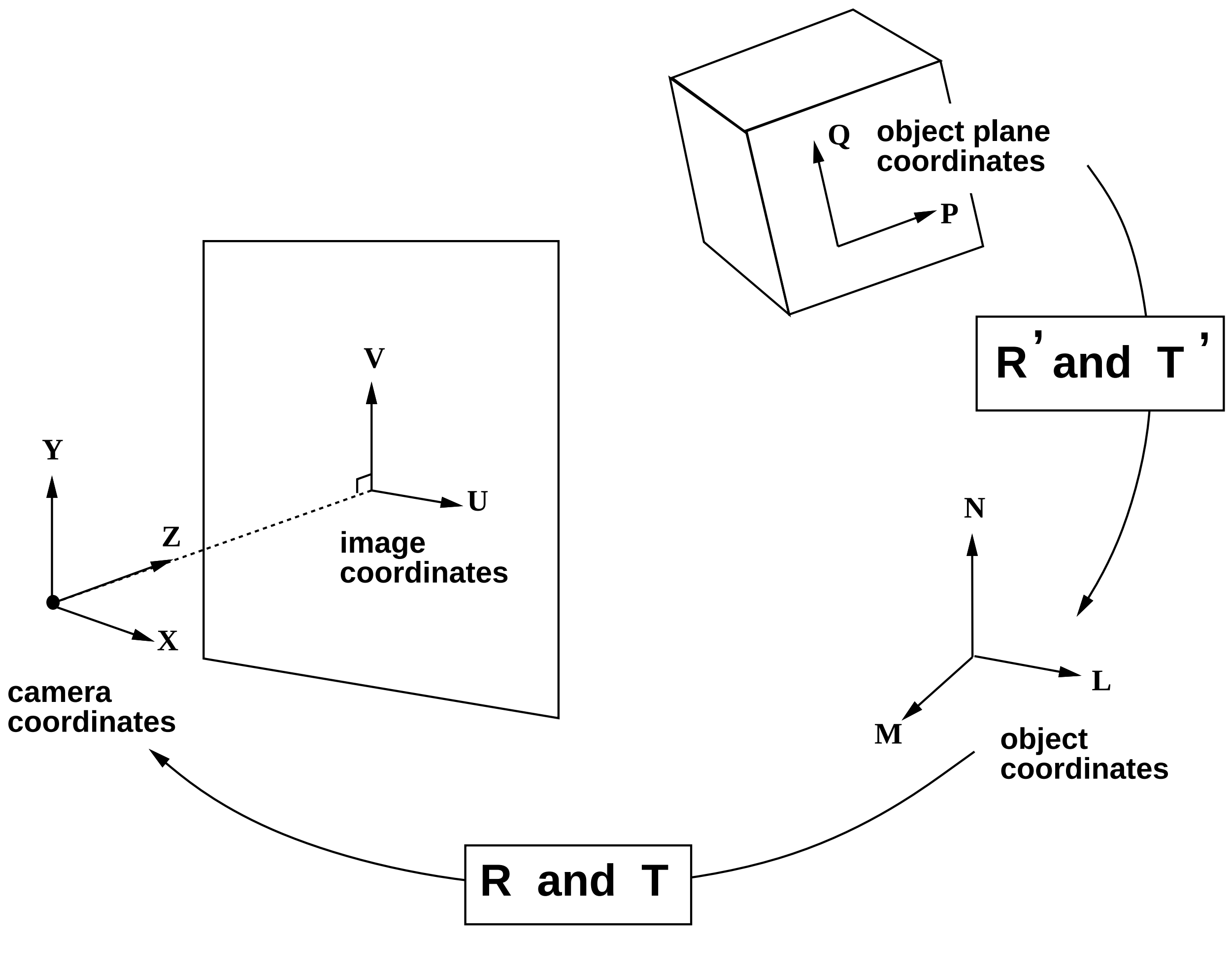


Figure 1.2: Coordinate frames for pinhole projection.

coordinates P-Q to image coordinates U-V is found to be

$$k \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} r'_{11} & r'_{12} & t'_1 \\ r'_{21} & r'_{22} & t'_2 \\ r'_{31} & r'_{32} & t'_3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p \\ q \\ 1 \end{bmatrix} \tag{1.4}$$

$$= \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} p \\ q \\ 1 \end{bmatrix} . \tag{1.5}$$

For this mapping to be nonsingular, and therefore invertible, it suffices that the object plane P-Q, when infinitely extended, does not pass through the origin of the camera coordinate system. The coordinates of object plane points can then be derived via backprojection from the positions of points in the image.

1.1.2 Planar Homographies

It is evident from equation (1.5) that the pinhole camera projection from an object plane to the image plane is a *homography,* that is, an invertible linear transformation in homogeneous coordinates.[1] Once this observation is made, powerful mathematical tools can be used to analyze images containing planar patch objects. For example, the set of all planar homographies forms a group. This means that every homography is invertible and closed under composition (the mapping of object points to image points in a picture of a picture of a planar object is still a homography). Because they are linear, invertible and closed under composition, homographies greatly simplify the analysis of piecewise planar environments. Figure 1.3 shows a familiar computer vision scenario where a single planar surface patch is imaged by two cameras from two different viewpoints. If the homographies relating object

[1]Pinhole projection is actually a special type of homography called a *perspectivity* – Appendix A.1 contains a review of these terms.

plane coordinates to pinhole image coordinates for the first and second cameras are labeled $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ respectively, then the homography relating points from the second pinhole image to their corresponding positions in the first is simply $\boldsymbol{H_1}\boldsymbol{H_2}^{-1}$.

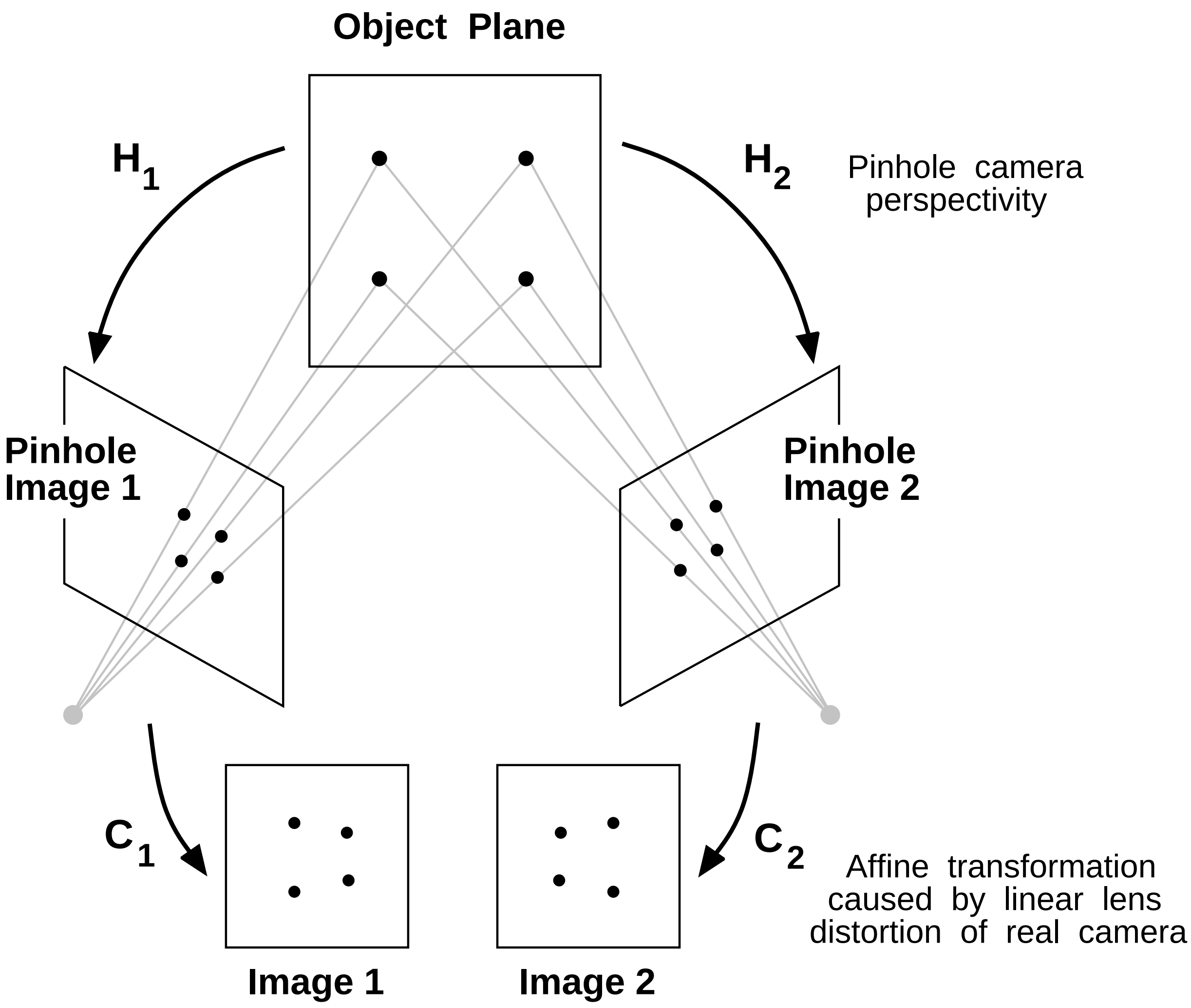


Figure 1.3: Some planar homographies in computer vision. Here, an object plane is viewed by two cameras whose linear camera calibration parameters distort the pure pinhole perspective image by an affine transformation. Corresponding points in any two planes in this diagram are related by a homography.

In actuality, the pure pinhole camera model does not adequately characterize images produced by real cameras. Disregarding nonlinear lens aberrations for the moment, the deviation of a particular camera from the pinhole model is governed by a set of linear internal calibration parameters. The internal or *intrinsic* camera parameters include the effective focal length, image aspect ratio, position of the principle point, and image axis skew. The effects of these parameters combine to produce an observed image that is some 2D affine deformation of the pure pinhole image [Horn86]. The affine deformations due to camera calibration parameters are labeled $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ in Figure 1.3. Affine transformations are also homographies, therefore the composition of a pinhole projection $\boldsymbol{H}$ followed by an affine deformation $\boldsymbol{C}$ yields a new homography $\boldsymbol{CH}$. Geometric constructions that are invariant under homographies will therefore remain valid regardless of whether or not the camera is calibrated correctly. This result is of great practical importance since accurate camera calibration is a demanding task, and in some situations (such as interpreting photos taken from a book or magazine) estimates of the camera parameters are not available at all.

Since all the plane to plane transformations in Figure 1.3 are homographies, it is easy to derive the transformation between any two planes in the diagram; for instance the transformation mapping points in image 2 into corresponding points in image 1 is $\boldsymbol{C}_1\boldsymbol{H}_1\boldsymbol{H}_2^{-1}\boldsymbol{C}_2^{-1}$. The homography $\boldsymbol{H}_1\boldsymbol{H}_2^{-1}$ relating pinhole images of coplanar points has appeared in the computer vision literature, and it has been shown that a decomposition of the homography matrix allows recovery of the relative positions of the two cameras with respect to each other and to the object plane [Tsai82, Faugeras88]. More recent work 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 calibration parameters [Mohr91, Collins92].

Unfortunately, some camera images show evidence of nonlinear lens distortion, and any analysis based on projective transformations may be invalid. Under radial or "barrel" distortion, for example, the images of colinear points may no longer be colinear [ASP80], and the mapping from object to image plane can therefore no longer be described by a homography. In practice it is common to ignore nonlinear camera parameters, or to assume that a preprocessing step has been performed to remove their effects [Grosky90]. In this thesis, the term "calibrated camera" is used when enough information is known about the intrinsic camera parameters, both linear and nonlinear, to recover the pure pinhole camera image. The results cited for calibrated cameras are therefore general. In contrast, results presented for uncalibrated cameras hold only when the unknown intrinsic camera parameters are linear – nonlinear lens effects must either be negligible, or known well enough to be corrected.

1.2 Stochastic Projective Geometry

Geometric measurements made by real camera systems are not exact. Indeed, geometric primitives extracted from real images suffer from a variety of degradations with causes ranging from random sensor noise to the idiosyncracies of individual feature extraction algorithms. For geometric constructions based on these inaccurate primitives to remain valid, a mechanism must be in place that can represent and manipulate uncertain geometric objects, allowing redundant measurements to be combined to achieve more robust results.

Stochastic Geometry provides a general approach to representing, manipulating, comparing and combining uncertain geometric features [Durrant-Whyte88]. In this formalism, uncertain geometric features are represented by probability distributions over a parameterized space of geometric objects. Each point in the parameter space represents a possible instantiation of one type of geometric object; different geometric classes are described by different parameter spaces. A probability density function over points in a parameter space represents the likelihood of observing a particular instance of that object class. The manipulation of uncertain geometric structures then reduces to well-defined problems involving the transformation and combination of probability density functions.

The appropriate parameter spaces for representing projective geometric objects relevant to computer vision are the projective line, the projective plane and projective 3-space. These projective spaces have a different global structure than the corresponding Euclidean spaces, therefore probability distributions appropriate for representing geometric features in Euclidean space will not necessarily be valid in projective space. Section 1.2.1 first discusses the topological difference between projective and Euclidean space. Then, in Section 1.2.2 homogeneous coordinates are used to develop a model in Euclidean space that *is* topologically correct. Projective n-space will be modelled by the surface of the Euclidean unit sphere with antipodal points equated. Suitably defined probability distributions on the sphere may then be used as distributions on projective space. This is the approach taken in Chapter 3.

1.2.1 Topology of the Projective Plane

Modeling the image formation process as a projective transformation greatly simplifies the analysis of piecewise planar scenes. This simplification comes at a price, however. In order to make a projective mapping bijective (one-to-one and onto), points at infinity must be added to the domain and range. Both image and object plane are now *projective* planes.

The addition of points at infinity gives the projective image plane quite different properties than the familiar Euclidean image plane. The most important difference is that the projective plane is topologically compact while the Euclidean plane is not. This means that a single cluster of points centered around a point at infinity in the projective plane appears as two clusters infinitely far apart in the Euclidean plane. Any estimation technique based on "averaging" these points in the Euclidean plane will produce bad results. Proper handling of point clusters at infinity is not just of theoretical interest; such clusters **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 (see Chapter 4).

1.2.2 A Euclidean Model of the Projective Plane

Projective spaces of n-dimensions are not adequately modeled by their n-dimensional Euclidean counterparts because of the need to explicitly represent points at infinity. In this section, homogeneous coordinates are used to build models in Euclidean $(n+1)$-space that can be used to visualize projective n-space.

Recall that homogeneous coordinates represent a point in projective n-space by an $(n+1)$-dimensional vector of numbers, not all zero, and that any two homogeneous coordinate vectors are equivalent if one is a nonzero scalar multiple of the other. For concreteness, consider the projective plane, where 2D points and lines are represented by homogeneous coordinate vectors in $(R^3 - \{0\})$. Let $\sim$ be an equivalence relation over points in $(R^3 - \{0\})$ such that

$$\boldsymbol{x} \sim \boldsymbol{y} \quad \text{iff} \quad (x_1, x_2, x_3) = (ky_1, ky_2, ky_3), k \neq 0. \tag{1.6}$$

The projective plane P^2 is defined as the quotient space $(R^3 - \{0\})/\sim$. Viewing R^3 as Euclidean space, each equivalence class of homogeneous coordinates can be thought of as an infinite line through the origin, excluding the origin itself which can not correspond to any homogeneous coordinate.

Figure 1.4 develops the relationship between homogeneous coordinates and perspective camera geometry The homogeneous coordinates for a point $p = (u, v)$ in the image plane are $k(u, v, 1)$. To view this geometrically, place the image plane at $Z = 1$. The set of equivalent homogeneous coordinates for point p is now a line h_p passing through the focal point and point p in the image plane. The homogeneous coordinates of a line segment l lying along line $aU + bV + c = 0$ in the image are $k(a, b, c)$. This equivalence class of homogeneous coordinates is visualized as line h_l passing through the focal point perpendicular to the plane containing the focal point and line segment l. Points at infinity are easily represented using homogeneous coordinates. A point at infinity in the projective image plane is represented as $k(u, v, 0)$ in homogeneous coordinates, and is visualized as a line passing through the origin parallel to the image plane. The homogeneous coordinates of the line at infinity in the projective plane are $k(0, 0, 1)$. This is viewed as a line passing through the origin perpendicular to the image plane. Since every line through the origin of R^3 corresponds to a unique point (and line) in the projective plane, and vice versa, the two sets of elements are isomorphic. In fact it can be shown that the two sets are homeomorphic – they have the same topological structure [Boothby86].

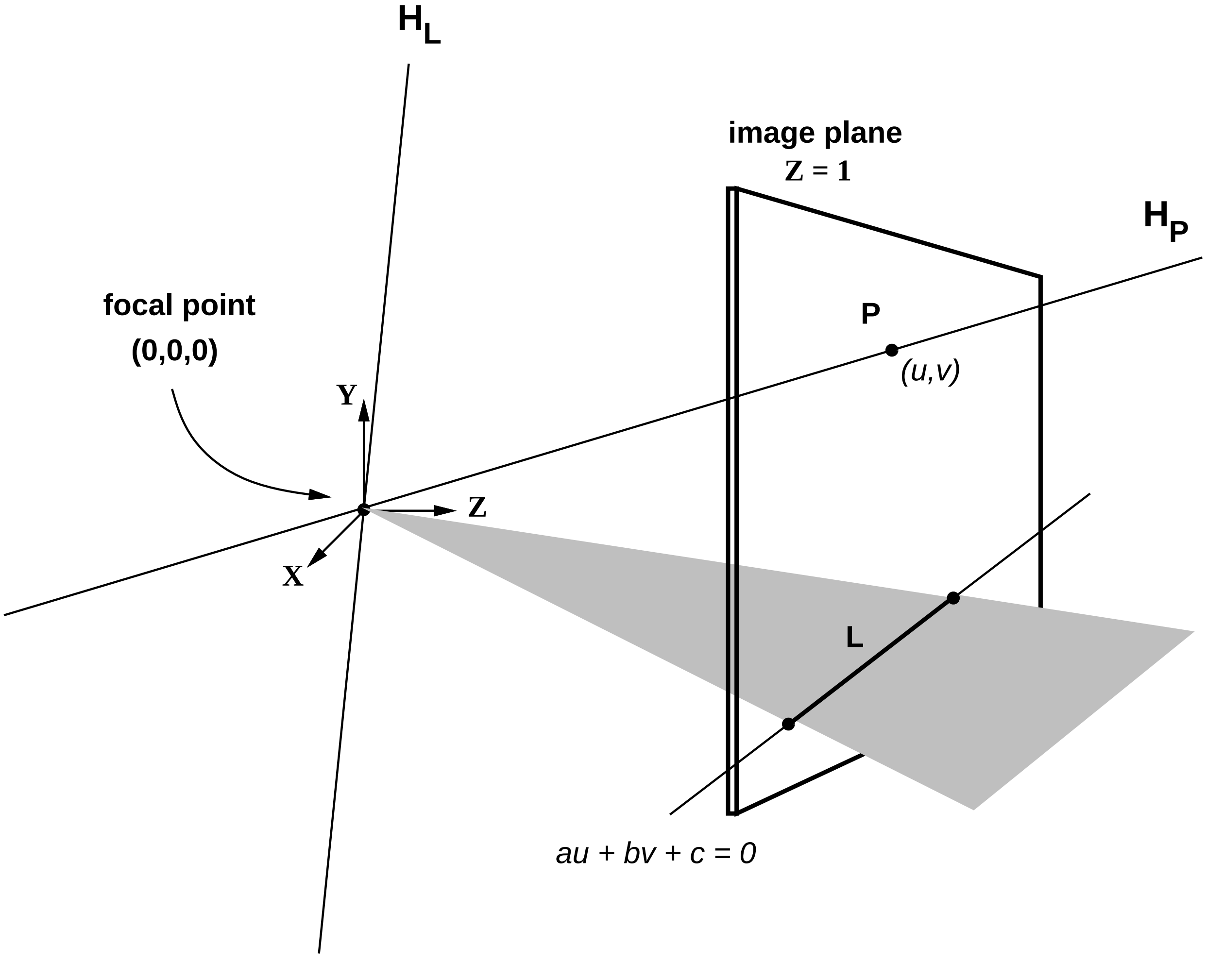


Figure 1.4: Homogeneous coordinates of points and lines in the image. Each equivalence class of homogeneous coordinates can be viewed as an infinite line through the origin of the camera coordinate system.

Although the homogeneous coordinates for points and lines in an image are not necessarily related to viewing rays towards 3D features in the world, for pinhole camera images they are directly related. This explains the emphasis that has been placed on accurate camera calibration for 3D scene reconstruction. **Associated with every calibrated camera is a known transformation that brings the space of homogeneous image coordinates for that camera into coincidence with the space of 3D view rays.** This transformation is applied to an observed image to recover the pure pinhole camera image. After applying this transform, the homogeneous coordinate vector of a point in the image gains a second, physical interpretation as the projection line containing the camera focal point, a point in the image, and the corresponding 3D point in the scene. Similarly, the homogeneous coordinate vector of a line segment in the image can be interpreted as a normal vector to the projection plane containing the focal point, image line, and corresponding 3D scene line.

Rather than dealing with a whole line through the origin, any point of which is equivalent to any other, it is easier computationally to scale homogeneous coordinate vectors to have unit length, resulting in points on the unit sphere. Since vector $\boldsymbol{x}$ is equivalent to vector $-\boldsymbol{x}$ in homogeneous coordinates, the resulting unit vector can have either sign, thus a point on one side of the sphere is indistinguishable from a point on the other. More formally, consider 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 $S^2 / \sim$, where $\sim$ is the equivalence relation defined in equation (1.6). Each equivalence class of this quotient space contains one pair of diametrically opposite points. Equating these equivalence classes with homogeneous coordinates in the obvious way shows that the surface of the unit sphere with antipodal points equated is isomorphic to the projective plane. This mapping also preserves the topological structure of the projective plane [Boothby86].

The sphere is a particularly appropriate model for the projective plane since it makes explicit the fact that both are compact surfaces. Furthermore, by identifying points and lines of the projective plane with pairs of antipodal points on the Euclidean unit sphere, a metric is induced on the projective plane. The introduction of a metric in the projective plane allows probability distributions on the sphere to be reinterpreted as distributions on 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 in a small neighborhood at any point on the sphere must be the same as at the diametrically opposite point.

The image plane is not normally identified with the projective plane. Indeed, many vision applications have developed plausible probabilistic measures of geometric uncertainty that treat the image plane as Euclidean. When such an error model is available, it may be mapped to the projective plane in a straightforward manner, again using a geometric interpretation of homogeneous coordinates. As shown in Figure 1.5, a probability density function (p.d.f.) representing an uncertain point feature in the image plane induces a p.d.f. over the unit vectors representing the homogeneous coordinates of that point, and thus a p.d.f. on the sphere. Likewise, a p.d.f. representing the uncertain endpoints of an extracted line segment induce a p.d.f. over the unit vectors normal to the projection plane of the line segment, again resulting in a p.d.f on the sphere. The definition and propagation of probability density functions on the sphere is the subject of Chapter 3.

Although this section has dealt primarily with the projective plane, certain aspects of the approach are extendible to other projective spaces. Obviously, the representation of uncertain points in the projective plane as random point processes on the sphere can be extended to any dimension. A point in projective n-space is represented by an antipodal

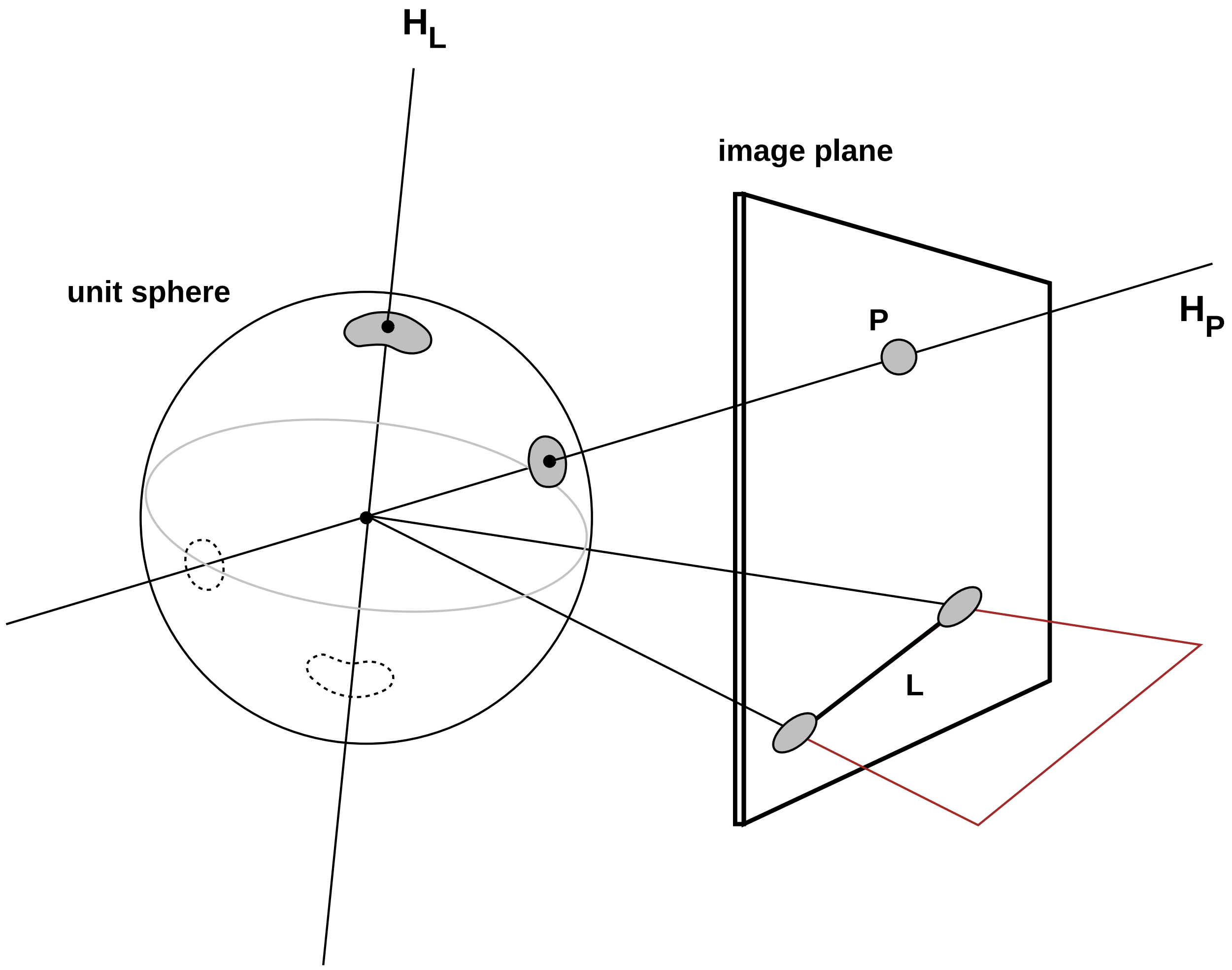


Figure 1.5: Mapping uncertainty from the image to the sphere. Probability distributions over points and lines in the image plane induce probability distributions over the space of homogeneous coordinates, thereby inducing antipodally symmetric distributions on the unit sphere.

point pair on the surface of the unit sphere in $(n + 1)$-dimensions. For example, points on the projective line are isomorphic to antipodal pairs on the unit circle, and points in projective 3-space can be represented as antipodal pairs on the unit hypersphere in R^4. The duality between points and lines in the projective plane (Appendix A.2), manifest by their essentially interchangeable homogeneous coordinate representations, is what allows 2D lines to be represented also by antipodal pairs of points on the sphere. In projective 3-space, points and planes are duals. Therefore, random point processes on the hypersphere apply equally well for planes as they do for points in projective 3-space. Lines in projective 3-space, however, are self-duals, and cannot be represented easily in this framework. The problem of representing uncertain infinite lines in projective 3-space is still an open problem.

1.3 Overview of the Thesis

The rest of this thesis is organized as follows. Chapter 2 reviews related work on geometric reasoning and the representation of geometric uncertainty. Chapter 3 presents an approach to uncertain geometric reasoning in projective space. The approach is called *stochastic projective geometry*, since geometric reasoning about homogeneous coordinate vectors is reduced to well-defined manipulations on probability density functions. Applications of this framework are presented in Chapter 4. Three applications are considered: deriving line and plane orientations using vanishing point analysis, partitioning scene features into planar patches using line correspondence stereo, and extending a partial model of planar surface structure using projective invariants. Chapter 5 concludes this thesis with a summary and some thoughts on future research directions.

CHAPTER 2

Related Work

No dissertation is written in a vacuum. This thesis owes much to the scholarship of others; the study of previous work has uncovered strengths to emulate and weaknesses to avoid. Along the way, ideas from seemingly disparate fields have been gathered and combined to form a novel, and hopefully enlightening, approach to managing the geometric uncertainty underlying visual scene reconstruction. This chapter traces the intellectual predecessors of this dissertation and points out concurrent related work. No attempt has been made to provide an exhaustive literature review. The works discussed here were chosen because they are either representative of a particular approach, illustrate an important point, or have contributed significantly to the development of this thesis.

The review is divided into two sections. Section 2.1 traces various stages in the development of tools for reasoning about visual geometry. It is seen that there has been a steady evolution of ideas, placing more and more emphasis on projective geometry: from the early block's world work on representing geometric constraints relating lines in the image plane with the orientation of lines and surfaces in the scene under orthographic projection, through the use of simple projective properties such as vanishing points to infer line and plane orientations, up to recent work involving the explicit use of projective invariants to derive 3D scene structure. In Section 2.2 the focus shifts to the representation and manipulation of geometric uncertainty, in particular the flexible and powerful probabilistic approaches currently in favor. In most computer vision applications, uncertainty in geometric features is represented and propagated using Gaussian probability density functions in a linearized space of object parameters. We show that this is not always an appropriate representation, particularly in projective space, and instead turn to probabilistic models of uncertainty on the sphere.

2.1 Geometric Reasoning

The pinhole camera model describes how a well-calibrated camera forms an image of the world. Under the pinhole camera model, point and line features in the image plane are directly related to the orientation of viewing rays towards corresponding scene features. Since the pinhole camera model is essentially a perspective projection of the scene onto the image plane, for calibrated cameras the space of homogeneous coordinates of features in the image coincides with the space of viewing ray orientations. The homogeneous coordinate vector of a point in the image plane represents the orientation of the ray in camera coordinate space that passes through the scene point in the world. The homogeneous coordinate vector of a line in the image is perpendicular to the projection plane containing the focal point, image line and 3D scene line. This implies that the homogeneous coordinate vector of a line in the image is perpendicular to the 3D orientation of the corresponding line in the scene.

The nice overlap of projective space and orientation space for the calibrated pinhole camera model perhaps explains why computer vision researchers have neglected the study of projective geometry until recently. This section traces a little of the evolution of thought in computer vision that has eventually teased apart the notions of orientation vs. projective space. Section 2.1.1 begins with a look at early systems for reasoning about simple block's world scenes, and how they represented geometric constraints between the orientations of 3D lines and planes in the scene, and the orientations of 2D lines in the image. Two approaches to representing orientations and their constraints were developed; gradient space and the Gaussian sphere. The work on vanishing point analysis described in Section 2.1.2 represents a move away from reasoning purely about orientations towards the simple, intuitive use of projective invariance for deriving scene structure. The original goal of vanishing point analysis was to recover line and plane orientations using a calibrated camera, so the difference between orientation space and projective space was still blurred. However, the fact that vanishing points can still be detected in images formed by uncalibrated cameras was soon exploited to perform the intrinsic camera calibration needed to bring orientation space into alignment with projective camera space. Finally, in Section 2.1.3 recent work is described that embraces projective geometry as a flexible reasoning tool. Geometric constructions in projective space are no longer required to have physical interpretations in terms of orientation space. The result is a powerful approach to deriving 3D scene structure in a manner invariant to viewing location or camera calibration parameters. The downside is that outside constraints must be provided to interpret a model built in projective space as a physical object in Euclidean space.

2.1.1 Reasoning about Orientations

Recovering the 3D geometry of a scene from a set of 2D geometric primitives in the image has long been a goal of computer vision research. Several early successes were achieved by systems for interpreting line drawings of polyhedral objects. The first systems for interpreting "block's world" images assigned symbolic labels – convex, concave, object boundary, or shadow – to the line segments in a drawing by determining sets of possible line labels at each vertex, then propagating label constraints to adjacent vertices until a consistent labeling of the whole edge graph emerged [Huffman71, Waltz72]. Two problems soon became apparent with this qualitative approach: objects with different shapes could be given the same set of edge labels, and consistent labelings could be generated for nonsensical objects that could not exist as polyhedra.

To overcome these difficulties, quantitative information about surface orientation was introduced into the interpretation process. Mackworth introduced the notion of *gradient space*, based on earlier work by Huffman, to represent constraints between plane and line orientations in the world, and their relation to line segment orientations in the image [Huffman71, Mackworth73].

Gradient Space

Gradient space represents an orientation (a, b, c) by the 2D point $(-a/c, -b/c)$. Geometrically, this can be interpreted in the pinhole camera coordinate system of Figure 1.1 as intersecting the orientation ray $k\,(a, b, c)$ with the plane $Z = -1$. General references to gradient space and its use in representing geometric constraints between the orientations of lines and surfaces can be found in [Shafer83] and [Kender80].

Gradient space has one serious flaw: it can not represent orientations perpendicular to the line of sight, namely those orientation vectors where $c = 0$. The representation is thus "blind" to an important subset of surfaces, namely surfaces parallel to the line of sight. For instance, when the camera is facing straight down a corridor, the side walls, the floor and the ceiling are all parallel to the line of sight, and thus not representable in gradient space. This was not a problem in the early block's world work where objects had limited extent in depth with respect to the viewing distance and orthographic (parallel) projection could be assumed. The orthographic projection of a plane nearly parallel to the line of sight is singular anyway, in that the plane projects to a single line in the image. However, when larger structures are viewed under perspective projection, planes parallel to the line of sight are often seen in the image and the limitations of gradient space become apparent. A different representation is needed to represent orientations under perspective projection.

Gaussian Sphere

Intuitively, a 3D orientation is most naturally represented as a unit vector having that orientation. This observation motivated Barnard to introduce the Gaussian sphere representation [Barnard83]. The Gaussian sphere is a unit sphere, centered about the origin of the camera coordinate system. A 3D orientation is represented as the point on the sphere where a unit vector having that same orientation touches the surface of the sphere.

Barnard found the Gaussian sphere to be a more convenient space to express geometric constraints under perspective projection than gradient space had been. He listed many reasons for preferring the Gaussian sphere over gradient space:

- all orientations are represented,
- the sphere is a finite (closed) space whereas gradient space is infinite (open), hence histogram-based methods become feasible,
- the sphere has the same symmetry as perspective projection (symmetric with respect to the focal point) whereas gradient space has the symmetry of orthographic projection (symmetric with respect to the line of sight), and

- gradient space is the limiting case of the Gaussian sphere representation as the focal length of the camera approaches infinity and perspective projection approaches orthography.

Every gradient space construction can be performed on the sphere by just mapping the relevant entities from the gradient space plane $Z = -1$ onto the sphere using central projection. That is, the point (p, q) in gradient space becomes point $(p, q, -1)/\sqrt{p^2 + q^2 + 1}$ on the sphere. Lines of points in gradient space map to great circles of points on the sphere. What becomes apparent after working with the sphere for a short while is that many relations have a clear, intuitive geometric meaning in terms of vectors, great circles, and perpendicularity. Once these intuitions on the sphere have been built up, the corresponding geometric interpretations in gradient space seem needlessly opaque.

Kender uses both gradient space and the Gaussian sphere in his work [Kender80]. He cites the biggest drawback of the Gaussian sphere as being the complicated form that equations take on when written in spherical coordinates. This is a common complaint, even among advocates of the Gaussian sphere representation. For example, McGee and Aggarwal use a clustering method on the sphere to detect vanishing point locations [Magee84]. Their measurement of arc distance between two points with azimuth-elevation coordinates (α_1, β_1) and (α_2, β_2) is the nonintuitive formula

$$\delta = \arccos\{\cos(90 - \beta_1)\cos(90 - \beta_2) + \sin(90 - \beta_1)\sin(90 - \beta_2)\cos(\alpha_1 - \alpha_2)\}.$$

Criticism of the Gaussian sphere based on complicated trigonometric expressions misses its mark, however, since most of the relevant relationships on the sphere can be written simply and concisely in terms of unit vector notation, rather than using spherical coordinates. For instance, the arc distance equation given above can be written much more simply as the unit vector formula

$$\delta = \arccos(\boldsymbol{u} \cdot \boldsymbol{v})$$

where $\boldsymbol{u}$ and $\boldsymbol{v}$ are unit vectors representing two points on the sphere.

2.1.2 Vanishing Point Analysis

Under perspective projection, 3D parallel lines in the world project to converging 2D lines in the image plane. The common point of intersection, perhaps at infinity, is called the *vanishing point*. Vanishing points provide strong cues for inferring 3D scene structure from only a single view [Haralick80, Horaud87, Shakunaga92, Straforini92]. For calibrated cameras, the vanishing point of a group of parallel lines determines their orientation in space. Two or more vanishing points from lines known to lie in a plane establish a *vanishing line,* which completely determines the orientation of the plane. Analysis of vanishing points yields simple methods for finding the relative orientation of the viewer with respect to the world [Bellutta89, Li90, Wang91], and for determining some of the intrinsic camera calibration parameters [Caprile90, Kanatani92].

By tracing the development of vanishing point algorithms in some detail, we hope to illustrate the interplay between orientation space and projective space in computer vision. The history of vanishing point analysis also provides a nice case study of how heuristic approaches for dealing with geometric uncertainty have given way to more formal statistical estimation techniques.

Vanishing point geometry is easiest to describe for calibrated cameras (see Figure 2.1). The normal vector of the projection plane of a 3D line is perpendicular to the orientation of the line. The projection plane normals of parallel lines in space are therefore all perpendicular

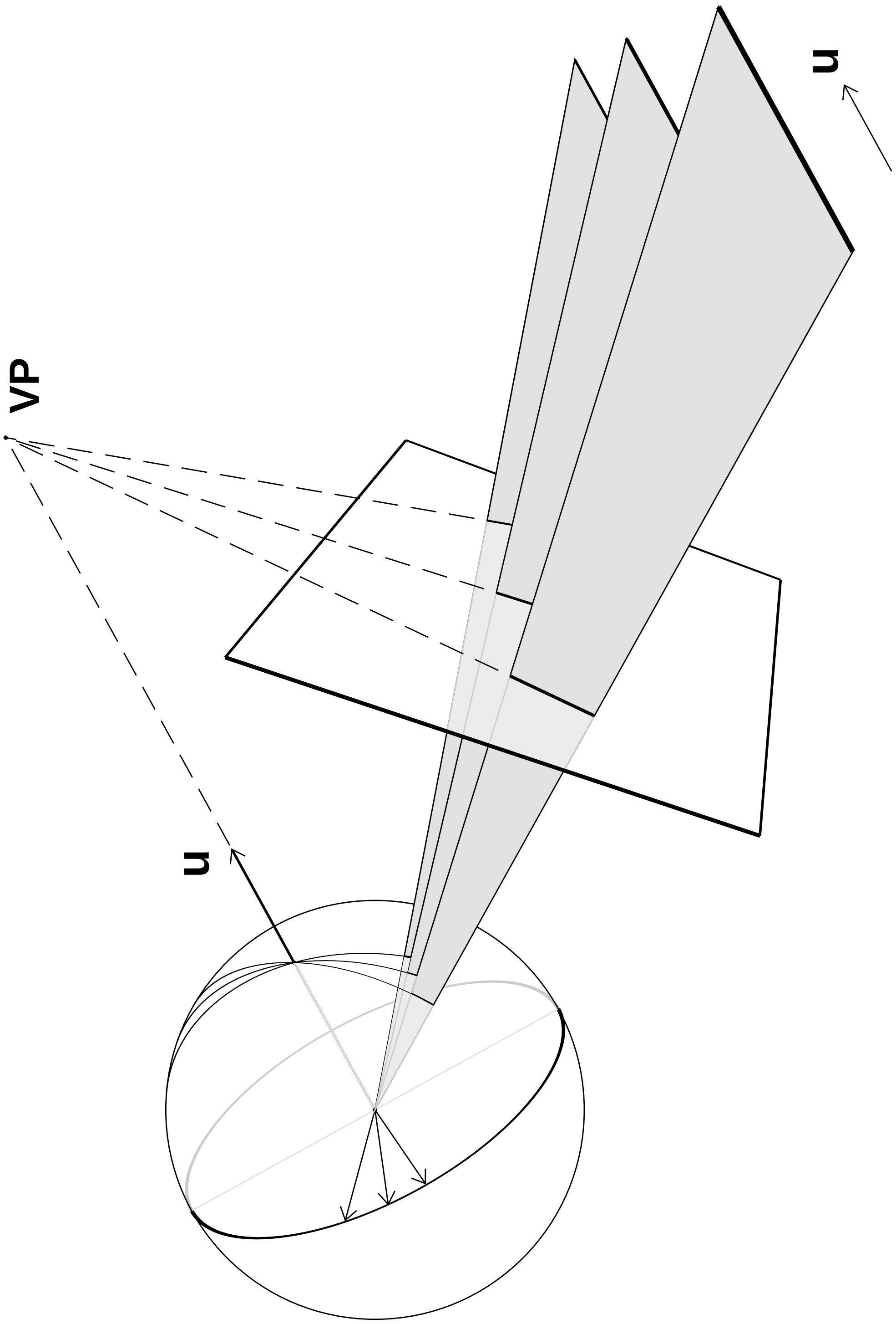


Figure 2.1: Vanishing point geometry.

to the shared line orientation $\boldsymbol{u}$. As a result, the projection planes of parallel lines are constrained to form a *pencil of planes*, all intersecting in the line through the origin having orientation $\boldsymbol{u}$. The intersection of this pencil of planes with the image plane yields a *pencil of 2D lines* intersecting at a single point, the vanishing point. The vanishing point associated with lines of orientation $\boldsymbol{u}$ is the image of the point at infinity in projective space where all parallel 3D lines of orientation $\boldsymbol{u}$ meet.

The pencil of projection planes for lines of orientation $\boldsymbol{u}$ intersects the Gaussian sphere in a pencil of great circles. All great circles in this pencil intersect at two antipodal points marking the tips of unit vectors $\pm\boldsymbol{u}$. This simple geometric constraint leads to an efficient clustering technique for detecting pencils of lines converging to a vanishing point. In Barnard [Barnard83], great circles representing the projection planes of line segments in the image are mapped onto a histogram of the Gaussian hemisphere, partitioned by azimuth and elevation. Each histogram bucket maintains a count of the number of great circles passing through it. Potential vanishing points are detected as peaks in the histogram, corresponding to areas where several great circles intersect.

While the computational complexity of Barnard's algorithm is $\mathcal{O}(n)$ in the number of lines processed, vanishing points are located only to within the boundaries of a histogram bucket. This makes the approach ill-suited for accurately determining the true vanishing point location. Usually, the center of a histogram bucket is chosen as the vanishing point location estimate, and a confidence region is constructed having roughly the same size and shape as the bucket. Since the surface of the sphere is not partitioned uniformly by azimuth and elevation, this method estimates vanishing points more accurately in some directions than in others. If high accuracy is required over the whole space of vanishing point orientations, the spherical histogram must be partitioned very finely, negating some of the computational benefits of the approach. This drawback can be overcome by using a more sophisticated algorithm, such as the hierarchical refinement method of [Quan89], a more regular tessellation of the sphere, or a different accumulator space altogether where the likelihood of finding a vanishing point in each bucket remains roughly constant [Brillault91]. Other problems (inherent to the Hough transform) remain [Grimson90]; for example, when the true vanishing point falls near a histogram boundary, candidates which should be grouped together may fall into separate buckets. In summary, while algorithms based on the Hough transform excel at quickly clustering line segments into convergent groups, the final estimate of vanishing point location and variance should be based on the line segments themselves rather than the arbitrary bucket boundaries of a histogram data structure.

In contrast to line-based Hough transform approaches, Magee and Aggarwal compute intersections of pairs of line segments directly, thus determining pairwise vanishing points [Magee84]. A line intersection is represented as the point on the unit sphere where a ray directed from the focal point towards the intersection point pierces the sphere. This ray can be easily computed as the cross product of the projection plane normals of the two lines. Vanishing points are detected as clusters of intersection points on the sphere. Since computed points are maintained and distances compared uniformly over the entire surface, vanishing points can be estimated with greater accuracy. One drawback, of course, is that examining all pairs of line segments yields a complexity of $\mathcal{O}(n^2)$. Magee and Aggarwal further observe that the true focal length of the camera does not have to be known to find vanishing point clusters, although it *is* necessary for deriving line orientation. This is because a line pencil is a geometric configuration that is invariant under projective transformations caused by changing the intrinsic camera parameters. If the wrong focal length is used, all vanishing points can still be located, but their projection rays will no longer be aligned with the true Euclidean orientations of lines in the scene.

A practical algorithm for finding vanishing points must address two issues: how to cluster line segments going to a single vanishing point, and how to accurately estimate the vanishing point location from a given line cluster. Barnard's histogram method provides an effective solution to the first subproblem, based on the constraint that the projection planes of a pencil of lines slice through the Gaussian sphere in a pencil of great circles. Collins and Weiss [Collins89] present an efficient solution to the estimation subproblem based on the dual constraint: that the normal vectors of a pencil of planes must themselves lie in a plane, which implies that the unit normal vectors on the sphere span a great circle. They use the normal vector to the orthogonal least-squares plane-fit as the estimate of the true vanishing point direction. This approach maintains the $\mathcal{O}(n)$ efficiency of Barnard's method but yields vanishing point estimates having far greater accuracy. In particular, Barnard's histogram is applied as an initial clustering method and as an efficient spatial access mechanism on the sphere, while the final estimation of vanishing point location is performed on the underlying data.

In real images, line segments are perturbed by errors both in the imaging process and in the method of line segment extraction. The uncertainty in line segment position and slope leads to uncertainty in the true vanishing point location and hence in derived 3D line orientation. As mentioned above, histogram-based methods locate each vanishing point within a histogram bucket on the sphere, the shape of the histogram bucket itself providing a discrete error bound on the true position. Weiss, Nakatani and Riseman [Weiss90] represent uncertainty in 3D line orientation as a circular error bound around the nominal vanishing point position on the sphere. This constrains any vector perpendicular to an uncertain line orientation to lie somewhere within a "belt" of constant width on the sphere, and therefore the surface normal of a planar patch containing two or more uncertain line orientations lies somewhere within the intersection of their constraint belts.

Collins and Weiss [Collins90] present a formal statistical approach to estimating uncertain line orientations from vanishing points. The approach is described more fully in Chapter 4. Assuming that image line segments have been previously grouped into pencils, the projection plane normals for each group form a set of points on the sphere clustered about a great circle perpendicular to the true vanishing point location. Collins and Weiss treat this cluster as a random sample from an equatorial distribution on the sphere, and estimate the vanishing point location as the *polar axis* of least dispersion. Under the assumption that the sample is distributed according to Bingham's distribution on the sphere, the polar axis can be found using standard maximum likelihood techniques [Bingham74]. A second, nonparametric approach is based on the method of moments, requiring only the assumption that the distribution is antipodally symmetric. Both approaches yield the eigenvector associated with the smallest eigenvalue of the sample second moment matrix as an estimate of the axis of least dispersion, which interestingly enough is the same result produced by a least-squares perpendicular error plane fit to the heads of the projection plane normals [Collins89]. Both distributional assumptions provide confidence regions for the estimated axis that are fairly similar in size and shape.

The work of Collins and Weiss represents the first application of probability distributions on the sphere to estimate vanishing point locations. However, it is based on the unrealistic assumption that the projection plane normals of a pencil of line segments are independent and identically distributed according to an equatorial distribution on the sphere. This is unrealistic for two reasons. First, longer line segments are measured much more accurately than shorter ones, and should thus be given more credence when estimating their intersection. Yet even if a weighted variant of maximum likelihood estimation or the method of moments is substituted, a more subtle problem remains. Each projection plane normal on the sphere represents a noisy observation of an unknown, true value. Collins and Weiss assume these true values are themselves random samples from an equatorial distribution. There is no reason to believe the true values are distributed according to any particular distribution, however. What is instead the case is that the true values are constrained to lie on a great circle by the geometry of vanishing point formation. In formal terms [Anderson84] the tendency of projection plane normals to cluster about a great circle should be treated as a *functional* relationship (*i.e.* a geometric constraint) rather than a *structural* relationship (*i.e.* evidence of an underlying stochastic process).

Kanatani [Kanatani93b] removes some of these troublesome assumptions. To reflect the differing accuracy of individual lines and their projection plane normals, each vector is weighted by an estimate of its uncertainty, represented as a 2D Gaussian distribution in the tangent plane to the sphere centered about the tip of the nominal vector location. These projection plane uncertainties are derived from an error model of line segment extraction in the image. An iterative search finds the axis that minimizes a least-squares, statistically weighted dispersion across the cluster of points on the sphere. Uncertainty in this computed vanishing point location is approximated by a Gaussian distribution in the tangent plane to the estimated axis. Although this approach removes many of the defects of the earlier work of Collins and Weiss, it does so at the cost of oversimplifying the representation of uncertain unit vectors. What should ideally be a probability distribution on the sphere is instead represented in a linearized parameter space, namely a tangent plane to the sphere. While there is little difference for small errors, larger errors typified by vanishing points far from the image plane will be underestimated, because the inherent curvature of the parameter space of 3D orientations is being ignored.

In Chapter 4 we revisit the classic problem of vanishing point estimation. As in [Collins90] we use probability distributions on the sphere to characterize uncertainties in unit vector orientations. As in [Kanatani93b] we allow each projection plane normal to have an independent estimate of accuracy, and make no assumptions about the distribution of sample points except that they are noisy observations of true values that lie on a great circle whose axis identifies the location of the vanishing point. This axis is estimated using a Bayesian nonlinear regression technique.

2.1.3 Projective Geometry

Due to the projective nature of computer vision, it is not surprising that past work has been implicitly based upon the principles of projective geometry. However, only in the past five years have appreciable numbers of researchers explicitly focused on projective geometry as a problem solving framework. Among the first were Naeve and Eklundh, who argue that projective space is the proper place to formulate questions about scene geometry [Naeve87]. They point out that representations in terms of homogeneous coordinates yield geometric equations of particularly simple form, usually linear. As a result, problems solved in projective form tend to make explicit the degrees of freedom of the solution. Finally, projective spaces are compact, so the search space for analytic solutions is bounded, and histogram-based methods for grouping data become possible. Naeve illustrates these advantages by developing projective coordinates for representing 3D lines in the world, and analyzing what this coordinate system implies about solutions to common problems such as finding parallel 3D lines from a 2D image (the vanishing point problem), and finding pairs of lines in the image which form right angles in the world.

Both the INRIA and LIFIA research teams in France made significant early progress in applying projective geometry for use in vision applications. Quan and Mohr use perspective invariants such as colinearity and convergence to dictate how to group 2D line segments into larger structures relevant for matching against 3D object models [Quan89]. The main idea is that initial matching hypotheses are more likely to succeed when they are based on geometric structures which remain intact throughout the image formation process. They also discuss the use of the cross ratio as a consistency check to prune out bad matches. Faugeras and Lustman show that when coplanar points (or lines) are viewed from two different positions, the relation between their corresponding 2D locations in the image planes is a homography [Faugeras88]. They use iterative mean square techniques to estimate the 3×3 homography matrix from several hypothesized point (line) matches. They show how to decompose the homography matrix to solve for the relative pose of the two viewing positions, and discuss ambiguities in the solution. The results roughly duplicate those of Tsai, Huang, and Zhu, who solved the same problem but did not discuss the problem in projective geometric terms [Tsai82]. Finally, Mohr and Morin use projective invariant point and line coordinates for determining 3D scene structure without laborious calibration of the camera [Mohr91].

Forsyth *et.al.* investigate the use of projective invariant shape descriptors to index a model base of curved planar objects [Forsyth91]. Object curves are fitted with conics, and then labeled with the projective invariants of those conics. Models in the database are stored according to those invariants as well. This effectively decouples the problem of object identification from that of pose determination. After extracting the appropriate object model from the database, the accurate 3D object pose can be determined by finding the perspective transformation that maps two known conics on the object plane to two conics found in the image plane. Follow-up work has investigated invariant indexing functions for other planar feature sets such as a conic and three lines, and five coplanar lines [Rothwell92].

The early successes in object recognition based on planar invariants sparked considerable interest in developing more general 3D-to-2D invariants. However, work by Burns shows that there is no holy grail — there are no 3D-to-2D invariants for general configurations of points in space [Burns93]. This has been widely misinterpreted as meaning that <u>no</u> 3D-to-2D invariants exist for any object. Fortunately, this is not the case, as Rothwell *et.al.* show by exhibiting invariants for sets of points lying on the vertices of a polyhedron, and for objects having bilateral symmetry [Rothwell93].

Projective invariants are insensitive not only to camera position, but also to the linear intrinsic parameters of the camera itself. This has led researchers to consider what 3D structure can be recovered from views of a scene taken by uncalibrated cameras. One fundamental result is now known: given an uncalibrated stereo pair, i.e. two views taken from unknown viewing position by two different unknown cameras, the 3D structure of the world can still be recovered *up to an unknown projective transformation* [Faugeras92a, Hartley93, Mohr92, Shashua93]. The recovery and use of non-metric structure is currently a subject of intense interest in the field.

2.2 Geometric Probability

The ability to represent and propagate uncertainty in geometric models, sensed data, and estimated transformations is an important requirement for robust vision and robotics systems. This issue pervades numerous application areas such as mobile robot navigation, automated assembly planning, dexterous hand manipulation, and geometric model acquisition.

In this section standard techniques for handing uncertainty in vision and robotic systems are examined. Most geometric reasoning tasks can be formulated as the problem of estimating a state vector of parameters from a set of noisy observations related to the parameters by nonlinear constraint equations. Statistical methods based on the Gaussian probability density are the most prevalent and powerful approaches for solving this estimation problem. Such methods include the popular Kalman filter. At the end of this section we analyze the assumption of Gaussian errors in the context of uncertain projective geometric reasoning and find the Gaussian distribution lacking. In the next Chapter, we explore alternatives to the Gaussian that make sense in projective space.

2.2.1 Stochastic Projective Geometry

Geometric features are often represented algebraically as parameterized functions. Particular instances of these geometric features can then be represented as points in the appropriate space of parameters. For example, the equation $x \cos\theta + y \sin\theta - \rho = 0$ represents a line of infinite extent in the affine plane. The parameter space $\{(\theta, \rho) | 0 \leq \theta < 2\pi, 0 \leq \rho\}$ represents the space of all possible lines in the plane; a particular line is singled out by choosing one pair of values for θ and ρ.

One simple, yet popular approach to representing uncertain geometric features is to specify a a region in parameter space within which the true parameter value is known to lie [Brooks84, Grimson90]. Since the true value can be anywhere within its allowed bounds, this method implies a uniform uncertainty over the circumscribed volume of parameter space. Uncertain information is combined by intersecting corresponding volumes. However, the geometric volume representing the result of even one or two intersections can become very difficult to describe. In practice, the uncertainty bound approach often degenerates into a crude, worst-case analysis using volumes of simple shape, but imprecise content.

A more powerful representation for geometric uncertainty is based on *stochastic geometry* [Baddeley82, Stoyan87]. Uncertainty in a geometric object is represented as a probability density function over the space of object parameters, higher probabilities representing more likely parameter values. Applications of this approach in robotics and computer vision typically use the Gaussian distribution, so that uncertainty in a parameter is represented by a mean value plus an associated covariance matrix [Durrant-Whyte88]. Durrant-Whyte considers two geometric reasoning tasks : transferring uncertain features into a new coordinate system (to bring measurements made from two camera positions into a common reference frame, for example), and building new geometric object descriptions from existing uncertain features (*e.g.* aggregating two uncertain point features into an uncertain line passing through them). Both types of problem are solved by computing the mean and covariance of parameter vector $\boldsymbol{y} = f(\boldsymbol{x})$ where $\boldsymbol{x}$ is a multivariate Gaussian variable. When f is a linear function, $\boldsymbol{y}$ will also be Gaussian. In general f will be nonlinear, but by expanding the function in a Taylor series about the mean parameter value for $\boldsymbol{x}$, f can be approximated by a linear function. Propagating the mean and covariance of $\boldsymbol{x}$ through this "linearized" version of f results in the approximate mean and covariance of $\boldsymbol{y}$.

Crowley notes that estimates of Gaussian uncertainty can aid in determining compatible feature matches via the Mahalanobis measure of statistical distance in parameter space [Crowley87]. Gaussian uncertainty propagation can therefore form the basis of a cyclic processing paradigm where an uncertain prior model of the world guides the interpretation of sensed data, which then is used to constrain the estimated location of the observer, which in turn is then used to further refine the world model. This *cycle of uncertainty and constraint* lies at the core of many current robotic navigation systems [Matthies87a].

A notable example of the cycle of uncertainty and constraint is the work of Ayache and Faugeras [Ayache91], who describe a successful model-based stereo navigation system. The strength of this system is its unified, homogeneous treatment of uncertainty in all stages of model construction, image interpretation and motion determination. All tasks are reduced to the estimation of a state vector of parameters given noisy observations that are related to the current system parameters via multiple, nonlinear constraint equations. All observation errors are assumed to be Gaussian, unbiased and independent of each other, with known covariance. Estimation of the state parameters is performed within the framework of Kalman filtering, an iterative version of the weighted least-squares estimation procedure that incrementally improves parameter estimates as more observations are built up over time.

2.2.2 Problems with the Gaussian Assumption

Parameter estimates produced by the Kalman filter are provably optimal provided the equations relating observations with parameters are linear and the measurement noise is truly Gaussian, When nonlinear constraint equations are used (the usual case) they must first be linearized about current parameter estimates. The implicit linearization of the parameter space due to the assumption of Gaussian noise, plus the enforced linearization of the constraint equations, make the Kalman filter in practice an approximate method. In particular, approaches to uncertain geometric reasoning based on Gaussian covariance propagation will not work well when

- the uncertainty is large, or
- the parameter spaces or transformations considered are highly nonlinear, or

- the true error distribution is significantly non-Gaussian.

Indeed, when these three constraints are not met, problems have been noted in the literature [Matthies87b, Maybank90b].

A further problem with the Gaussian assumption, at least in the context of this thesis, is that the Gaussian is not *stochastically well-conditioned* under projective transformations [Durrant-Whyte88]. This means that a Gaussian density function is no longer Gaussian after being subjected to a projective transformation, and approximating the resulting true density function using linear covariance propagation can yield extremely misleading results. Not only is a projectively transformed Gaussian no longer Gaussian, the constant probability contours of the transformed function are no longer ellipses, or even conic sections. Worst of all, the resulting density function may even turn bimodal. These statements can be derived formally using results presented in Appendix C, example C.9.

Gaussian covariance propagation will perform badly under projective transformations whenever the transformed mean approaches infinity. This is primarily due to the fact that the projective plane is compact, while the affine plane (where the Gaussian is defined), is not. A density function that is unimodal in the projective plane is bimodal in the affine plane whenever an appreciable amount of probability mass straddles the line at infinity. Since the Gaussian density function is defined over the affine plane, and is unimodal, it cannot adequately approximate these situations. This problem is solved in the next chapter by using probability density functions defined over the projective plane instead.

CHAPTER 3

Stochastic Projective Geometry

Geometric measurements made by real camera systems are inherently uncertain, due to causes ranging from random sensor noise to grouping and fitting errors in the low-level feature extraction routines. Algorithms for scene reconstruction that do not take this uncertainty into account run the risk of generating invalid or poor results. In this chapter we present an inference engine for performing uncertain geometric reasoning during the scene reconstruction process. We call this formalism *stochastic projective geometry* to highlight the similarity between this approach and the popular *stochastic geometry* method described in Section 2.2.1. Just as in that work, geometric objects are represented by points in a parameter space, and geometric uncertainty is represented by probability distributions over the space of object parameters. Geometric reasoning about uncertain objects reduces to well-defined manipulations of the underlying probability distributions.

The distinguishing character of *this* work, that makes it *stochastic projective geometry,* is that the geometric objects under consideration are projective objects and the appropriate parameter space for representing them is projective space. Earlier chapters have shown that projective space has a different global structure than its affine and Euclidean counterparts, and that probability distributions in these spaces, particularly the Gaussian, are not appropriate in topologically compact projective space.

In this chapter the fundamental principles of stochastic projective geometry are defined. The goal is to develop an uncertainty calculus in projective space analogous to the Gaussian uncertainty calculus in affine space. Above all, we seek to maintain the main strength of the Gaussian calculus, namely its uniform treatment of uncertainty in all stages of the geometric reasoning process. At the same time, the limitations of the Gaussian as a representation of uncertainty in projective space are removed.

For concreteness, the methods of this chapter are developed specifically for the projective plane, but the results are easily generalized to projective spaces of other dimension. In Chapter 1 the projective plane was shown to be topologically equivalent to the quotient space formed from the surface of the Euclidean unit sphere modulo the set of infinite lines through the sphere's center. This quotient space amounts to equating antipodal pairs of points on the surface of the sphere. By identifying the projective plane with this quotient space, antipodally symmetric probability density functions on the sphere to be reinterpreted as density functions on projective space. In Section 3.1 a family of exponential distributions on the sphere is considered, and Bingham's distribution is chosen from this family to represent uncertainty over points and lines in the projective plane.

Once a suitable representation of uncertainty has been chosen over a parameter space, there are two fundamental types of operations of interest to an uncertain geometric reasoning system. The first is determining how the chosen representation of uncertainty transforms under a change of coordinates or under a change in parametric representation. The coordinate transformations of interest in plane projective geometry are the nonsingular planar projective transformations or homographies. Section 3.2 shows that a projective transformation, although linear in homogeneous coordinates, induces a nonlinear bijective transformation on the surface of the sphere. Using a method called *transformation of variables*, the change in point and line uncertainty due to a projective coordinate transformation is determined by propagating the Bingham probability density function through this nonlinear transformation. Also considered is the initialization of Bingham uncertainty measures by mapping representations of uncertain points and lines in the image plane onto the sphere via central projection.

We shall see that Bingham's distribution does not remain invariant under projective transformations. In order to maintain a uniform representation of uncertainty, it will be necessary to approximate an analytic density function computed by the transformation of variables technique by an appropriate Bingham density. Three methods for approximating a given density function by a Bingham density are explored in Section 3.3. These methods are compared in terms of the mean-squared error metric.

After several geometric observations from multiple sensors and/or multiple images have been brought into a comparable coordinate system and found to represent separate measurements of the same phenomenon by some test, a second fundamental geometric reasoning task involves *fusing* these multiple estimates into a single, more accurate geometric description. Section 3.4 examines the incidence relations of projective geometry, reformulating them to take into account uncertainty in the underlying point and line primitives. Bayesian parameter estimation is used on the sphere to combine multiple measurements of incident points and lines to derive more accurate feature descriptions. This chapter closes with a brief recap of the main principles underlying stochastic projective geometry as an approach to geometric reasoning in projective space.

In an attempt to make this document as self-contained as possible, basic background material regarding probability density functions on the sphere is presented in Appendix B.

A description of the transformation of variables technique, and its use in deriving the results referred to in the text, can be found in Appendix C.

3.1 Probability in the Projective Plane

In Section 1.2.2 it was shown via homogeneous coordinates that the projective plane can be visualized as the Euclidean unit sphere with antipodal points equated. By identifying the projective plane with the quotient space $S^2/\sim$, antipodally symmetric probability distributions on the sphere may be reinterpreted as distributions in the projective plane. A useful characterization of probability distributions on the sphere is presented by Beran [Beran79]. Beran considers a family of probability density functions of the form $exp\{P\}$, with P being some polynomial evaluated over the surface of the sphere (see Appendix B). Although limiting the density functions considered to those having exponential form may appear restrictive at first, it is not in fact a very severe restriction, since any strictly positive analytic function F on the sphere can be represented as $exp\{log\{F\}\}$. Probability density functions having exponential form play an important role in statistics due to their ease of use in maximum likelihood estimation [Mendel87].

Beran chooses the spherical harmonic polynomial basis for writing polynomial P, yielding a polynomial decomposition on the sphere analogous to the decomposition of polynomials in Euclidean space using Fourier analysis. Let Y_n denote a spherical harmonic polynomial consisting only of terms of degree n. If a density function on the sphere is required to have antipodal symmetry, all the odd order harmonic polynomials in its expansion must be identically zero. This leaves an expression $exp\{Y_0 + Y_2 + Y_4 + \ldots\}$. The zeroth-order harmonic Y_0 is just a constant term, so $exp\{Y_0\}$ can be factored out of the expression and absorbed into the density function's normalization constant. Therefore, the lowest-order approximation to any antipodally symmetric density function on the sphere will have the form $C exp\{Y_2\}$. A distribution having this form has already been studied in the statistical literature, where it is called Bingham's distribution [Bingham74].

Principle 1: Uncertainty in the projective plane will be represented as a Bingham probability density function, because this is a computationally convenient approximation to any antipodally symmetric density function on the sphere.

The following two subsections provide a review of Bingham's distribution, and a comparison with alternative distributions that could have been used instead.

3.1.1 Bingham's Distribution

Bingham's distribution is a standard probability distribution for representing both bipolar and equatorial clusters of points on the sphere [Bingham74, Mardia72]. It is a second-order approximation to any antipodally symmetric probability distribution on the sphere, and plays a role analogous to the Gaussian distribution on the sphere and the projective plane [Breitenberger63]. A Bingham random process can be constructed by taking $N_3(\boldsymbol{x}; 0, \boldsymbol{\Sigma})$, a trivariate Gaussian with zero mean and arbitrary covariance matrix, and conditioning variable $\boldsymbol{x}$ to have unit length. Bingham's distribution thus represents the portion of a trivariate Gaussian distribution that intersects the surface of the unit sphere, normalized so that the total probability density over the sphere integrates to one. Varying ellipsoidal shapes of the underlying Gaussian contours produce a variety of distributional forms on the sphere.

The real symmetric covariance matrix $\boldsymbol{\Sigma}$ of a trivariate Gaussian can always be decomposed as $\boldsymbol{U}\boldsymbol{S}\boldsymbol{U}^t$, where $\boldsymbol{U} = [\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}]$ is a 3×3 orthonormal matrix and $\boldsymbol{S} = diag(\sigma_1^2, \sigma_2^2, \sigma_3^2)$ is a diagonal matrix of variances. Without loss of generality, assume that $\sigma_1^2 \leq \sigma_2^2 \leq \sigma_3^2$. Letting $\boldsymbol{K} = diag(\kappa_1, \kappa_2, \kappa_3) = -\frac{1}{2}\boldsymbol{S}^{-1}$ and constraining $\boldsymbol{x}$ to have unit length yields the Bingham density function

$$\begin{aligned} \mathcal{B}(\boldsymbol{x}; \boldsymbol{U}, \boldsymbol{K}) &= C(\boldsymbol{K})\, exp\{\boldsymbol{x}^t \boldsymbol{U}\boldsymbol{K}\boldsymbol{U}^t \boldsymbol{x}\} \qquad (3.1) \\ &= C(\boldsymbol{K})\, exp\{\boldsymbol{x}^t\, (\kappa_1\, \boldsymbol{u_1}\boldsymbol{u_1^t} + \kappa_2\, \boldsymbol{u_2}\boldsymbol{u_2^t} + \kappa_3\, \boldsymbol{u_3}\boldsymbol{u_3^t})\, \boldsymbol{x}\} \\ &= C(\boldsymbol{K})\, exp\{\kappa_1(\boldsymbol{u_1^t}\boldsymbol{x})^2 + \kappa_2(\boldsymbol{u_2^t}\boldsymbol{x})^2 + \kappa_3(\boldsymbol{u_3^t}\boldsymbol{x})^2\}\ , \end{aligned}$$

where $C(\boldsymbol{K})$ is the normalizing constant required to make $\int \mathcal{B}(\boldsymbol{x}; \boldsymbol{U}, \boldsymbol{K})\, dS^2 = 1$.

It is easy to verify that $\mathcal{B}(-\boldsymbol{x}; \boldsymbol{U}, \boldsymbol{K}) = \mathcal{B}(\boldsymbol{x}; \boldsymbol{U}, \boldsymbol{K})$, showing that the density function is antipodally symmetric, and that it is appropriate to interpret $\boldsymbol{x}$ as an undirected axis, or as a point in the projective plane. Further analysis shows that the "shape" parameters κ_i are determined only up to an additive constant. This is so because points on the surface of the unit sphere really have only two degrees of freedom. For uniqueness it is customary to subtract out the largest κ_i, in this case κ_3, leaving $k_i = \kappa_i - \kappa_3$ with $k_1 \leq k_2 \leq k_3 = 0$.[1] The relative values of the canonical Bingham shape parameters k_1 and k_2 characterize the different distributional forms of Figure 3.1.

In geometric terms, the additive indeterminacy of the Bingham shape parameters means that there are an infinite number of trivariate zero-mean Gaussians that produce the same

[1]This is an arbitrary decision. The additive indeterminacy of the shape parameters could also be removed by requiring $\kappa_1 + \kappa_2 + \kappa_3 = 1$, for example. Subtracting out the largest parameter κ_3 is the convention that has been adopted in the literature.

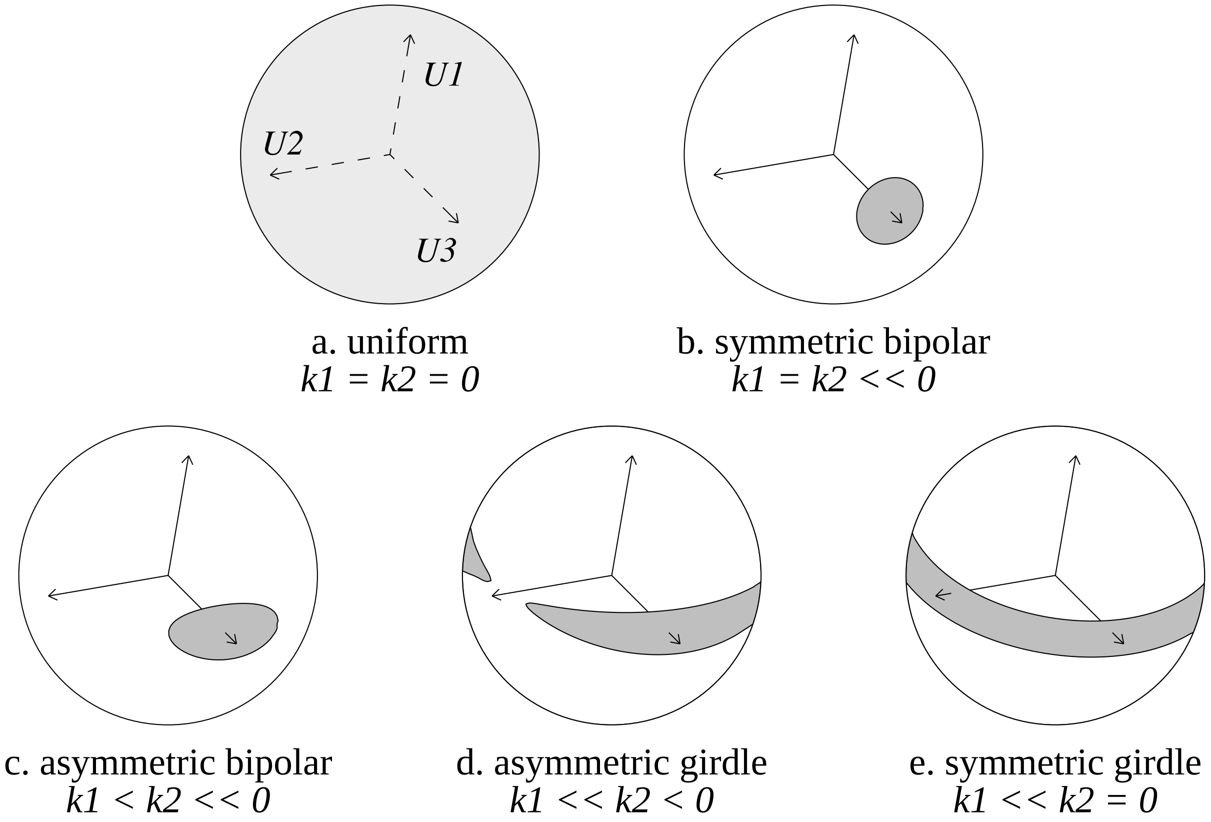


Figure 3.1: Bingham's distribution. Representative contours are shown for varying magnitudes of the canonical Bingham shape parameters k_1 and k_2.

Bingham density function when normalized over the surface of the sphere. Removing this indeterminacy by constraining the largest parameter k_3 to be zero corresponds to picking a singular Gaussian as the unique representative for that Bingham density. The Gaussian is called *singular* because the determinant of its inverse covariance matrix is zero. In terms of variance, this means that the representative Gaussian has infinite variance along (at least) one of its axes.

Bingham's distribution can easily be generalized to spheres of arbitrary dimension. For the unit sphere S^n, a Bingham density function is defined by intersecting an $(n+1)$-dimensional zero-mean singular Gaussian with the surface of the sphere, then renormalizing so that the density on the sphere integrates to one. The resulting distribution is described by an $(n+1) \times (n+1)$ orthonormal matrix of orientation parameters, and n independent shape parameters $k_1 \leq \cdots \leq k_n \leq 0$. Bingham distributions on the unit circle, sphere, and hypersphere represent probability distributions over points on the projective line, the projective plane, and projective three-space, respectively.

Example 3.1 *Bingham's distribution on S^1 is formed by normalizing a zero-mean bivariate Gaussian over the unit circle. Let the Gaussian have variances ${\sigma_1}^2 <= {\sigma_2}^2$ along minor and major axes $\boldsymbol{u_1}$ and $\boldsymbol{u_2}$ respectively. The corresponding Bingham distribution on the circle has modal axis $\boldsymbol{u_2}$ and a single canonical shape parameter $k_1 = -w = -\frac{1}{2}(\frac{1}{{\sigma_1}^2} - \frac{1}{{\sigma_2}^2})$. Choose a spherical coordinate system indexed by angle θ, aligned with the major and minor axes such that any unit vector $\boldsymbol{x}$ is represented as $\sin\theta \boldsymbol{u_1} + \cos\theta \boldsymbol{u_2}$. The density function maxima then occur at θ values of 0 and π, and minima occur at $\pi/2$ and $3\pi/2$ (Figure 3.2). Within this*

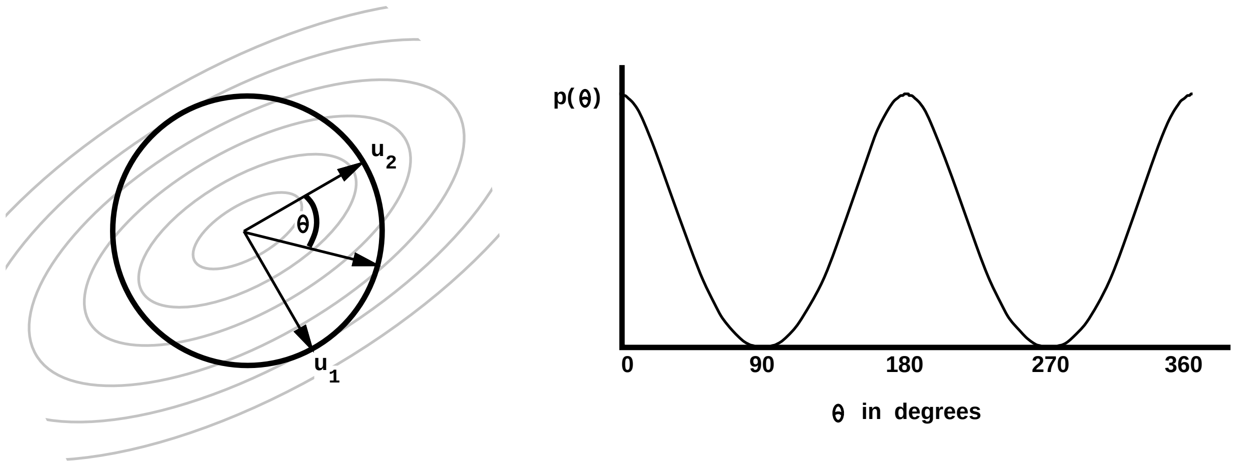


Figure 3.2: Bingham's distribution on the circle. Left shows iso-density contours of a zero-mean bivariate Gaussian superimposed over the unit circle. A Bingham density function is formed by intersecting the Gaussian density with the circle and renormalizing to get total probability one. At right is a plot of the resulting bimodal Bingham density function, coordinatized by circular angle θ from one of the modes.

spherical coordinate system the Bingham density function on the circle is written

$$C \; exp\{\, -w \, \sin^2 \theta \,\} \;\; . \tag{3.2}$$

The constant C is computed by normalizing this expression over the unit circle. Using the trigonometric identity $\cos 2\theta = 1 - 2\sin^2\theta$, we find

$$\begin{aligned} 1/C &= \int_0^{2\pi} exp\{\, -w \, \sin^2 \theta \,\} \, d\theta & (3.3) \\ &= exp\{-\frac{w}{2}\} \int_0^{2\pi} exp\{\, \frac{w}{2} \cos 2\theta \,\} \, d\theta & (3.4) \\ &= exp\{-\frac{w}{2}\} \, 2\pi \, I_0(\, \frac{w}{2} \,) \;\; , & (3.5) \end{aligned}$$

with I_0 being standard notation for the modified Bessel function of the first kind and order 0. The last integral was evaluated by noting that $\int exp\{\frac{w}{2}\cos 2\theta\}$ is the inverse normalization constant for an axial von-Mises distribution with concentration parameter $k = w/2$

([Mardia72], Section 2.6). This derivation shows that Bingham's distribution on the circle is equivalent to the axial von-Mises distribution.

■

Example 3.2 *A rotationally-symmetric version of Bingham's distribution (such as that shown in Figure 3.1b) is obtained by letting* $k_1 = k_2 = -w$, *and implicitly* $k_3 = 0$. *Since the Bingham shape parameters* k_i *are defined only up to an additive constant, an equivalent density function is obtained by letting* $k_i^* = k_i + w$, *so that* $k_1^* = k_2^* = 0$, *and* $k_3^* = w$. *The Bingham density function of equation 3.1 simplifies to the form*

$$\mathcal{B}(\boldsymbol{x};\boldsymbol{U},\boldsymbol{K}) \;=\; C\ exp\{\,\boldsymbol{x}^t\,(w\,\boldsymbol{u_3}\boldsymbol{u_3^t})\,\boldsymbol{x}\,\} \;=\; C\ exp\{\,w\,(\boldsymbol{u_3^t}\boldsymbol{x})^2\,\}\ . \tag{3.6}$$

This probability distribution is more commonly known as the Dimroth-Watson distribution [Mardia72, Upton89]. The value of the density function depends only on the arc distance of a point from the poles $\pm\boldsymbol{u_3}$, *and thus the distribution is symmetric (invariant) under rotations about polar axis* $\boldsymbol{u_3}$. *Iso-density contours are concentric circles on the sphere. The single shape parameter* w, *called the concentration parameter, governs the concentration of probability mass with respect to the poles. When* $w > 0$, *the probability mass is attracted to the poles and the distribution is bimodal. A high concentration value results in sharp, well-defined antipodal peaks, while a low concentration value leads to flatter, more diffuse modal peaks. For* $w < 0$ *the probability mass is repelled from the poles and clusters about the great circle halfway between them. The sharpness of this modal belt is again directly proportional to the absolute value of the concentration parameter. When* $w = 0$ *the distribution reduces to the uniform distribution on the sphere.*

■

3.1.2 Bingham Maximum Likelihood Estimation

Maximum likelihood estimation is commonly used to estimate the parameters of a distribution from a set of sample observations. Maximum likelihood estimation of Bingham parameters from a random sample is performed in the usual way, by setting up the likelihood equations for an independent and identically distributed (iid) sample of size N, then solving for the parameters that maximize the likelihood.

Assume a set of N unit vectors $\boldsymbol{x}_i = \langle x_i, y_i, z_i \rangle$ that are distributed according to a Bingham density. Based on general results for estimating the parameters of a density function in the exponential family [Beran79, Clutton-Brock90], the parameters that maximize the likelihood function must be chosen so that the second moments of the estimated density are equal to the second moments of the sample. The maximum likelihood result obtained in this situation will therefore be equivalent to that found by the method of moments parameter estimation procedure.

A sufficient statistic for the orientation and shape parameters $\boldsymbol{U}$ and $\boldsymbol{K}$ is the sample second moment or *scatter* matrix

$$\boldsymbol{M} = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}\boldsymbol{x}^t) = \frac{1}{N}\begin{bmatrix} \sum {x_i}^2 & \sum x_i y_i & \sum x_i z_i \\ \sum x_i y_i & \sum {y_i}^2 & \sum y_i z_i \\ \sum x_i z_i & \sum y_i z_i & \sum {z_i}^2 \end{bmatrix}. \tag{3.7}$$

Since the scatter matrix is a symmetric real matrix, it can be decomposed into $\boldsymbol{M} = \boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{A}^t$, where $\boldsymbol{A} = [\boldsymbol{a_1}, \boldsymbol{a_2}, \boldsymbol{a_3}]$ is an orthonormal matrix of eigenvectors, and $\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, \lambda_3)$ is a diagonal matrix of corresponding eigenvalues with $\lambda_1 \leq \lambda_2 \leq \lambda_3$ summing up to 1. It can be shown [Bingham74, Mardia72] that the maximum likelihood estimate of the Bingham orientation matrix $\boldsymbol{U}$ is the matrix of eigenvectors $\boldsymbol{A}$. Maximum likelihood estimates of the shape parameters k_1 and k_2 are nontrivial functions of $\mathbf{\Lambda}$ that can be looked up in tables [Mardia77] or computed numerically [Kent87, Upton89].

3.1.3 A Note Regarding Alternative Distributions

We have chosen Bingham's distribution for representing uncertainty on the sphere and projective plane. In order to fully understand the strengths and weaknesses of this choice, this section briefly reviews some of the other possibilities. It is not meant to be a complete review of distributions on the sphere, but rather a look at those distributions that are used most often, and are therefore the most obvious possible alternatives. More details are available in [Mardia72, Upton89, Watson83, Jupp91].

In applied statistics on the sphere, the two most popular distributions in practice are the Fisher (or Langevin) distribution and the Dimroth-Watson distribution [Upton89]. The Fisher distribution is a higher-dimensional generalization of the Von-Mises distribution on the circle. Since it is unimodal, it is automatically disqualified as a representation of uncertainty in projective space where antipodal (bimodal) symmetry is required. An antipodally symmetric variant of the Fisher distribution can be defined; the result is the Dimroth-Watson distribution, which is an antipodally symmetric distribution with circular iso-density contours. The Dimroth-Watson distribution is actually a special case of Bingham's distribution with $k1 = k2$ (see Example 3.2). Since probability values depend only on arc distance from the mode (the point of highest probability mass), independent of rotational angle about the mode, this distribution is easier to use than the full Bingham model. However we shall see cases where accurate description of uncertainty requires the elongated, ellipsoidal iso-density contours available from the more flexible Bingham density.

Commonly used distributions on the sphere, including Bingham's distribution, share the property of remaining invariant in form under rotations of the underlying coordinate system. The set of all instantiations of a density function formed by giving specific values to its parameters is called a *family* of densities. The Bingham family is said to be *closed under rotations* since, although the values of the parameters may be different in the new coordinate system, no new parameters are needed and the density function retains the same basic functional form. This is an important property when reasoning about orientations because it guarantees that the inferences made are invariant to the (arbitrary) choice of the underlying orientation coordinate system.

When reasoning in projective space, the choice of the underlying coordinate system should likewise be arbitrary. Since a change of coordinates from one projective frame to another involves a general projective transformation, it is desirable that the family of density functions used to represent uncertainty in projective space be closed under projective transformations. Unfortunately, the Bingham family is not closed under projective transformations, as will be shown formally in the next section. Transforming a Bingham density function via a projective transformation does not yield a new Bingham density, and the process of making the resulting function Bingham entails an approximation process. Precision *may* be lost because this approximation deteriorates under very distorted coordinate systems, and as a result, the accuracy of the geometric inferences made *may* depend on which coordinate frame is used. This does not turn out to be an overwhelming problem in practice, since there is a natural coordinate frame to use for camera-center projective reasoning, namely the pinhole camera coordinate system in which the homogeneous coordinate vectors of image features are aligned with the orientations of 3D view rays towards corresponding features in the scene. Although the exact geometry of the pinhole coordinate system depends on camera parameters that may not be known exactly, such as focal length and image center, we hypothesize that any reasonable approximation to the true viewing geometry will yield equally valid statistical inferences. In practice, the non-closedness of Bingham's distribution under projective transformation merely means that extremely distorted coordinate systems must be avoided. When a coordinate system with a field of view angle of $2°$ is used when the angle is in fact $30°$, or an aspect ratio of $1000 : 1$ is specified when it is more like $1 : 1$, numerical problems may ensue.

In contrast, the Angular Central Gaussian distribution <u>**is**</u> closed under projective transformations [Tyler87]. The density function can be intuitively described as the marginal distribution of directions of a zero-mean, multivariate Gaussian distribution. That is, the value of the density function at a point on the sphere is the integral of a Gaussian over the ray starting from the center of the Gaussian and having the same direction as that defined by the given point on the sphere. Because the angular central Gaussian family is closed under projective transformations, statistical inferences based on this distribution are truly invariant to the choice of projective coordinate system. It thus seems to be the natural distribution to use for statistical reasoning in projective space. Unfortunately, this distribution is hard to use analytically. Even the simplest estimation tasks, such as maximum likelihood estimation of density parameters from a set of independent, identically distributed samples, must be performed using iterative numerical methods [Tyler87]. Our choice of Bingham's distribution over the angular central Gaussian is a case where computational expediency has been chosen over theoretical rigor.

Two approaches to representing uncertain orientation data in the vision literature, both using the Gaussian distribution, are worthy of mention here. Kanatani performs statistical reasoning tasks over uncertain points on the sphere for several vision applications, many of which are also treated in this thesis [Kanatani92, Kanatani93a, Kanatani93b]. He represents unit vector uncertainty using a 2D Gaussian distribution in a plane tangent to the sphere at the nominal position of the uncertain vector. This approach is asymptotically equivalent to using Bingham's distribution when the amount of uncertainty is very small, and thus the relevant portion of the sphere's surface is well-approximated by a tangent plane. The tangent plane approach will be less useful, however, when errors are large or sample sizes are small. In these cases the probability mass is spread out over larger portions of the sphere's surface, and the curvature of the sphere must be taken into account. For this reason, approaches based on tangent plane Gaussian approximations are always overly optimistic in the sense that the generated confidence regions are too small. This misrepresentation gets worse as variance increases.

Hel-Or and Werman present an interesting application of the singular 3D Gaussian distribution [Hel-Or92]. Their goal is to develop a unified approach to computing 3D camera pose given a set of 3D model points and a set of uncertain measured data points, which could come from either 3D range data, 2D image projections, or combinations of both. They unify the treatment of both 2D and 3D data by converting each 2D image point into a 3D point lying somewhere along a view ray. The total lack of knowledge in the actual position of the point on the ray is represented by setting the variance of the point position along the view ray to infinity. Earlier in this section it was shown that a singular Gaussian distribution having infinite variance along an axis passing through the center of projection can be thought of as a Bingham distribution. Hel-Or and Werman's treatment of 2D projected image data can thus be reinterpreted as representing the uncertainty in the orientation of the view ray towards the unknown 3D point. This is an interesting and perhaps fruitful way of looking at their work. In actuality, however, they do not treat the uncertain entity they produce as an orientation at all, but rather as a 3D location estimate that is *very* uncertain in one direction. The notion of statistical reasoning on the sphere never comes up, and is thus not addressed.

3.2 Transformations of Projective Uncertainty

At the heart of any uncertain reasoning system lie two tasks: transferring uncertain geometric features from one coordinate system to another, and fusing multiple feature observations to infer either new or more accurate geometric descriptions. This section addresses the former; the latter is the topic of Section 3.4.

Determining how geometric uncertainty changes under coordinate transformations is necessary for a variety of reasons. First, multiple measurements of one feature are often made in more than one coordinate frame (due to motion for example). It is required to bring these disparate measurements into a common reference frame so that they may be compared or fused. In addition to transforming the geometric object parameters themselves, any associated description of uncertainty must also be transformed into the new coordinate frame. A closely related task is reparameterization of feature descriptions, such as converting from Cartesian to polar coordinates. Finally, simple aggregations of low-level primitives into higher-level features can also be described as coordinate transformations. For instance, computing the feature description of the line passing through two points is a deterministic transformation amenable to the techniques of this section.

3.2.1 Transformation of Random Variables

When uncertain geometric objects are represented as random variables, changes in uncertainty due to coordinate transformations are reflected by changes in the probability density function. The parameter vectors of a geometric object before and after a change of coordinate frames or a reparameterization are related by the vector function $\boldsymbol{y} = T(\boldsymbol{x})$. If parameter vector $\boldsymbol{x}$ is a vector of random variables representing an uncertain geometric object, then the resulting parameter vector $\boldsymbol{y}$ is also a random vector, with a probability distribution determined both by the distribution of $\boldsymbol{x}$ and by the transformation T. For an aggregation of two geometric objects into a third, the transformation is written as $\boldsymbol{z} = T(\boldsymbol{x}, \boldsymbol{y})$ where vector function T describes how the parameters of $\boldsymbol{x}$ and $\boldsymbol{y}$ combine to produce the new object parameters $\boldsymbol{z}$. Once again, density functions attached to random vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ can be propagated through T to determine the probability density function of random vector $\boldsymbol{z}$.

The process used to compute the probability density function of a function of random variables is called the transformation of variables technique. If the transformation function is bijective (one-to-one and onto) as well as differentiable, computation of the resulting density function is a well-defined process. Consider a random vector $\boldsymbol{x}$ distributed according to $g(\boldsymbol{x})$, and the bijective transformation $\boldsymbol{y} = T(\boldsymbol{x})$. Since T is bijective the inverse transformation T^{-1} is defined, and the probability density function of the random vector $\boldsymbol{y}$ is known to be

$$h(\boldsymbol{y}) = g\left(T^{-1}(\boldsymbol{y})\right) J_{T^{-1}}(\boldsymbol{y}) \tag{3.8}$$

where the Jacobian $J_{T^{-1}}(\boldsymbol{y})$ specifies the change in area of a differential surface patch at point $\boldsymbol{y}$ under transformation T^{-1} (see Appendix C).

Principle 2: Changes in uncertainty due to coordinate transformations will be determined by computing a new probability density function using the transformation of variables technique.

The transformation of variables technique provides a powerful tool for determining how a probability density function changes as a result of a coordinate transformation. Unfortunately, it takes only a few transformations before exact computation of the resulting density function becomes complicated to the point of intractability [Springer64]. When uncertain parameters are Gaussian it is common practice to approximate the density resulting from a nonlinear transformation as another Gaussian by computing just the first and second order moments (mean and covariance) of the transformed density. This uniform representation of uncertainty provides a flexible and powerful framework for uncertain geometric reasoning. It makes good sense to adopt the same approach on the sphere, where Bingham's distribution provides a low-order approximation to any antipodally symmetric density function.

Principle 3: Any transformed density function that is not of Bingham form will be approximated as a Bingham density in order to maintain a uniform representation of uncertainty.

The rest of this section derives an analytic formula describing how a Bingham density function transforms under a projective coordinate transformation on the sphere. Since the resulting probability density functions is not always Bingham, Section 3.3 will discuss ways to approximate a computed density function by a Bingham density.

3.2.2 Projective Coordinate Transformations on the Sphere

The coordinate transformations of fundamental interest to projective geometry are of course the projective transformations. Let $\boldsymbol{A}$ be a nonsingular 3×3 matrix representing a plane-to-plane projective transformation in homogeneous coordinates. Identify each equivalence class of homogeneous coordinates with a line through the origin of R^3. This line intersects the unit sphere S^2 in a pair of antipodal points. Under a projective transformation, each infinite line through the origin is mapped to a new line through origin, which intersects the unit sphere in another pair of antipodal points. The operation of $\boldsymbol{A}$ upon the sphere is said to be the transformation P that maps a pair of antipodal points on the sphere before transformation $\boldsymbol{A}$ to the corresponding pair after transformation $\boldsymbol{A}$. This transformation is

$$\boldsymbol{y} \;=\; P(\boldsymbol{x}) \;=\; \frac{\boldsymbol{A}\boldsymbol{x}}{\|\boldsymbol{A}\boldsymbol{x}\|} \quad . \tag{3.9}$$

Notice that $P(-\boldsymbol{x}) = -P(\boldsymbol{x})$ so that antipodal pairs are indeed mapped to antipodal pairs. Further note that although a projective transformation is linear in homogeneous coordinates, it is not necessarily linear over points on the sphere.

The probability density of $\boldsymbol{y}$ is a function of the density $g(\boldsymbol{x})$ and of the inverse projective transformation matrix $\boldsymbol{B} = \boldsymbol{A}^{-1}$. The formula is derived in Appendix C (example C.11) as

$$h(\boldsymbol{y}) \;=\; g\left(\frac{\boldsymbol{B}\,\boldsymbol{y}}{\|\boldsymbol{B}\,\boldsymbol{y}\|}\right)\frac{|\,det\,\boldsymbol{B}\,|}{\|\boldsymbol{B}\,\boldsymbol{y}\|^3} \quad . \tag{3.10}$$

This equation will now be specialized to the case where g is a Bingham probability density function.

Transformation by Rotations and Reflections

A special case of equation 3.10 occurs when the transformation matrix is orthonormal. Let $\boldsymbol{A} = \boldsymbol{R}$ be an orthonormal matrix, so that $\boldsymbol{R}^t\boldsymbol{R} = I$ and $det\,\boldsymbol{R} = \pm 1$. Matrices of this form can be interpreted as a rotation or reflection about the origin of Euclidean space. Since vector length is preserved, formula 3.10 reduces to

$$h(\boldsymbol{y}) = g(\boldsymbol{R}^t\boldsymbol{y}). \tag{3.11}$$

Assuming $g(x)$ is a Bingham density function with orientation parameters $\boldsymbol{U}$ and shape parameters $\boldsymbol{K}$ (see equation 3.1)

$$g(\boldsymbol{x}) = \mathcal{B}(\boldsymbol{x};\boldsymbol{U},\boldsymbol{K}) = c(\boldsymbol{K})\,exp\{\boldsymbol{x}^t\boldsymbol{U}\boldsymbol{K}\boldsymbol{U}^t\boldsymbol{x}\}.$$

Substituting this into the previous equation yields the density function of $\boldsymbol{y}$

$$\begin{aligned} h(\boldsymbol{y}) &= c(\boldsymbol{K})\,exp\{(\boldsymbol{R}^t\boldsymbol{y})^t\boldsymbol{U}\boldsymbol{K}\boldsymbol{U}^t(\boldsymbol{R}^t\boldsymbol{y})\} \\ &= c(\boldsymbol{K})\,exp\{\boldsymbol{y}^t(\boldsymbol{R}\boldsymbol{U})\boldsymbol{K}(\boldsymbol{R}\boldsymbol{U})^t\boldsymbol{y}\} \\ &= \mathcal{B}(\boldsymbol{y};(\boldsymbol{R}\boldsymbol{U}),\boldsymbol{K}) \end{aligned} \tag{3.12}$$

No approximation is needed, since the result is a Bingham density function with orientation parameters $\boldsymbol{RU}$ and shape parameters $\boldsymbol{K}$, *i.e.* just a rotated or reflected version of the original Bingham density function. Bingham's distribution therefore remains invariant in form and shape under orthonormal projective transformations. This result also holds for scaled orthonormal mappings $k\boldsymbol{R}$.

General Projective Transformations

Let $\boldsymbol{A}$ be a general, nonsingular 3×3 matrix with inverse $\boldsymbol{B}$, and let $\boldsymbol{x}$ be a Bingham variable with orientation parameters $\boldsymbol{U}$ and shape parameters $\boldsymbol{K}$. Inserting the Bingham density function (3.1) into the transformation of variables equation (3.10) yields

$$h(\boldsymbol{y}) \;=\; c(\boldsymbol{K})\, exp\{(\tfrac{\boldsymbol{B}\,\boldsymbol{y}}{\|\boldsymbol{B}\,\boldsymbol{y}\|})^t \boldsymbol{U}\boldsymbol{K}\boldsymbol{U}^t(\tfrac{\boldsymbol{B}\,\boldsymbol{y}}{\|\boldsymbol{B}\,\boldsymbol{y}\|})\}\, \frac{|\,det\,\boldsymbol{B}\,|}{\|\boldsymbol{B}\,\boldsymbol{y}\|^3} \quad . \tag{3.13}$$

No significant simplification can be performed. This density function is clearly not of Bingham form, due to the nonconstant Jacobian and the nonlinear way variable $\boldsymbol{y}$ is involved in the exponential term. Propagating a Bingham density function through a general projective transformation on the sphere will therefore require an approximation step to reduce the resulting probability density function to a corresponding Bingham density.

3.3 Approximation by a Bingham Density

The transformation of variables technique is a powerful tool for deriving analytic descriptions of the probability density function resulting from a coordinate transformation. The resulting density functions can be quite complicated, however. To make the process of uncertainty propagation manageable, any computed probability density function not of Bingham form will always be approximated by a Bingham density. This ensures a uniform representation of uncertainty at all stages of the geometric reasoning process.

It remains to be seen how to actually determine the parameters of the Bingham density function that best approximates a given probability density function on the sphere. In this section three possible approaches are considered. The first is developed by analogy to maximum likelihood estimation of Bingham parameters from a set of samples. The strategy is to replace the sample moments in that procedure by the actual population moments computed analytically from the given density function. The remaining two approximation techniques are based on the observation that Bingham's distribution is formed as a zero-mean Gaussian intersected with the surface of the sphere. This suggests a Tayor-series expansion around the mode of the density function, in an effort to estimate the parameters of the underlying 3D Gaussian. This of course requires the mode to be found first, which will be done using a modification of the Newton-Raphson method of finding extrema of functions. The third approximation technique considered in this section abandons the surface of the sphere completely, relying instead on covariance propagation in the ambient Euclidean space. Rather than being given an analytic description of a density function on the sphere, this approach propagates a singular covariance matrix representing a Bingham density function through a linearized version of the given transformation to determine a new covariance matrix, and thereby a new approximate Bingham density function.

3.3.1 Maximum Likelihood Approximation

A method for estimating the parameters of the Bingham density that best approximates a given density function can be developed by analogy with maximum likelihood estimation of Bingham parameters from a random sample. The idea is to replace the sample statistics in the likelihood equations by the population statistics computed from the given density function.

The second moments of a given density function $f(\boldsymbol{x})$, where $\boldsymbol{x} = (x, y, z)$ is a point on the sphere, are the expected values of the quadratic monomials in x, y and z. The population second moment matrix is therefore written as (c.f. equation 3.7)

$$\boldsymbol{M} \;=\; E(\boldsymbol{x}\boldsymbol{x}^t) \;=\; \begin{bmatrix} E(x^2) & E(xy) & E(xz) \\ E(xy) & E(y^2) & E(yz) \\ E(xz) & E(yz) & E(z^2) \end{bmatrix}. \tag{3.14}$$

Each expected value is found by integration of the monomial times the density function over the surface of the sphere. For example, $E(x^2) = \int x^2 f(\boldsymbol{x})\, dS^2$ using the integral notation of Appendix B.2. Since $x^2 + y^2 + z^2 = 1$, one of the integrals can be avoided by computing $E(z^2) = 1 - E(x^2) - E(y^2)$. As before, to maximize the likelihood equations the Bingham orientation parameters are chosen to be the eigenvectors of the population second moment matrix, and the shape parameters are functions of the eigenvalues.

This modified maximum likelihood method for density approximation is not very useful in practice, however. The main drawback is that the surface integrals for computing population second moments often cannot be solved in closed form. In fact for most examples we have tried, the resulting integrals have had to be evaluated using numerical integration techniques – an approach that is much too slow to be included in a working vision system.

A second problem with this approach is that, occasionally, the quality of the approximations obtained is low, even given accurate computations. Classic maximum likelihood estimation assumes that the samples observed are drawn from the type of distribution whose parameters are being estimated. It is not necessarily designed to give good results when used for density approximation. It is known, for example, that within the exponential family of distributions the approximating distribution obtained using maximum likelihood estimation maximizes Kullback-Liebler divergence, an information-based measure of the difference between two distributions [McCulloch88]. However, it is also known that "goodness-of-fit" measures based on Kullback-Liebler divergence give too much weight to the tails of a distribution [Clutton-Brock90]. That is, the Kullack-Liebler loss function disproportionately rewards approximations that fit the target distribution accurately in places where the probability mass approaches zero.

We propose instead to use mean-squared-error (MSE) as a goodness-of-fit measure, because the MSE criterion rewards approximations that fit best where the probability density is highest. This is an important consideration when doing Bayesian analysis, an approach to statistical inference used in the next section. In most Bayesian work, the mode of the posterior density function is taken as an estimate of the most likely parameter value, and a confidence region for this estimate is constructed based on the shape of the density function near the mode. It is thus desirable that an approximate Bingham density function should fit the true density function best near the mode, where most of the probability mass is.

3.3.2 Taylor Series Expansion about the Mode

When first presented in Section 3.1.1, it was mentioned that the Bingham density function is proportional to a zero-mean, trivariate Gaussian density function intersected with the surface of the sphere. This suggests approaches based on estimating the covariance matrix of a zero-mean, trivariate Gaussian, keeping in mind that the only measurements available are on the sphere.

One standard approach to approximating a given density function by a Gaussian is based on Taylor series expansion of the log density function about the mode of the analytic density [Lindley65]. Taylor series is only a local approximation, but by expanding about the mode, the density function is best approximated where most of the probability mass is. An attractive feature of Taylor series expansion as an approximation technique is that it only requires taking derivatives of the analytic density function, rather than the numeric integration involved with maximum likelihood approximation.

The standard derivation runs as follows [Lindley65]. Let $f(\boldsymbol{x})$ be the probability density function to be approximated. A second order Taylor series expansion of $log\ f$ about $\boldsymbol{x_0}$ yields

$$log\ f(\boldsymbol{x}) \;\approx\; log\ f(\boldsymbol{x}_0) + \boldsymbol{\nabla}^t(\boldsymbol{x}-\boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^t\boldsymbol{\Omega}(\boldsymbol{x}-\boldsymbol{x}_0) \tag{3.15}$$

where $\boldsymbol{\nabla}$ and $\boldsymbol{\Omega}$ are the gradient vector and Hessian matrix of $log\, f$, evaluated at $\boldsymbol{x_0}$. A defining feature of the mode $\hat{\boldsymbol{x}}$ of $log\ f$ is that the gradient is 0 there. Evaluating the Taylor series of $log\, f$ at the mode, and exponentiating to recover an approximation of f thus yields

$$f(\boldsymbol{x}) \;\approx\; C\ exp\left\{\frac{1}{2}(\boldsymbol{x}-\hat{\boldsymbol{x}})^t\boldsymbol{\Omega}(\boldsymbol{x}-\hat{\boldsymbol{x}})\right\} \quad . \tag{3.16}$$

Comparing this equation with that of the Gaussian density function yields an approximate Gaussian mean of $\hat{\boldsymbol{x}}$ and inverse covariance matrix $-\boldsymbol{\Omega}$.

To tailor this derivation to approximate a bimodal, antipodally symmetric probability density function $f(\boldsymbol{x})$ on the sphere with a Bingham density function, let $\hat{\boldsymbol{x}}$ be one of the two antipodal modes of $f(\boldsymbol{x})$. The major axis of the underlying zero-mean, trivariate Gaussian density function will lie along $\hat{\boldsymbol{x}}$. Choose vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ so that the matrix $R \equiv [\boldsymbol{u}_1\,\boldsymbol{u}_2\,\hat{\boldsymbol{x}}]$ is an orthonormal rotation matrix. Vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ thus span the tangent plane to the unit sphere at point $\hat{\boldsymbol{x}}$. Let $\boldsymbol{\Omega}$ be the Hessian matrix of $log\ f$, evaluated at $\hat{\boldsymbol{x}}$, and define a *directional* second derivative of *log f* with respect to vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ to be the value $\boldsymbol{a}^t\boldsymbol{\Omega}\boldsymbol{b}$. The 2×2 matrix

$$\boldsymbol{\Omega}_T = \begin{bmatrix} \boldsymbol{u}_1^t\boldsymbol{\Omega}\boldsymbol{u}_1 & \boldsymbol{u}_1^t\boldsymbol{\Omega}\boldsymbol{u}_2 \\ \boldsymbol{u}_1^t\boldsymbol{\Omega}\boldsymbol{u}_2 & \boldsymbol{u}_2^t\boldsymbol{\Omega}\boldsymbol{u}_2 \end{bmatrix} \tag{3.17}$$

represents the Hessian matrix of second partial derivatives of *log f* restricted to the tangent plane at point $\hat{\boldsymbol{x}}$, and by the above argument $-\boldsymbol{\Omega}_T$ is the inverse covariance of a 2D Gaussian cross-sectional "slice" perpendicular to the major axis of the 3D Gaussian distribution being estimated (Figure 3.3).

All that remains to be estimated is the change in scale of the cross-sectional Gaussian slices along the major axis, i.e. the inverse variance along the major axis. An obvious measure is (-1) times the directional second derivative along that axis. One requirement must be imposed for this to work, however, namely that the variance along the major axis must be maximal (greater than the variance along any other axis). A simple constraint that maintains such consistency, and is fulfilled by all the density functions that arise in this thesis, is that $f(\boldsymbol{x})$, when treated as an unconstrained 3D function, has a single mode at the origin $\boldsymbol{x} = 0$ and decreases monotonically along any ray emanating out from the origin. This constraint can be relaxed to include the case where f has a mode consisting of a single line of points through the origin. The extent to which it can be relaxed still further has not been explored. Assume this consistency constraint is satisfied, and let $\omega = \hat{\boldsymbol{x}}^t\Omega\hat{\boldsymbol{x}}$ be the directional

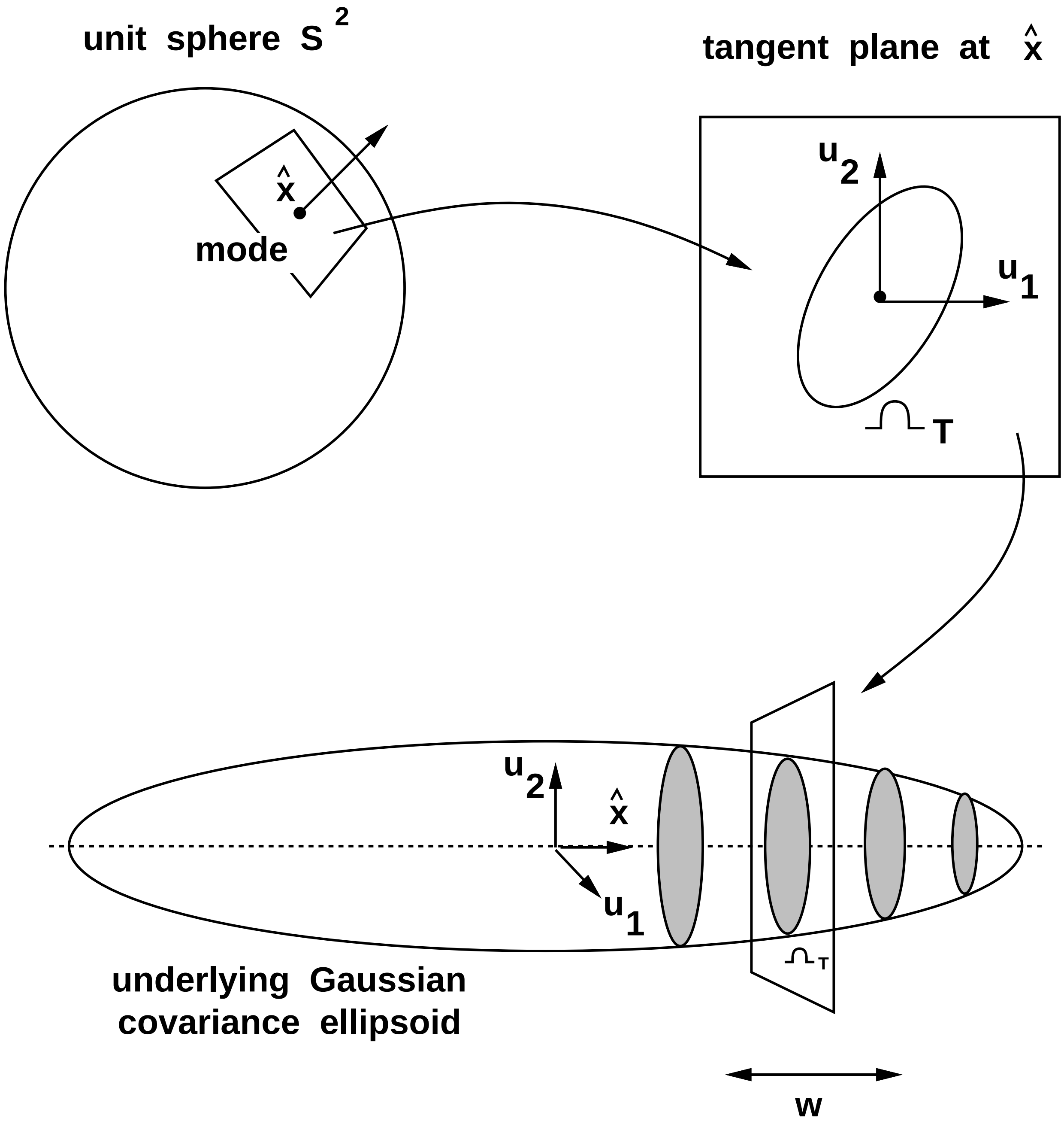


Figure 3.3: Taylor series approximation of a Bingham density. A second-order Taylor series expansion at the mode of the target density function on the sphere determines the parameters of a zero-mean, trivariate Gaussian distribution from which a Bingham density is constructed by normalizing over the surface of the sphere. Refer to the text.

second derivative along the major axis. The Taylor series Bingham approximation can then be written as

$$B(\boldsymbol{x}) \quad = \quad C \; exp\left\{ \boldsymbol{x}^t \; [\, \boldsymbol{u}_1 \; \boldsymbol{u}_2 \; \hat{\boldsymbol{x}} \,] \begin{bmatrix} \frac{1}{2}\Omega_T & & 0 \\ & & 0 \\ 0 & 0 & \frac{\omega}{2} \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1{}^t \\ \boldsymbol{u}_2{}^t \\ \hat{\boldsymbol{x}}^t \end{bmatrix} \boldsymbol{x} \right\} \tag{3.18}$$

This derivation assumes that the mode $\hat{\boldsymbol{x}}$ of density function f on the sphere is known. To locate the mode of a density function on the sphere, a modification of the Newton-Raphson method for finding the extrema of functions was developed. This modification constrains the search for extrema to the surface of the sphere. Newton-Raphson is a particularly appropriate algorithm for this application since the directional second derivatives necessary for Equation 3.18 are generated as a byproduct.

3.3.3 Singular Covariance Propagation

The final Bingham approximation method considered in this section is based on standard covariance propagation techniques. In Section 3.1.1 it was shown that Bingham's distribution on S^2 can be represented by a unique, singular Gaussian density function centered around the origin of the ambient space R^3. Let the inverse covariance matrix of this Gaussian be $\boldsymbol{\Gamma}$. Matrix $\boldsymbol{\Gamma}$ is always symmetric, and it is also singular (it has determinant 0) since one of its eigenvalues is 0. In terms of covariance this means that the Gaussian density function has infinite variance along at least one axis. Given the orientation matrix $\boldsymbol{U}$ and shape parameters k_1 and k_2 of a Bingham density function, the inverse covariance matrix of the unique singular Gaussian representing it can be reconstructed

$$\boldsymbol{\Gamma} \;=\; -2\,\boldsymbol{U} \begin{bmatrix} k_1 & & \\ & k_2 & \\ & & 0 \end{bmatrix} \boldsymbol{U}^t \tag{3.19}$$

Conversely, given a singular, symmetric matrix $\boldsymbol{\Gamma}$, the unique Bingham density function corresponding to it is written

$$B(\boldsymbol{x}) \;=\; C\ exp\left\{-1/2\ \boldsymbol{x}^t\,\boldsymbol{\Gamma}\,\boldsymbol{x}\right\} \tag{3.20}$$

Let transformation $\boldsymbol{y} = \boldsymbol{T}(\boldsymbol{x})$ be a nonlinear coordinate transformation on the sphere with inverse $\boldsymbol{T}^{-1}$, and let $\boldsymbol{\Gamma}$ be a singular, symmetric matrix representing a Bingham density function on the sphere. The Jacobian matrix $\boldsymbol{J}_{T^{-1}}$ of the inverse transformation specifies the linear transformation of R^3 that best approximates the nonlinear inverse transformation $\boldsymbol{T}^{-1}$ at the point at which the Jacobian is evaluated. An obvious choice is the point $\boldsymbol{T}(\hat{\boldsymbol{x}})$, the image of the mode of the original Bingham density. A linear transformation of R^3 maps a zero-mean Gaussian with inverse covariance matrix $\boldsymbol{\Gamma}$ to another zero-mean Gaussian having inverse covariance matrix

$$\boldsymbol{\Gamma}' \;=\; \boldsymbol{J}^t\,\boldsymbol{\Gamma}\,\boldsymbol{J}\ . \tag{3.21}$$

This new matrix is guaranteed to be symmetric. It is also singular, even if $\boldsymbol{J}$ is nonsingular, since

$$det\,\boldsymbol{\Gamma}' \;=\; (det\,\boldsymbol{J}^t)\,(det\,\boldsymbol{\Gamma})\,(det\,\boldsymbol{J})$$

and $\boldsymbol{\Gamma}$ has determinant 0. The unique Bingham density function associated with $\boldsymbol{\Gamma}'$ will be chosen as an approximation of the density obtained by propagating a Bingham density through the linearized version of the nonlinear transformation $\boldsymbol{T}$.

3.3.4 Comparing Approximations

Each of the three approximations considered here has good features and bad. The maximum likelihood approach to computing the parameters of the best-fitting Bingham approximation to an analytic density function has been theoretically analyzed in the statistical literature where it is shown that the approximations produced minimize Kullback-Liebler divergence, an information-based measure of the discrepancy between two distributions. Unfortunately, the integrals required for determining the population second moments are rarely solvable in closed-form, and evaluating them numerically incurs a large computational expense. Furthermore, the Kullback-Liebler criterion is known to favor approximations that fit well at the "tails" of a distribution, but not necessarily at the mode, where most of the probability mass is.

In contrast, an approach based on second-order Taylor series expansion of the log density at the mode of the analytic density function produces Bingham approximations that fit well near the mode. The mode of the approximating Bingham is located exactly at the mode of the analytic density. The Taylor series is computed entirely from derivatives of the analytic density function, rather than integrals, and thus the computational expense is considerable lower. The method does require predetermination of the mode, however. An iterative, Newton-Raphson-style algorithm has been adapted for use in finding the extrema of functions on the sphere. The Newton-Raphson algorithm has attractive convergence properties when started out close to the goal, but a good initial estimate of the mode is necessary to use this method efficiently.

Propagation of Gaussian covariance matrices through linear approximations of nonlinear functions is a powerful tool that underlies most current uncertainty analysis in computer vision. By noting the one-to-one and onto correspondence between zero-mean, singular Gaussian densities and Bingham density functions, an approximate Bingham propagation algorithm has been developed based on propagating the inverse covariance matrices of singular Gaussians through a local, linear approximation to a given transformation. In general, the mode of the approximating distribution will not necessarily coincide with the mode of the target density function. Yet, because the approximations are derived in *closed-form*, this is an extremely useful method for performing a first-order error analysis of the behavior of a Bingham density function with respect to various transformations on the sphere.

Figure 3.4 shows a comparison of these three different approximating strategies in terms of mean-squared error (MSE) for several different projective transformations of a Bingham density function on the circle (example 3.1). The original 2D Bingham density had modal axis

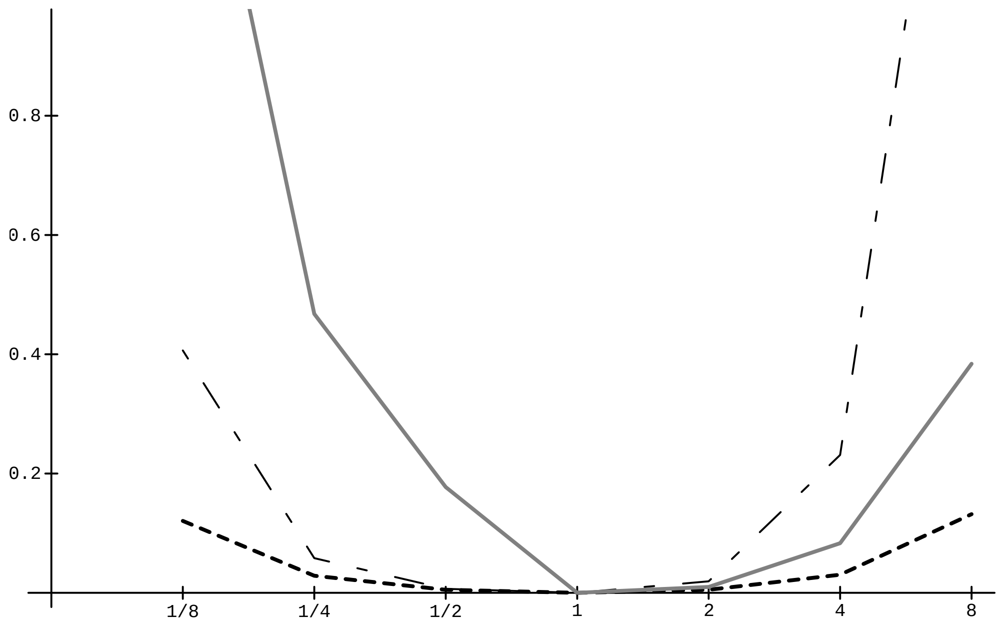


Figure 3.4: Comparison of approximating distributions in terms MSE. Refer to text for details. MLE = maximum likelihood, TAY = taylor series expansion, COV = singular covariance propagation.

$(x, y) = (\cos\frac{\pi}{6}, \sin\frac{\pi}{6})$ and concentration parameter $w = 5$. The projective transformations applied to it were of the form

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad .$$

Figure 3.4 shows the MSE values of each approximation, for values of S ranging from 1/8 to 8. All approximations have zero MSE for $S = 1$, since that indexes the identity transformation. As seen from the figure, approximation using singular covariance propagation consistently outperforms the other estimators in terms of MSE. Combined with its simplicity, singular covariance propagation is the most attractive of the three alternatives considered here.

3.4 Statistical Inference In Projective Space

The essence of geometric inference is the discovery of spatial relationships between objects. When geometric objects are represented as parameter vectors, spatial relationships between objects can be written as algebraic equations relating their parameters. Geometric reasoning over parameterized objects can therefore be given a firm algebraic foundation.

For projective spaces, the fundamental primitives are points, lines and planes represented by homogeneous coordinate parameter vectors, and the relevant spatial relationships are the incidence relations, which remain invariant under projective transformations. In the projective plane the relevant objects are points and lines, and there are two types of incidence relation (see Appendix A.3). Rank I incidence relations hold between objects of the same kind. Two points or two lines are said to be incident iff their homogeneous coordinate vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ are related by the equation $\boldsymbol{p} = k\boldsymbol{q}$. If two geometric elements of the same kind are incident, they are in fact the same object, in the sense that their parameter vectors lie in the same equivalence class of homogeneous coordinates. Rank II incidence relations hold between primitive elements of different kind. A point and a line are incident iff their homogeneous coordinate vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ are related algebraically by $\boldsymbol{p}^t\boldsymbol{q} = 0$. Multiple points incident to a single line are said to be *colinear*. Multiple lines incident to a single point form a line *pencil*.

When the projective plane is represented by the quotient space $S^2/\sim$, homogeneous coordinate vectors are represented by antipodal pairs of points on the unit sphere. Homogeneous coordinate vectors related by Rank I incidence are represented by the same antipodal pair of points on the sphere. Vectors related by Rank II incidence are represented by points separated on the surface of the sphere by an arc distance of $\pi/2$. Multiple points incident to a single line are represented by a great circle of points whose polar axis is aligned with the antipodal pair of points representing the line. Similarly for a set of lines incident to a single point.

These geometric definitions are not quite so crisp when geometric primitives are allowed to be uncertain. Multiple measurements of the same point or line may no longer exactly coincide on the sphere. Points that should lie on a great circle may no longer do so after being perturbed by measurement errors. When dealing with uncertain geometric objects it is best to make as many redundant measurements as possible, and then combine these observations into more accurate feature descriptions. This task is often called *data fusion* in the computer vision literature. How to fuse multiple observations related by projective incidence relations is the topic of this section. A general Bayesian mechanism for evidence combination on the sphere is presented in Section 3.4.1. The problem of combining uncertain projective object descriptions, using known incidence relations to estimate the parameters of new uncertain objects, is deferred until Section 3.4.2.

3.4.1 Bayesian Parameter Estimation

Data fusion over a space of uncertain, parameterized objects can be formalized as a parameter estimation problem. The task is to estimate the value of an unknown parameter vector $\boldsymbol{\theta}$ from a set of observed parameter vectors $\boldsymbol{x} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x}_m\}$. Ideally $\boldsymbol{x} = \boldsymbol{\alpha} = \{\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \ldots, \boldsymbol{\alpha}_m\}$, a set of unknown true values related to vector $\boldsymbol{\theta}$ through a set of implicit, possibly nonlinear, constraint equations

$$\boldsymbol{f}(\boldsymbol{\alpha}\,, \boldsymbol{\theta}) \;=\; \mathbf{0} \quad . \tag{3.22}$$

Due to noise, however, each observation $\boldsymbol{x}_i$ is corrupted by a probability density function with known parameters $\boldsymbol{M}_i$. Thus $\boldsymbol{x}_i$ is distributed according to a probability density function g that depends on $\boldsymbol{\alpha}_i$ and $\boldsymbol{M}_i$

$$p(\boldsymbol{x}_i) \;=\; \boldsymbol{g}(\, \boldsymbol{x}_i,\, \boldsymbol{\alpha}_i\,,\, \boldsymbol{M}_i\,) \quad .$$

Each observation $\boldsymbol{x}_i$ is either a direct measurement of the parameter vector $\boldsymbol{\theta}$, in which case $\boldsymbol{\alpha}_i = \boldsymbol{\theta}$, or is related indirectly through the constraint equations 3.22. In the latter case, the noise-free observables $\boldsymbol{\alpha}$ are called *nuisance parameters,* because estimating their value is not the primary goal of the exercise, but is nevertheless a necessary intermediate step for determining $\boldsymbol{\theta}$.

What has just been sketched is an implicit, nonlinear parameter estimation problem. This general form has widespread applications in the sciences [Britt73, Bates88, Seber89] and computer vision [Ayache91]. If equation (3.22) can be rewritten so that $\boldsymbol{\alpha}$ is expressed explicitly as a function of $\boldsymbol{\theta}$, the problem reduces in form to an explicit, nonlinear parameter estimation task [Britt73], which is sometimes called errors-in-variables nonlinear regression [Reilly81, Schwetlick85]. Although on the surface they appear different, solution techniques for implicit and explicit nonlinear models are essentially equivalent [Dolby76, Benichou89].

A notable feature of current geometric reasoning systems in vision and robotics is the cyclic nature of the processing involved [Ayache91, Matthies87a]. This cyclic style of processing lends itself well to sequential (*recursive*) parameter estimation techniques. When uncertainty in geometric features is represented using probability density functions, the natural choice for a sequential inference engine is Bayesian parameter estimation.

The heart of Bayesian analysis is of course Bayes rule, which states that after the observations have been made, the probability of any given parameter value being correct is proportional to the probability of seeing the observations given that value times the prior probability that the value is correct. Assuming there are no nuisance parameters $\boldsymbol{\alpha}_i$ for the moment, Bayes rule is written[2]

$$p(\boldsymbol{\theta}|\boldsymbol{x}) \quad \propto \quad p(\boldsymbol{x}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}) \ . \tag{3.23}$$

The unspecified constant of proportionality is the value necessary to make the resulting density integrate to one, namely the reciprocal of $\int p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$. Note that this is a constant that does not depend on the value of $\boldsymbol{\theta}$.

Bayes rule describes how a probability density function is modified by new observations to produce an updated density function. The three terms in Bayes rule are the prior density $p(\boldsymbol{\theta})$, the likelihood function $p(\boldsymbol{x}|\boldsymbol{\theta})$, and the posterior density $p(\boldsymbol{\theta}|\boldsymbol{x})$. The prior density describes the state of knowledge about the parameters $\boldsymbol{\theta}$ before any observations have been seen. The joint probability of seeing the observations given a particular value of $\boldsymbol{\theta}$, when interpreted as a function of $\boldsymbol{\theta}$, is called the *likelihood function.* The likelihood function encapsulates all the knowledge gained from the sample observations. These two terms combine to

[2]Technically, each term in this equation should also be conditioned on the given noise parameters $M = \{M_1, M_2, \ldots, M_m\}$. This fact has been omitted for notational clarity.

produce a function of $\boldsymbol{\theta}$ that is proportional to the posterior density. The posterior probability density function describes the updated state of knowledge about $\boldsymbol{\theta}$ after the observations $\boldsymbol{x}$ have been taken into account.

The Prior Density

The inclusion of the prior distribution has led to charges that Bayesian analysis is less objective than classical statistical methods such as maximum likelihood [Efron86]. We disagree, for several reasons. First, choosing a prior distribution based on experience is no more subjective than choosing the form of the underlying distribution for maximum likelihood estimation, or in choosing which features of a problem to model and which to ignore. Secondly, for recursive estimation formulations where uncertain observations are aggregated over time, the prior density is the natural place to accumulate partial results. Finally, if no convincing prior information is available, an uninformative prior density can be used, that is, a density function that gives no overwhelming amount of weight to any particular hypothesis. In that case the posterior density is dominated by the likelihood function, so that maximizing the posterior is equivalent to maximizing the likelihood. It is well known that for linear parameter estimation problems using multivariate Gaussian samples, Bayesian analysis using an uninformative prior yields the same results as classical maximum likelihood and least-square techniques [Gelb86].

Ironically, the use of uninformative priors that brings Bayesian estimation into accordance with classical estimation methods leads to another criticism. In order to accurately reflect the absence of information over an infinite domain, an *improper prior* that is not a well-defined probability density function must be used. For instance, there is no uniform distribution over the whole real number line, since there is no constant value C such that $\int 1/C \, dx = 1$. This is not a problem in the application considered here, however, since the projective parameter spaces are compact and therefore a proper uniform prior exists. Furthermore, since the uniform density function on the sphere can be written in the form $exp\, C$ for some constant C (see Appendix B.2), the uniform prior on the sphere is actually a Bingham density with shape parameters $k1 = k2 = k3 = 0$.

Principle 4: Uncertain parameter values will be estimated using Bayesian estimation on the sphere. When no convincing prior information is available a proper uninformative prior density will be used, specifically the uniform distribution on the sphere.

The Likelihood Function

The likelihood function relates any value of the unknown parameter vector $\boldsymbol{\theta}$ to the probability of seeing the observations $\boldsymbol{x}$. In the absence of overwhelming prior evidence, Bayes rule states that parameter values that maximize the probability of having seen the available observations are more likely to be the correct values. Each observation may be treated as independent, so that

$$p(\boldsymbol{x}|\boldsymbol{\theta}) \;=\; \prod_{i=1}^{m} p(\boldsymbol{x}_i|\boldsymbol{\theta}) \;\; . \tag{3.24}$$

It is assumed that sample vector $\boldsymbol{x}_i$ is generated by a bipolar Bingham noise process with modal axis aligned with $\boldsymbol{\theta}$. The probability of observing an individual $\boldsymbol{x}_i$ given $\boldsymbol{\theta}$ is therefore

$$p(\boldsymbol{x}_i|\boldsymbol{\theta}) \;=\; g(\boldsymbol{x}_i, \boldsymbol{\theta}, \boldsymbol{M}_i) \;=\; C \; exp\{\, \boldsymbol{x}_i^t \; \boldsymbol{M}_i(\boldsymbol{\theta}) \; \boldsymbol{x}_i \,\} \;\; , \tag{3.25}$$

where the Bingham parameter matrix M_i has been written as a function of θ to make that relationship explicit. It needs to be stressed that this is *not* an additive noise model; samples are not generated by taking the true parameter value $\boldsymbol{\theta}$ and adding a noise vector. Since different Bingham processes are allowed for each measurement error, it is necessary either to know *a priori* each Bingham parameter matrix $\boldsymbol{M}_i$, or to have several replications of each observation [Reilly81]. Since each $\boldsymbol{x}_i$ comes from a previous Bayesian inference or from a transformation of random variables, it already has an uncertainty estimate associated with it, but represented as a Bingham density function with modal axis aligned with $\boldsymbol{x}_i$.

An analogous situation arises in engineering applications, where errors are assumed to be Gaussian. There, the variance of the Gaussian process that generated $\boldsymbol{x}_i$ is estimated by the variance of the Gaussian process that would generate observations if $\boldsymbol{x}_i$ were the true value [Bevington69, Press90]. The former is the variance of an unknown Gaussian density centered around $\boldsymbol{\theta}$, while the latter variance is taken from a Gaussian density centered about $\boldsymbol{x}_i$, computed from Monte Carlo-like perturbations of previous observed values or via covariance propagation. For the computed variance in $\boldsymbol{x}_i$ to be a relevant indicator of the measurement noise incurred in observing $\boldsymbol{\theta}$, either the observed value $\boldsymbol{x}_i$ must indeed be near the true value $\boldsymbol{\theta}$, or the effects of random errors upon experimentally measured values must vary slowly with respect to the position of $\boldsymbol{\theta}$. Given a possible true value $\boldsymbol{\theta_0}$, the Gaussian density function centered about $\boldsymbol{x}_i$ is transformed via translation into a Gaussian with modal value $\boldsymbol{\theta_0}$, which is then used to compute the likelihood of observing $\boldsymbol{x}_i$ given $\boldsymbol{\theta_0}$.

On the sphere, where Bingham's distribution is an analogue of the Gaussian, the analogue of translation is *parallel transport.* Parallel transport (sometimes called parallel displacement or parallel translation) is a general mechanism for "translating" vectors on a manifold [Boothby86]. The general idea is that a vector $\boldsymbol{v}$ at point $\boldsymbol{a}$ is transported to point $\boldsymbol{b}$ by sliding it along a geodesic curve from $\boldsymbol{a}$ to $\boldsymbol{b}$ while maintaining a constant angle between the moving vector $\boldsymbol{v}$ and the local tangent vector to the curve at every intermediate point (see Figure 3.5). On the sphere, parallel transport from point $\boldsymbol{a}$ to point $\boldsymbol{b}$ is performed by a rotation about axis $\boldsymbol{a} \times \boldsymbol{b}$ through angle $ArcCos(\boldsymbol{a}^t\boldsymbol{b})$. This is a *minimal rotation*, in the sense that the angle of rotation is smaller than any other (axis,angle) combination that carries point $\boldsymbol{a}$ to point $\boldsymbol{b}$.

On the sphere, an estimate of the uncertainty in the sample value $\boldsymbol{x}_i$ is provided by a Bingham density function with modal axis aligned with $\boldsymbol{x}_i$. Let the known Bingham parameter matrix for this density function be the covariance matrix $\boldsymbol{M}_i(\boldsymbol{x}_i)$. It is desired to transport this density to have an arbitrary modal axis $\boldsymbol{\theta}$ so that the likelihood equation 3.25 can be computed for each possible value of $\boldsymbol{\theta}$. Let $\boldsymbol{R}(\boldsymbol{x}_i, \boldsymbol{\theta})$ be the minimal rotation that

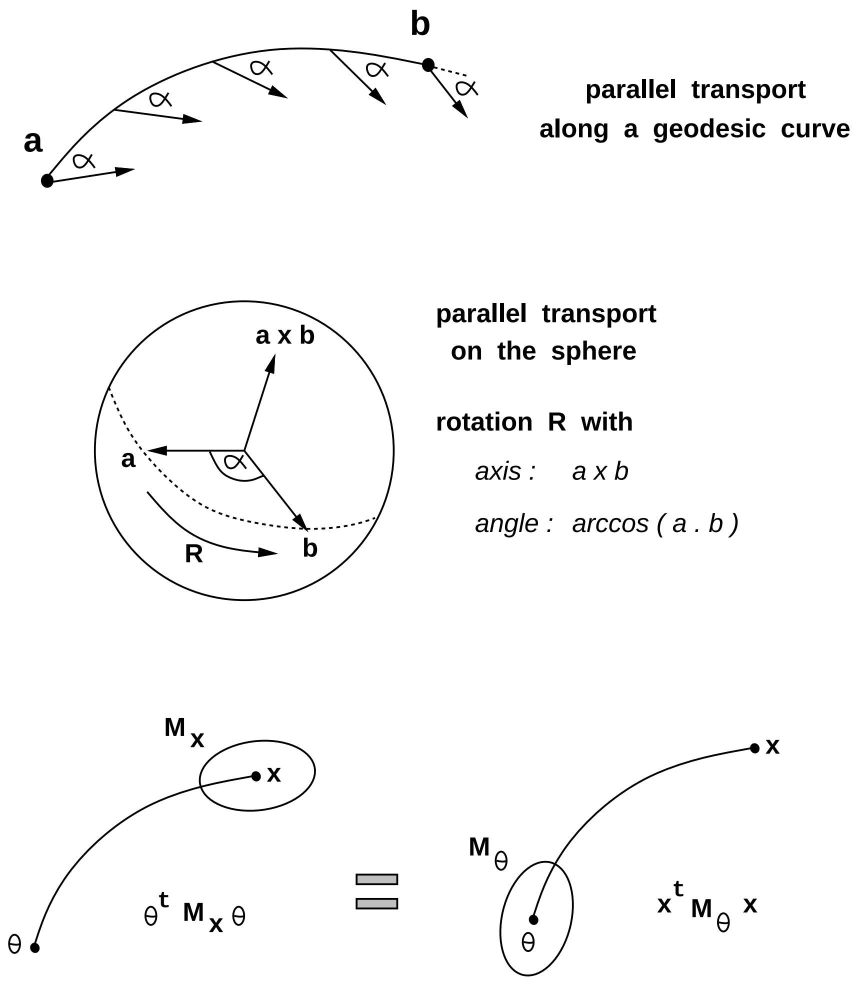


Figure 3.5: Parallel transport on the sphere. Parallel transport provides a general mechanism for "translating" vectors on the surface of the sphere. Transport from point $\boldsymbol{a}$ to $\boldsymbol{b}$ is achieve by performing the minimal rotation taking $\boldsymbol{a}$ to point $\boldsymbol{b}$. The likelihood value computed by performing parallel transport of a Bingham density function from $\boldsymbol{x}_i$ to $\boldsymbol{\theta}$, then evaluating the probability at point $\boldsymbol{x}_i$, is the same as the value returned by the original density function centered around $\boldsymbol{x}_i$, when evaluated at point $\boldsymbol{\theta}$.

transports axis $\boldsymbol{x}_i$ to axis $\boldsymbol{\theta}$. In Section 3.2.2 it was shown that under a coordinate rotation $\boldsymbol{R}$, a Bingham density function with parameter matrix $\boldsymbol{M}$ maps to new Bingham density with parameter matrix $\boldsymbol{R}\boldsymbol{M}\boldsymbol{R}^t$. In the case under consideration, the result is the parameter matrix $\boldsymbol{M}_i(\boldsymbol{\theta})$ of a new Bingham density function with modal axis aligned with $\boldsymbol{\theta}$, just as required to compute the likelihood function. The likelihood equation (3.25) becomes

$$p(\boldsymbol{x}_i|\boldsymbol{\theta}) \; = \; C \; exp\{ \, \boldsymbol{x}_i^t \; \boldsymbol{R}(\boldsymbol{x}_i, \boldsymbol{\theta}) \; \boldsymbol{M}_i(\boldsymbol{x}_i) \; \boldsymbol{R}^t(\boldsymbol{x}_i, \boldsymbol{\theta}) \; \boldsymbol{x}_i \, \} \; . \tag{3.26}$$

The ellipsoidal symmetry of Bingham's distribution allows an important simplification to be made to equation (3.26). In particular,

$$\boldsymbol{x}_i^t \; \boldsymbol{R}(\boldsymbol{x}_i, \boldsymbol{\theta}) \; \boldsymbol{M}_i(\boldsymbol{x}_i) \; \boldsymbol{R}^t(\boldsymbol{x}_i, \boldsymbol{\theta}) \; \boldsymbol{x}_i \quad = \quad \boldsymbol{\theta}^t \; \boldsymbol{M}_i(\boldsymbol{x}_i) \; \boldsymbol{\theta} \tag{3.27}$$

so that

$$p(\boldsymbol{x}_i|\boldsymbol{\theta}) \; = \; C \; exp\{ \, \boldsymbol{\theta}^t \; \boldsymbol{M}_i(\boldsymbol{x}_i) \; \boldsymbol{\theta} \, \} \; . \tag{3.28}$$

In other words, the likelihood value computed by performing parallel transport of a Bingham density function from $\boldsymbol{x}_i$ to $\boldsymbol{\theta}$, then evaluating the probability at point $\boldsymbol{x}_i$, is the same as the value returned by the original density function centered around $\boldsymbol{x}_i$, when evaluated at point $\boldsymbol{\theta}$. This observation is fundamental to the closed-form solution for inferring rank I incidence relations reported in Section 3.4.2.

The Posterior Density

The difference between Bayesian vs. classical approaches to uncertain reasoning is that all Bayesian inferences are based on the posterior density function. It is nevertheless helpful to summarize the information contained in the density function by specifying the point most likely to be the true value, and regions that contain the true value with some degree of confidence. A common point estimate for Bayesian analysis is the mode of the posterior density function, know as the maximum *a posteriori* (MAP) estimator. There are many ways to construct confidence regions containing a specified amount of probability mass. The most intuitive is the highest posterior density (HPD) region. The choice of this region is based on the principle that the probability density of any point inside the region is at least as large as the density at any point outside the region. The boundary of the region obviously follows an iso-density contour of the posterior density function. It can also be shown that the region constructed occupies the smallest possible volume of any region containing the specified amount of probability mass [Box73].

When both the likelihood function and the prior density function are Bingham, the resulting posterior density is also Bingham. In Bayesian terminology the Bingham density function is *conjugate* or *closed* with respect to Bayes rule [Zacks81]. However, the likelihood function need not always be Bingham, and thus the posterior density may have a different form. When this non-Bingham function is used as a prior density in new inferences, the result can be a cascade of increasingly more complicated distributional forms. This situation is similar to that described under transformation of variables, and the solution to this potential problem is the same.

Principle 5: To keep the propagation of uncertainty through Bayes rule manageable, any posterior density not of Bingham form will be approximated by a Bingham density

For a bipolar Bingham posterior density function, the MAP estimator will obviously coincide with the two antipodal modes. An HPD region constructed from a bipolar Bingham posterior will consist of two disjoint regions, bounded by Bingham iso-density contours containing each of the two antipodal modes. These contours are the space-curves formed by the intersection of a zero-mean Gaussian iso-density ellipsoid with the surface of the sphere.

Nuisance Parameters

When the observations $\boldsymbol{x}$ are not direct measurements of the unknown vector $\boldsymbol{\theta}$, but rather of parameters $\boldsymbol{\alpha}$ related indirectly to $\boldsymbol{\theta}$ via the constraint equation (3.22), the nuisance parameters $\boldsymbol{\alpha}$ become additional unknowns that must also be estimated. Bayes rule is modified to incorporate the extra unknowns as

$$p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{x}) \;\; \propto \;\; p(\boldsymbol{x} | \boldsymbol{\theta}, \boldsymbol{\alpha}) \, p(\boldsymbol{\theta}, \boldsymbol{\alpha}) \;\; . \tag{3.29}$$

The posterior is now a function of variables $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$. To obtain a posterior density that is a function of $\boldsymbol{\theta}$ only, the nuisance parameters are integrated out to reveal the marginal posterior density

$$p(\boldsymbol{\theta} | \boldsymbol{x}) \;\; = \;\; \int p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{x}) \, d\boldsymbol{\alpha} \;\; . \tag{3.30}$$

This topic is continued in the next section, under the solution for inferring rank II incidence relations.

3.4.2 Bayesian Estimation of Incidence Relations

The section provides solutions to the problem of inferring accurate homogeneous coordinate descriptions of points and lines in the projective plane by combining multiple, noisy observations related by known incidence relations. The parameter space is the projective plane, represented via normalized homogeneous coordinate vectors on the manifold $S^2/\sim$.

All observations and unknowns are unit vectors; two unit vectors differing only in sign are considered equivalent.

Two general problems are considered. The first is data fusion of points or lines under Rank I incidence relations. This problem involves combining multiple, noisy observations of the same entity to produce a more accurate homogeneous coordinate description. The input is a (bipolar) cluster of observed points on the sphere, along with a Bingham parameter matrix describing the measurement uncertainty in each point. The task is to estimate the single point that best characterizes the cluster, along with a Bingham parameter matrix summarizing its associated uncertainty.

The second problem is data fusion of points and lines under Rank II incidence relations. Multiple, colinear points are fused into a single line description, or multiple, converging lines are fused into a single point describing their intersection. The input is a set of points on the sphere that are clustered about a great circle, with a Bingham parameter matrix quantifying the measurement uncertainty in each point. The task is to estimate the polar axis of the great circle, and to summarize its uncertainty with a Bingham parameter matrix. As an intermediate step, estimates of the true positions of the observables are recovered – all lying exactly on a great circle.

For both problems, a prior estimate of the location and uncertainty of the parameter vector being estimated can be specified by a Bingham prior density function. If no prior estimate is given, the uniform density function on the sphere is used.

Inferring Rank I Incidence

Let $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}$ be a bipolar cluster of observed unit vectors on the sphere. The goal is to estimate the vector $\boldsymbol{\theta}$ that best describes the axis of this cluster. A prior estimate of $\boldsymbol{\theta}$ is given by the prior Bingham density function

$$p(\boldsymbol{\theta}) \;=\; C \; exp\{ \boldsymbol{\theta}^t \; \boldsymbol{M}_0 \; \boldsymbol{\theta} \} \; . \tag{3.31}$$

Associated with each observation $\boldsymbol{x}_i$ is an estimate of its measurement uncertainty. This uncertainty is represented by a bipolar Bingham probability density function with modal axis aligned with $\boldsymbol{x}_i$. Let the known Bingham parameter matrix for $\boldsymbol{x}_i$ be $\boldsymbol{M}_i$. By (3.28) the log likelihood for $\boldsymbol{x}_i$ is, up to constant term C,

$$log\, p(\boldsymbol{x}_i|\boldsymbol{\theta}) \;=\; C \;+\; \boldsymbol{\theta}^t \; \boldsymbol{M}_i \; \boldsymbol{\theta} \;\; , \tag{3.32}$$

and the log likelihood of all observations jointly is

$$\begin{aligned} log\, p(\boldsymbol{x}|\boldsymbol{\theta}) \quad &= \quad log \prod_{i=1}^{m} p(\boldsymbol{x}_i|\boldsymbol{\theta}) \;\; = \;\; \sum_{i=1}^{m} log\, p(\boldsymbol{x}_i|\boldsymbol{\theta}) \\ &= \quad C \;+\; \sum_{i=1}^{m} \boldsymbol{\theta}^t \; \boldsymbol{M}_i \; \boldsymbol{\theta} \end{aligned} \tag{3.33}$$

Combining the prior density (3.31) with the log likelihood (3.33) results in the log posterior density

$$\begin{aligned} log\, p(\boldsymbol{\theta}|\boldsymbol{x}) \quad &= \quad C \;+\; \boldsymbol{\theta}^t \; \boldsymbol{M}_0 \; \boldsymbol{\theta} \;+\; \sum_{i=1}^{m} \boldsymbol{\theta}^t \; \boldsymbol{M}_i \; \boldsymbol{\theta} \\ &= \quad C \;+\; \boldsymbol{\theta}^t \left(\sum_{i=0}^{m} \boldsymbol{M}_i \right) \boldsymbol{\theta} \end{aligned} \tag{3.34}$$

Evidently, the posterior probability density function for $\boldsymbol{\theta}$ is also a Bingham density function, with parameter matrix $\boldsymbol{M} = \sum_{i=0}^{m} \boldsymbol{M}_i$. The MAP estimate for the axis $\boldsymbol{\theta}$ of the cluster of observations is the unit vector $\hat{\boldsymbol{\theta}}$ that maximizes $\hat{\boldsymbol{\theta}}^t \boldsymbol{M} \hat{\boldsymbol{\theta}}$. This is found computationally as the eigenvector associated with the largest eigenvalue of $\boldsymbol{M}$. Boundaries of the HPD confidence regions coincide with the iso-density contours of the Bingham density function.

Example 3.3 *Let the uncertainty in each observation $\boldsymbol{x}_i$ be represented by a rotationally symmetric Bingham density function with modal axis $\boldsymbol{x}_i$ and concentration parameter w_i (Example 3.2). The log likelihood for each observation is then*

$$log\, p(\boldsymbol{x}_i|\boldsymbol{\theta}) \;=\; C \;+\; \boldsymbol{\theta}^t \, (w_i \, \boldsymbol{x}_i \boldsymbol{x}_i^t \,) \, \boldsymbol{\theta} \;\; . \tag{3.35}$$

Assuming a uniform prior density, the log posterior is

$$log\, p(\boldsymbol{\theta}|\boldsymbol{x}) \;\;=\;\; C \;+\; \boldsymbol{\theta}^t \; \left(\sum_{i=1}^{m} w_i \, \boldsymbol{x}_i \boldsymbol{x}_i^t \right) \; \boldsymbol{\theta} \;\; . \tag{3.36}$$

The MAP estimator for the bipolar axis of the cluster of observations is the eigenvector associated with the largest eigenvalue of $(\sum w_i \, \boldsymbol{x}_i \boldsymbol{x}_i^t \,)$. This is also recognized to be the weighted, least-squares solution; that is, the MAP estimator selects the line through the center of the sphere that minimizes the sum of squares of the weighted orthogonal distances to the observed points.

Inferring Rank II Incidence

Let $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}$ be an equatorial cluster of unit vectors on the sphere. Each $\boldsymbol{x}_i$ is assumed to be a noisy observation of some true point $\boldsymbol{\alpha}_i$ lying exactly on a great circle with polar axis $\boldsymbol{\theta}$. The goal is to estimate $\boldsymbol{\theta}$. Each true value $\boldsymbol{\alpha}_i$ is a nuisance parameter related to the desired vector $\boldsymbol{\theta}$ by the constraint equation $\boldsymbol{\theta}^t\boldsymbol{\alpha}_i = 0$. In the presence of nuisance parameters, equation (3.29) is the relevant form of Bayes rule to use.

Constrained Bayesian estimation can usually be handled by incorporating the constraints into the prior density so that parameter values that don't satisfy the constraint have zero prior probability [Box73]. Consider the log prior

$$\begin{aligned} log\, p(\boldsymbol{\theta}, \boldsymbol{\alpha}) &= C + log\, p(\boldsymbol{\theta}) + log\, p(\boldsymbol{\alpha}|\boldsymbol{\theta}) & (3.37) \\ &= C + \boldsymbol{\theta}^t\, \boldsymbol{M}_0\, \boldsymbol{\theta} + \sum_{i=1}^{m} log\, p(\boldsymbol{\alpha}_i|\boldsymbol{\theta}) & (3.38) \\ &= C + \boldsymbol{\theta}^t\, \boldsymbol{M}_0\, \boldsymbol{\theta} + \sum_{i=1}^{m} \boldsymbol{\alpha}_i{}^t\, (-\infty\, \boldsymbol{\theta}\boldsymbol{\theta}^t)\, \boldsymbol{\alpha}_i & (3.39) \end{aligned}$$

where the prior density for $\boldsymbol{\theta}$ is specified as a Bingham density function with parameter matrix $\boldsymbol{M}_0$ and $-\infty$ is shorthand notation meaning $-k$ as $k \to \infty$. Recall from example 3.2 that $exp\,\{\boldsymbol{\alpha}_i{}^t\,(-k\boldsymbol{\theta}\boldsymbol{\theta}^t)\,\boldsymbol{\alpha}_i\}$ is a rotationally symmetric Bingham density with probability mass clustered around $G(\boldsymbol{\theta})$, the great circle of points perpendicular to $\boldsymbol{\theta}$. In the limit as k goes to infinity, the probability density goes to zero everywhere except on the great circle of points $G(\boldsymbol{\theta})$. This limiting prior density function therefore constrains every $\boldsymbol{\alpha}_i$ to lie on a great circle perpendicular to $\boldsymbol{\theta}$.

Each observation $\boldsymbol{x}_i$ is a noisy observation of some true value $\boldsymbol{\alpha}_i$. As before, let an estimate of the uncertainty with which $\boldsymbol{x}_i$ is measured be represented as a bipolar Bingham probability density function with modal axis aligned with $\boldsymbol{x}_i$, and with known parameter matrix $\boldsymbol{M}_i$. By (3.28) the log likelihood of $\boldsymbol{x}_i$ given a true value $\boldsymbol{\alpha}_i$ is

$$log\, p(\boldsymbol{x}_i|\boldsymbol{\theta}, \boldsymbol{\alpha}_i) = C + \boldsymbol{\alpha}_i^t\, \boldsymbol{M}_i\, \boldsymbol{\alpha}_i\ , \qquad (3.40)$$

and the log likelihood of all observations jointly is

$$log\, p(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{\alpha}) \quad = \quad C + \sum_{i=1}^{m} \boldsymbol{\alpha}_i^t\, \boldsymbol{M}_i\;\; \boldsymbol{\alpha}_i \tag{3.41}$$

Combining the log prior density (3.39) with the log likelihood (3.41) results in the log posterior density

$$log\, p(\boldsymbol{\theta},\boldsymbol{\alpha}|\boldsymbol{x}) \quad = \quad C + \boldsymbol{\theta}^t\, \boldsymbol{M}_0\, \boldsymbol{\theta} + \sum_{i=1}^{m} \boldsymbol{\alpha}_i^t\, (\, \boldsymbol{M}_i \; - \infty\, \boldsymbol{\theta}\boldsymbol{\theta}^t\,)\, \boldsymbol{\alpha}_i \tag{3.42}$$

The joint posterior density function for $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ is therefore proportional to

$$C\;\; exp\{\, \boldsymbol{\theta}^t\, \boldsymbol{M}_0\, \boldsymbol{\theta}\,\} \prod_{i=1}^{m} exp\{\, \boldsymbol{\alpha}_i^t\, (\, \boldsymbol{M}_i \; - \infty\, \boldsymbol{\theta}\boldsymbol{\theta}^t\,)\, \boldsymbol{\alpha}_i\,\} \tag{3.43}$$

To make inferences about $\boldsymbol{\theta}$ alone, the effects of the nuisance parameters $\boldsymbol{\alpha}$ must be integrated out, leaving a marginal posterior density for $\boldsymbol{\theta}$. The range of integration for each $\boldsymbol{\alpha}_i$ is the surface of the sphere S^2. However, by construction the joint posterior density is zero everywhere except for the great circle of points $G(\boldsymbol{\theta})$. Therefore integration over S^2 reduces to integration over $G(\boldsymbol{\theta})$. That is

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{x}) &= \int p(\boldsymbol{\theta},\boldsymbol{\alpha}|\boldsymbol{x})\; d\boldsymbol{\alpha} \\ &= C\;\; exp\{\, \boldsymbol{\theta}^t\, \boldsymbol{M}_0\, \boldsymbol{\theta}\,\} \prod_{i=1}^{m} \int_{S^2} exp\{\, \boldsymbol{\alpha}_i^t\, (\, \boldsymbol{M}_i \; - \infty\, \boldsymbol{\theta}\boldsymbol{\theta}^t\,)\, \boldsymbol{\alpha}_i\,\}\; dS^2 \qquad (3.44) \\ &= C\;\; exp\{\, \boldsymbol{\theta}^t\, \boldsymbol{M}_0\, \boldsymbol{\theta}\,\} \prod_{i=1}^{m} \int_{G(\boldsymbol{\theta})} exp\{\, \boldsymbol{\alpha}_i^t\, \boldsymbol{M}_i\, \boldsymbol{\alpha}_i\,\}\; dG \qquad (3.45) \end{aligned}$$

The MAP estimate is the value $\hat{\boldsymbol{\theta}} \in S^2$ that maximizes equation (3.45). An exact, closed-form solution seems unlikely.

We now develop an approximate MAP estimator based on two assumptions : 1) the uncertainty in each observation is small, and 2) the uncertainty is rotationally symmetric. Assumption 2 can be enforced in practice by choosing $w_i = -k_2$. This creates a rotationally symmetric density function that consistently overestimates the variance in the measurement errors, a situation that is certainly less harmful than underestimating it. These two assumptions specify that the Bingham density function representing measurement uncertainty in $\boldsymbol{x}_i$ has a parameter matrix of the form $\boldsymbol{M}_i = w_i \boldsymbol{x}_i \boldsymbol{x}_i^t$ with w large. For notational convenience, a uniform prior density function will be assumed for $\boldsymbol{\theta}$, so that $\boldsymbol{\theta}^t \ \boldsymbol{M}_0 \ \boldsymbol{\theta} = 0$.

When w_i is large, the probability mass of the rotationally symmetric bipolar Bingham density function is tightly concentrated in two sharp antipodal peaks. In this case each integral in equation (3.45) can be approximated ([Box73], Section 1.6.2) as

$$\int_{G(\boldsymbol{\theta})} exp\{ \boldsymbol{\alpha}_i^t \ \boldsymbol{M}_i \, \boldsymbol{\alpha}_i \} \, d\boldsymbol{\alpha}_i \ \approx \ 2 \sup_{G(\boldsymbol{\theta})} \ exp\{ \boldsymbol{\alpha}_i^t \ \boldsymbol{M}_i \, \boldsymbol{\alpha}_i \} \quad , \tag{3.46}$$

where $\sup_{G(\boldsymbol{\theta})}$ stands for the maximal value attained when $\boldsymbol{\alpha}_i$ ranges over the great circle of points $G(\boldsymbol{\theta})$.

Substituting in the definition of $\boldsymbol{M}_i$ yields

$$2 \sup_{G(\boldsymbol{\theta})} \ exp\{ \boldsymbol{\alpha}_i^t \ \boldsymbol{M}_i \, \boldsymbol{\alpha}_i \} \ = \ 2 \sup_{G(\boldsymbol{\theta})} \ exp\{ \boldsymbol{\alpha}_i^t \ \boldsymbol{w}_i \boldsymbol{x}_i \boldsymbol{x}_i^t \, \boldsymbol{\alpha}_i \} \tag{3.47}$$

Since the density function specified by $\boldsymbol{M}_i$ is rotationally symmetric the maximum value is attained where $\boldsymbol{G(\theta)}$ comes closest to $\boldsymbol{x}_i$, namely at the point

$$\hat{\boldsymbol{\alpha}}_i \ = \ \frac{\boldsymbol{x}_i - (\, \boldsymbol{\theta}^t \boldsymbol{x}_i \,) \, \boldsymbol{\theta}}{1 - (\, \boldsymbol{\theta}^t \boldsymbol{x}_i \,)^2} \quad . \tag{3.48}$$

Inserting $\hat{\boldsymbol{\alpha}}_i$ into equation 3.47 yields a final simplification

$$2 \sup_{G(\boldsymbol{\theta})} exp\{\, \hat{\boldsymbol{\alpha}}_i{}^t \; w_i \boldsymbol{x}_i \boldsymbol{x}_i^t \, \hat{\boldsymbol{\alpha}}_i \,\} \;=\; exp\{\, 2 \, w_i (1 - (\, \boldsymbol{\theta}^t \boldsymbol{x}_i \,)^2) \,\} \quad . \tag{3.49}$$

Inserting this result into equation (3.45) shows that an approximate log posterior density function for the polar axis $\boldsymbol{\theta}$ of an equatorial cluster of observations $\boldsymbol{x}_i$ is given by

$$log\, p(\boldsymbol{\theta}|\boldsymbol{x}) \;\approx\; C + \sum_{i=1}^{m} 2w_i \{1 - (\, \boldsymbol{\theta}^t \boldsymbol{x}_i \,)^2\} \tag{3.50}$$

$$= \; C \;-\; \boldsymbol{\theta}^t \left(\sum_{i=1}^{m} w_i \, \boldsymbol{x}_i \boldsymbol{x}_i^t \right) \boldsymbol{\theta} \quad . \tag{3.51}$$

Let W be the following weighted scatter matrix

$$\boldsymbol{W} = \sum_{i=1}^{m} w_i \, \boldsymbol{x}_i \boldsymbol{x}_i^t \quad . \tag{3.52}$$

The MAP estimate for polar axis $\boldsymbol{\theta}$ of the great circle of observations is the unit vector $\hat{\boldsymbol{\theta}}$ that *minimizes* $\hat{\boldsymbol{\theta}}^t \boldsymbol{W} \hat{\boldsymbol{\theta}}$. This is found computationally as the eigenvector associated with the smallest eigenvalue λ_1 of $\boldsymbol{W}$.

Confidence regions for this estimator can be derived by analogy to a method presented in [Watson83] for testing coplanarity of random samples from the Fisher distribution on the sphere. The following approximate results hold when the concentration parameter w_i is large:

$$2 \, w_i (\, \boldsymbol{\theta}^t \boldsymbol{x}_i \,)^2 \quad \sim \quad \chi_1^2 \tag{3.53}$$

$$2 \sum_{i=1}^{m} w_i (\, \boldsymbol{\theta}^t \boldsymbol{x}_i \,)^2 \quad \sim \quad \chi_m^2 \tag{3.54}$$

$$2 \sum_{i=1}^{m} w_i(\hat{\boldsymbol{\theta}}^t \boldsymbol{x}_i)^2 \quad \sim \quad \chi^2_{m-2} \tag{3.55}$$

where χ^2_d is a chi-squared random variable with d degrees of freedom.

Then, by analogy with the analysis of variance (ANOVA), an approximate confidence region for a proposed polar axis θ is

$$\frac{\sum_{i=1}^{m} w_i(\boldsymbol{\theta}^t \boldsymbol{x}_i)^2 \; - \; \sum_{i=1}^{m} w_i(\hat{\boldsymbol{\theta}}^t \boldsymbol{x}_i)^2}{\sum_{i=1}^{m} w_i(\hat{\boldsymbol{\theta}}^t \boldsymbol{x}_i)^2} \quad \leq \quad \frac{2\, F_\alpha(2, m-2)}{m-2} \quad , \tag{3.56}$$

where $F(2, m-2)$ is an F variable with 2 and $m-2$ degrees of freedom. This can be written more compactly in matrix form as

$$\frac{\boldsymbol{\theta}^t \boldsymbol{W} \boldsymbol{\theta} - \lambda_1}{\lambda_1} \quad \leq \quad \frac{2\, F_\alpha(2, m-2)}{m-2} \tag{3.57}$$

where W is the scatter matrix of equation 3.52 and λ_1 is its smallest eigenvalue.

CHAPTER 4

APPLICATIONS

Chapter 3 developed a probabilistic framework for uncertain geometric reasoning in projective space. Concrete applications of this approach are presented in this chapter. Three computer vision applications are addressed : the derivation of line and plane orientations using vanishing point analysis, partitioning of scene features into planar patches using line correspondence stereo, and extending a partial model of planar surface structure using projective invariants.

Each application considered here involves the estimation of a vector of homogeneous coordinates from a set of noisy observations related to the desired value by either Rank I or Rank II incidence constraints. Since these incidence constraints are invariant under affine and projective transformations, the results derived will be valid regardless of camera calibration or pose. However, the derivation of projective structure is more easily comprehended and evaluated when converted into Euclidean form. This is possible when the intrinsic camera calibration parameters are known. The ideal pinhole image can then be recovered, and the homogeneous coordinate vectors of points and lines gain additional interpretations as 3D scene orientations. Validity of projective inferences can then be evaluated in terms of Euclidean relationships such as perpendicularity that are known to hold. For this reason, information about the camera calibration parameters has been used whenever possible to convert derived homogeneous coordinate vectors into camera-centered estimates of 3D orientation.

4.1 Vanishing Point Analysis

Under perspective projection, parallel lines in three-space project to converging lines in the image plane. The common point of intersection, perhaps at infinity, is called the *vanishing point*. Vanishing point analysis provides strong cues for inferring the 3D structure of a scene from only a single view [Haralick80, Horaud87]. Under known camera geometry, the vanishing point of a group of parallel lines determines their orientation in three-space. Two or more vanishing points from lines known to lie in a single 3D plane establish a *vanishing line,* which completely determines the orientation of the plane. A more complete overview and a review of previous work was presented in Section 2.1.2.

Vanishing point analysis requires parallel, linear structures to be present in the scene, and their projections to be identified in the image. Figure 4.1 shows a set of straight line segments extracted from an outdoor building scene by the Boldt straight line extraction algorithm [Boldt89]. Identifying converging lines in the image is easiest when there many lines in each convergent cluster, since an efficient clustering mechanism (to be described shortly) is applicable. However, as lines become more sparse the clustering method begins to break down, to the point where it is hard to distinguish vanishing point clusters from accidental line intersections. Although these requirements limit vanishing point analysis to scenes containing lots of parallel linear structure, the techniques that are used, most notably detecting clusters of converging lines using a histogram data structure on the unit sphere, and the statistical estimation of the axis of a great circle of points, are applicable to more general cases.

Figure 4.1: Straight line segments from an outdoor building scene. These line segments were computed by the Boldt straight line extraction algorithm [Boldt89]. There were 2372 lines extracted for this image.

In computer vision, the pure geometry of vanishing point analysis must be applied to noisy observations of the physical world. The slopes of line segments in the image are perturbed by noise both in the imaging process and in the method of line segment extraction. Finding the vanishing point from a set of converging lines becomes a problem of estimation. A practical algorithm for finding vanishing points from a set of line segments in an image must therefore address two issues : how to cluster line segments going to a single vanishing point, and how to estimate an accurate vanishing point from a given line cluster.

4.1.1 Vanishing Point Detection

Vanishing point detection involves finding clusters of lines that nearly converge to a single point of intersection. A Hough-transform approach originally due to Barnard excels at quickly clustering line segments into potentially convergent groups [Barnard83]. Line segments in the image are mapped onto a histogram representing the surface of the unit sphere $x^2 + y^2 + z^2 = 1$ centered about the camera focal point. In practice, only the positive hemisphere $z > 0$ needs to be represented, and the surface of the hemisphere is partitioned by longitude and colatitude (see Figure 4.2). The sphere is a more appropriate histogram space than the image plane because the sphere is a compact, finite surface, while the image plane is not.

Each line segment in the image, taken together with the camera focal point, forms a *projection plane* which intersects the unit (hemi)sphere in a great (semi)circle. Each histogram bucket maintains a count of the number of great circles passing through it. Potential vanishing points are detected as peaks in the histogram, corresponding to areas where several great circles intersect, and the vanishing point location is chosen as the center of the histogram bucket containing a peak.

The implementation used in this thesis contains some modifications to the basic algorithm described above. The most fundamental change is that the histogram data structure is applied here as an initial clustering method and as an efficient spatial access mechanism, but the final analysis of vanishing point location is performed on the underlying data. To

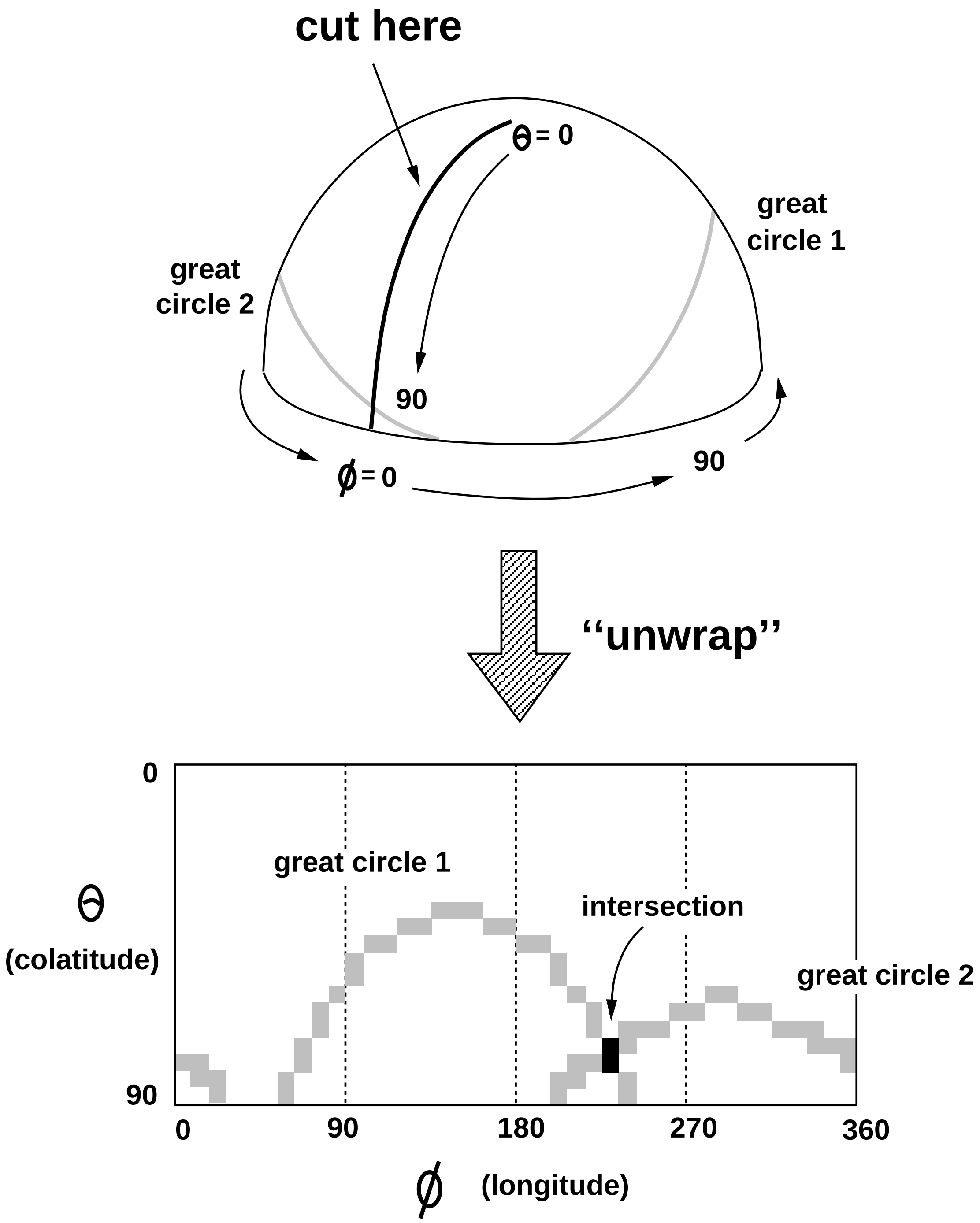


Figure 4.2: Barnard's histogram method for finding vanishing points. A hemispherical histogram is partitioned by longitude and colatitude. For each line segment in the image, a great circle of histogram cells is incremented. Potential vanishing points are detected as peaks in the histogram, corresponding to areas where several great circles intersect.

achieve this, each histogram bucket maintains a list of the line segments that pass through it in addition to the count. After a peak is detected, the line segments within it are retrieved and statistical estimation techniques are applied to derive a more accurate estimate of vanishing point position.

Peak selection in Hough-transform methods is inherently problematic [Grimson90]. In particular, when the true position of a vanishing point on the sphere falls near a histogram boundary, candidate line segments that should be grouped together fall into separate buckets. For this reason, the set of line segments forming a convergent cluster is actually collected from a whole neighborhood of buckets surrounding the peak. The neighborhood is specified as a circular region with some given radius ρ on the sphere, where ρ is a user-specified parameter that is typically set between 3 and 5 degrees. The cluster selection mechanism works as follows. Let $\boldsymbol{x}$ be a unit vector representing the projection plane normal of a line segment in the image, and $G(\boldsymbol{x})$ represent the great circle of points having polar axis $\boldsymbol{x}$. Once a peak is found, the set of n projection plane normals $\{\boldsymbol{x}_i\}$ corresponding to the n great circles passing through the peak bucket are selected, and the direction along which the dispersion of these vectors is minimal is computed as the eigenvector $\hat{\boldsymbol{u}}$ associated with the smallest eigenvalue of the scatter matrix $\sum \boldsymbol{x}_i \boldsymbol{x}_i^t$. Vector $\hat{\boldsymbol{u}}$ is simultaneously an estimate of the polar axis of the great circle that best fits the set of vectors $\{\boldsymbol{x}_i\}$, and an estimate of the point of intersection of the set of great circles $\{G(\boldsymbol{x}_i)\}$. Given the estimated intersection point $\hat{\boldsymbol{u}}$, all histogram buckets that intersect the circular area $|\hat{\boldsymbol{u}} \cdot \boldsymbol{v}| \geq \cos\rho$ are activated, and all projection plane normals $\boldsymbol{x}_i$ within them are subjected to the boolean filter

$$|\, \hat{\boldsymbol{u}} \, \cdot \, \boldsymbol{x}_i \,| \; \leq \; \sin\rho \quad . \tag{4.1}$$

Every projection plane normal $\boldsymbol{x}_i$ that passes the filter is guaranteed to lie within a belt of radius ρ around great circle $G(\hat{\boldsymbol{u}})$. Each corresponding great circle $\mathrm{G}(\boldsymbol{x}_i)$ is therefore known to pass through a circular area of radius ρ around the proposed intersection point $\hat{\boldsymbol{u}}$. Line segments in the image corresponding to these converging great circles are returned as a cluster.

The procedure just outlined relies on the ability to determine which set of buckets in the hemispherical histogram are neighbors of any given peak bucket. This must be done with care, however. Representing the surface of the unit (hemi)sphere as a 2D rectangular array of grid cells is not faithful to the topology of the sphere, and must by necessity "break" the neighborhood structure in some places. Let the spherical histogram be indexed by longitude $0 \leq \phi < 360°$ and by colatitude $0 \leq \theta \leq 90°$ such that the corresponding point on the sphere has coordinate vector $(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$. Referring back to Figure 4.2, the process of clustering adjacent accumulator cells is complicated by at least three factors

wraparound : All arithmetic involving longitude ϕ is modulo 360 degrees, i.e. $(\phi = 0, \theta = \alpha)$ is the same point as $(\phi = 360, \theta = \alpha)$. Thus pairs of cells on the extreme right and left sides of the ϕ–θ array are adjacent.

singularity : The ϕ–θ coordinate system has a singularity at the north pole, since that point is represented by $\theta = 0$ degrees while ϕ may take any arbitrary value. Therefore the whole row of cells along the top of the array are actually adjacent to each other.

instability : When lines are parallel, their intersection lies along the equator $\theta = 90$, but the intersection can be at either $\phi = +\alpha$ or $\phi = -\alpha$. Therefore, along the bottom row of the accumulator array, pairs of points differing by 180 degrees (modulo 360) are adjacent.

A routine has been implemented to iteratively visit every grid cell falling within a circle of radius ρ from a given point on the hemisphere, taking into account all of the above topological considerations.

After the largest cluster of converging lines is detected, all lines contained within it are deleted from the histogram, and all bucket counts are updated. The highest remaining peak is taken as the second cluster, and so on. Multiple peaks are thus detected in decreasing order of number of lines contributing to them. The number of vanishing point clusters to look for is a user-supplied parameter – usually less than three appear in any image.

Figure 4.3 illustrates the extraction of the three largest clusters of converging line segments from Figure 4.1. Notice that the problem of false peaks associated with coincidental line intersections becomes apparent after the lines associated with the largest two peaks are removed. It is not uncommon to have such false peaks, particularly when repetitive, "checkerboard" patterns are seen, such as the pattern formed by the windows on the wall of the far building in this image.

4.1.2 Vanishing Point Estimation

Once image line segments have been clustered into convergent groups (line pencils), the problem remains of estimating the correct vanishing point location. The geometry of vanishing point estimation was described in detail in Section 2.1.2. The homogeneous coordinate vectors of every line in a line pencil, when normalized to have unit length, form a great circle

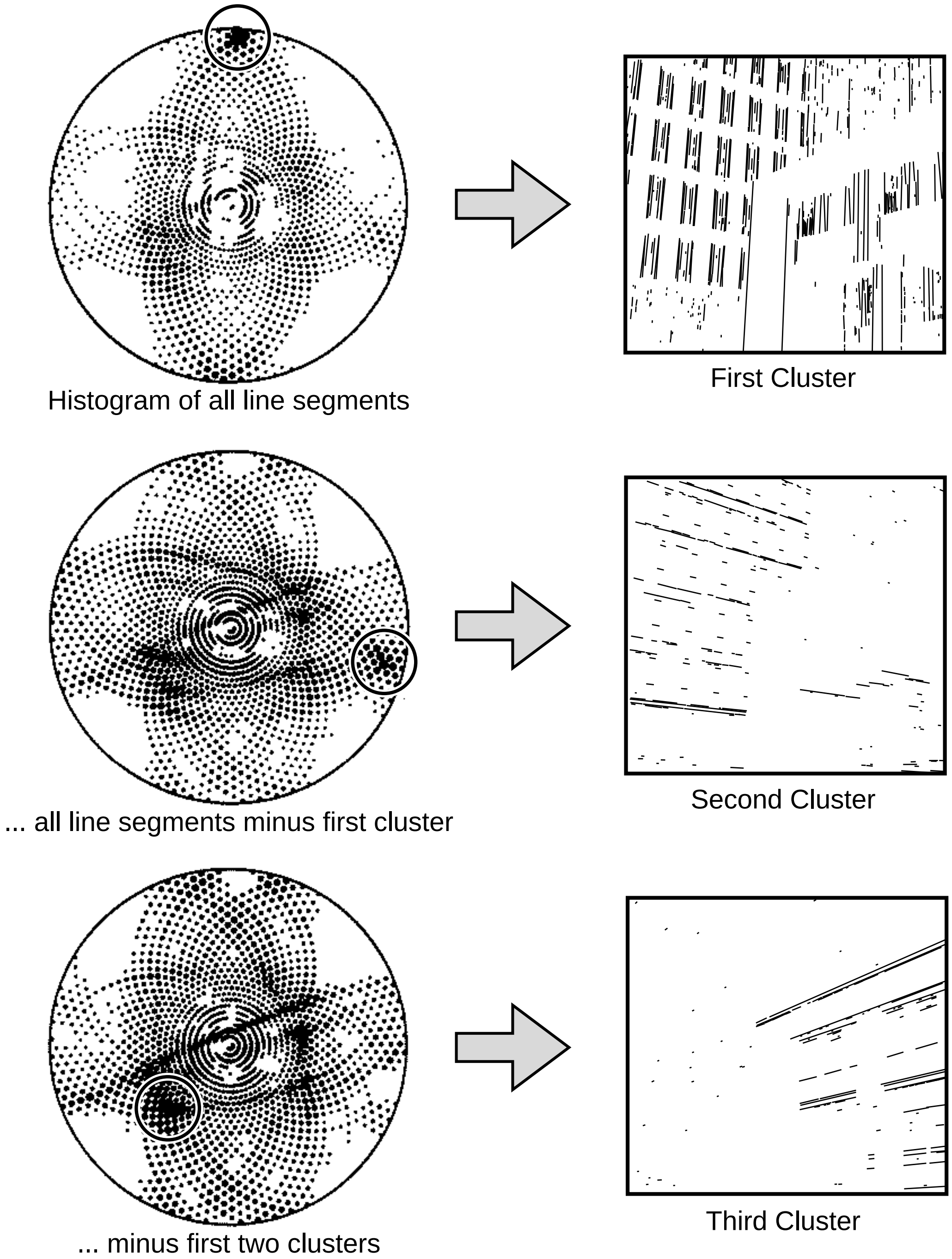


Figure 4.3: The three largest clusters found using Barnard's histogram method. Each histogram bucket has been mapped back onto its proper position on a hemisphere, and drawn as a black dot with radius proportional to the number of lines in the bucket (relative to the number of lines in the highest peak). The number of lines in peaks 1–3 is 536, 216, and 146, respectively.

of points on the sphere. The polar axis of this great circle coincides with the homogeneous coordinate vector of the vanishing point of the line pencil. Due to noise in the extracted image line segments, coplanarity of their normalized coordinate vectors will not be exactly satisfied. The corresponding points on the sphere no longer lie exactly on a great circle, but instead form a cluster around it. Computing the polar axis of this great circle, and hence of the vanishing point, becomes a problem of statistical estimation on the sphere (see Figure 4.4).

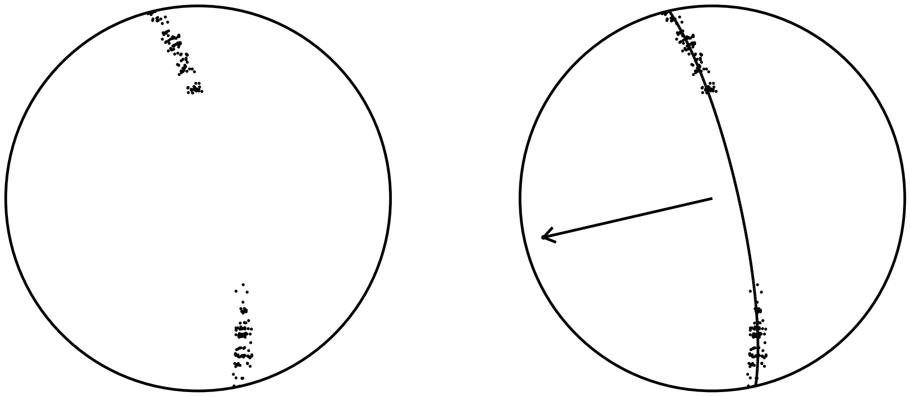


Figure 4.4: Vanishing point estimation on the sphere. The homogeneous coordinate vectors of converging line segments cluster around a great circle on the unit sphere. The polar axis of the best-fitting great circle coincides with the homogeneous coordinates of their vanishing point.

The purpose of this section is to introduce three statistical estimators of the polar axis of a great circle of noisy sample points. Experimental evaluations of these estimators are presented in the following sections. The estimators considered can be broken into two groups on the basis of how knowledge that the given samples cluster about a great circle is incorporated into the estimation process. Under the *functional model* this knowledge is treated as a geometric constraint, while under the *structural model* it is treated as evidence of an underlying probability distribution. To illustrate the difference, the approach to Bayesian inference under Rank II incidence presented in Chapter 3 is a functional approach. The sample points are treated as noisy observations whose unknown true values all lie exactly on some great circle. No other relationship between the points is assumed, and in particular, each true value may be perturbed by unrelated noise processes having completely different covariances. In contrast, [Collins90] presents a structural approach to vanishing point calculation. Each sample point is assumed to be an observation from some equatorial probability density function on the sphere. Let X be a random variable over the surface of the unit sphere, and $f(X)$ be an antipodally symmetric probability density function. In the structural model, the set of sample points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ forms an independent and identically distributed random sample from $f(X)$.

The three estimators to be considered differ not only in their assumptions, but also in their methods of solution. The three estimation frameworks considered are :

- **Bingham structural model** – All observed points $\boldsymbol{x}_i$ are assumed to be independent, identically distributed (i.i.d.) samples from an equatorial Bingham distribution. The parameters of the polar axis of this great circle are estimated using **maximum likelihood** estimation.

- **Nonparametric structural model** – All observed points $\boldsymbol{x}_i$ are assumed to be i.i.d. samples from a general antipodally symmetric distribution. The parameters of the polar axis are estimated using the **method of moments.**

- **Bingham functional model** – This approach was detailed in Section 3.4. The true values $\boldsymbol{\alpha}_i$ of each sample point $\boldsymbol{x}_i$ are assumed to lie exactly on a great circle. Each observed point is generated by an independent, bipolar Bingham density function with known parameter matrix. The polar axis of the great circle of true values is estimated using **Bayesian MAP** estimation.

Bingham Structural Model

Maximum likelihood estimation given a set of i.i.d. samples from a Bingham density function was described in Section 3.1. To restate the relevant results, a sufficient statistic for the orientation and shape parameters $\boldsymbol{U}$ and $\boldsymbol{K}$ is the sample second moment matrix

$$\boldsymbol{M} \;=\; \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i\boldsymbol{x}_i^t) \;=\; \boldsymbol{A\Lambda A}^t, \tag{4.2}$$

factored into $\boldsymbol{A} = [\boldsymbol{a_1}, \boldsymbol{a_2}, \boldsymbol{a_3}]$, an orthogonal matrix of eigenvectors, and diagonal matrix $\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, \lambda_3)$ of corresponding nondecreasing eigenvalues. The maximum likelihood estimate of the orientation matrix $\boldsymbol{U}$ is the matrix of eigenvectors $\boldsymbol{A}$. In particular, the maximum likelihood estimate of the polar axis of both the equatorial and asymmetric bipolar forms of Bingham's distribution is $\hat{\boldsymbol{u}}_1 = \boldsymbol{a}_1$. Maximum likelihood estimates of the nonzero shape parameters k_1 and k_2 are nontrivial functions of Λ. We compute them using a program provided by Dr. John Kent that performs the asymptotic series expansions developed in [Kent87].

For large sample sizes, an approximate $1-\alpha$ confidence region (confidence cone) for $\hat{\boldsymbol{u}}_1$ is the set of unit vectors [Bingham74]

$$\left\{ \boldsymbol{v} : \quad \boldsymbol{v}^t \, \boldsymbol{R} \, \boldsymbol{v} \quad \leq \quad \frac{\chi^2_{2,\alpha}}{2\,n} \right\}, \tag{4.3}$$

where $\boldsymbol{R}$ is the matrix

$$[\boldsymbol{a}_2,\, \boldsymbol{a}_3] \begin{bmatrix} (k_1-k_2)(\lambda_1-\lambda_2) & 0 \\ 0 & k_1(\lambda_1-\lambda_3) \end{bmatrix} [\boldsymbol{a}_2,\, \boldsymbol{a}_3]^t$$

and $\chi^2_{2,\alpha}$ is the upper α critical point of the χ^2 distribution with 2 degrees of freedom.

Nonparametric Structural Model

Prentice develops a nonparametric estimation procedure for antipodally symmetric distributions on the sphere [Prentice84]. His argument, based on the method of moments, is that the population second moment matrix $E(\boldsymbol{x}\boldsymbol{x}^t) = \boldsymbol{U}\boldsymbol{K}\boldsymbol{U}^t$ can be estimated by the sample second moment matrix (4.2). Equating matrix factors in the obvious way leads to estimators $\hat{\boldsymbol{u}}_i = \boldsymbol{a}_i$, so that again the desired polar axis estimate is $\boldsymbol{a}_1$.

Variances of the computed parameters are estimated from the quantities

$$c_{jk} \;=\; \frac{1}{n}\sum_{i=1}^{n}(\hat{\boldsymbol{u}}_j^t\,\boldsymbol{x}_i)^2\,(\hat{\boldsymbol{u}}_k^t\,\boldsymbol{x}_i)^2, \qquad 1 \leq i, j \leq 3$$

which are the nonzero sample fourth moments in the coordinate system of the principle axes. Again invoking the method of moments, and assuming large sample sizes, an approximate $1-\alpha$ confidence region for $\hat{\boldsymbol{u}}_1$ is found to be

$$\left\{\boldsymbol{v}:\;\; \boldsymbol{v}^t\,\boldsymbol{Q}\,\boldsymbol{v} \;\;\leq\;\; \frac{\chi^2_{2,\alpha}}{n}\right\}, \tag{4.4}$$

where $\boldsymbol{Q}$ is the matrix

$$[\boldsymbol{a}_2,\,\boldsymbol{a}_3]\begin{bmatrix}(\lambda_1-\lambda_2)^2/c_{12} & 0\\ 0 & (\lambda_1-\lambda_3)^2/c_{13}\end{bmatrix}[\boldsymbol{a}_2,\,\boldsymbol{a}_3]^t$$

and $\chi^2_{2,\alpha}$ is as before.

Bingham Functional Model

This problem was the treated in Section 3.4.2 using Bayesian MAP estimation. Exact computation of the resulting estimator turned out to be intractable. Instead, a computationally simple approximation to the MAP estimator was developed. Let l_i be the length of the observed image line segment associated with sample point $\boldsymbol{x}_i$. A *weighted* scatter matrix is then written as

$$\boldsymbol{W} \;=\; \sum_{i=1}^{n} l_i^2\,(\boldsymbol{x}_i\,\boldsymbol{x}_i^t) \quad . \tag{4.5}$$

Let λ_1 be the smallest eigenvalue of W, a_1 be its associated eigenvector, and let m be the number of lines in the sample cluster. Then an approximate MAP estimator for the polar axis of the great circle of sample points is $\hat{\boldsymbol{u}} = \boldsymbol{a}_1$. Based on an analogy to the analysis of variance procedure, an approximate $1-\alpha$ confidence region for $\hat{\boldsymbol{u}}$ was found to be

$$\frac{\boldsymbol{\theta}^t \boldsymbol{W} \boldsymbol{\theta} - \lambda_1}{\lambda_1} \leq \frac{2\,F_\alpha(2, m-2)}{m-2} \quad , \tag{4.6}$$

where $F(2, n-2)$ is an F variable with 2 and $n-2$ degrees of freedom.

4.1.3 A Numerical Example

Figure 4.1 at the beginning of this section displayed a set of straight line segments extracted from an outdoor building scene. The 512×512 image was taken using a camera with a field of view of roughly 30 degrees. Figure 4.3 showed the largest three line pencils found. The largest cluster contains 536 line segments. Its sample second moment matrix is (all quantities rounded to four decimal places)

$$\boldsymbol{M} = \frac{1}{536} \left| \begin{array}{rrr} 521.8195 & -19.7644 & 2.3812 \\ -19.7644 & 2.1703 & -3.4969 \\ 2.3812 & -3.4969 & 12.0102 \end{array} \right| ,$$

which yields eigenvector and eigenvalue matrices

$$\boldsymbol{A} = \left| \begin{array}{rrr} 0.0305 & -0.0154 & 0.9993 \\ 0.9588 & -0.2817 & -0.0380 \\ 0.2820 & 0.9594 & 0.0049 \end{array} \right|$$

$$\Lambda = \frac{1}{536} \left| \begin{array}{ccc} 0.4189 & 0 & 0 \\ 0 & 12.9985 & 0 \\ 0 & 0 & 522.5825 \end{array} \right| ,$$

ordered so that the eigenvalues appear in increasing magnitude. Both of the structural estimators yield $\boldsymbol{a}_1$, the first column of $\boldsymbol{A}$, as the estimate of the polar axis. The unit vector

(0.0305, 0.9588, 0.2820) is therefore the estimate for the direction $\boldsymbol{u}$ of the original 3D line segments.

Under the Bingham structural assumption, eigenvalues $\lambda_1 = 7.816E^{-4}$, $\lambda_2 = 0.0243$, and $\lambda_3 = 0.975$ yield maximum likelihood estimates $k_1 = -640.23$ and $k_2 = -21.16$ (see [Kent87]). An approximate 95% confidence region (4.3) for $\boldsymbol{u}$ is thus the ellipse centered at $\boldsymbol{a}_1$, with axes of half length

$$b_{12} = \sqrt{\frac{5.991}{15607.822}}\text{rad} = 1.12\text{deg} \quad b_{13} = \sqrt{\frac{5.991}{668631.96}}\text{rad} = 0.17\text{deg},$$

directed along great circles towards axes $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$ respectively.

For the Prentice nonparametric structural estimator, confidence regions are based on the sample fourth moment matrix $\boldsymbol{y}_i = \boldsymbol{A}^t\boldsymbol{x}_i$. This turns out to be

$$\boldsymbol{C} = \begin{vmatrix} 1.5048E^{-6} & 2.327E^{-5} & 7.5682E^{-4} \\ 2.327E^{-5} & 1.0098E^{-3} & 2.3218E^{-2} \\ 7.5682E^{-4} & 2.3218E^{-2} & 9.5099E^{-1} \end{vmatrix}.$$

The confidence region (4.4) is located and oriented the same as in the Bingham structural case, and has axes of half length

$$p_{12} = \sqrt{\frac{5.991}{12740.426}}\text{rad} = 1.24\text{deg} \quad p_{13} = \sqrt{\frac{5.991}{672178.85}}\text{rad} = 0.17\text{deg}.$$

quite similar to the Bingham half lengths.

Line orientation estimates and associated confidence regions were similarly computed for the other two convergent line clusters in Figure 4.1. For the second largest cluster the estimated line orientation vector was (0.775, -0.1762, 0.6069). An approximate 95% confidence ellipse on the sphere for the Bingham assumption has major and minor axis half lengths of 1.35 deg and 0.23 deg. The Prentice nonparametric method yields corresponding half lengths of 1.41 deg and 0.23 deg. For the third largest cluster the estimated 3D line orientation of (-0.6, -0.234, 0.765) is computed. The Bingham 95% confidence ellipse has half lengths of 1.25 deg and 0.2 deg, while the corresponding Prentice half lengths are 1.14 deg and 0.2 deg. No asolute ground truth orientation data is available for this image, however an idea of the relative accuracy of the derived vanishing point orientations is provided by examining the angles between pair of derived vectors. These angles are 87.7°, 88.5°, and 91.5°. The true relative angles are presumably 90°.

Estimates of vanishing point location and confidence were also computed under the Bingham functional model. The orientation estimate for the largest cluster was found to be $(0.041, 0.968, 0.25)$ with 95% confidence region half lengths of .65 deg and .1 deg. For the second largest cluster, the orientation estimate was $(0.8, -0.19, 0.573)$ with confidence half lengths on the sphere of 0.85 deg and .13 deg. Finally, the orientation estimate for the third cluster was $(-0.6, -0.24, 0.76)$ with computed half lengths of .57 deg and .06 deg. As mentioned above, no ground truth orientation data is available for this image, but the relative angles between these computed estimates are 90.2°, 93.6°, and 90.0°. A test on images with ground truth data is described in the next section.

One notable difference between the functional model and the structural model is that the confidence regions for the functional model are much smaller in size than those computed under the structural model. Estimators with small confidence regions are said to be more *statistically efficient* than those with larger regions. Efficient estimators are desireable because they are less likely to accept false positives.

4.1.4 RADIUS Model Board 1 Experiment

The goal of this experiment is to compare 3D line orientations computed via vanishing point analysis against absolute ground truth orientations measured independently. Good ground truth measurements of camera orientation are hard to come by. This experiment uses data supplied through the Arpa-ORD RADIUS project (Research and Development for Image Understanding Systems) [Gerson92, Gee93]. The focus of the RADIUS project is automated site model update and construction from aerial imagery. The images and camera parameters used in this experiment comprise the "model board 1" data set distributed with version 1.0 of the RCDE (RADIUS Common Development Environment) software package [MM93].

Figure 4.5 shows a single image from the eight model board images J1–J8. The scene is a 1:500 inch scale model of an industrial site, built on a table top that can be raised and tilted to simulate a variety of camera altitudes and orientations. For model board images J1–J8 the table was set to simulate aerial photographs taken with a ground sample distance of 18 inches, that is, pixels near the center of the image backproject to quadrilaterals on the ground plane with sides approximately 18 inches long (which in scale model coordinates is 18 / 500 = .036 inches). Each image contains approximately 1320×1035 pixels, with about 11 bits of grey level information per pixel. The dimensions of each image vary slightly because the images have been resampled, and unmodeled geometric and photometric distortions have been introduced to more accurately reflect actual operating conditions.

Buildings in urban or industrial areas are often oriented with respect to an underlying orthogonal grid plan (city blocks, for example). When seen from the air, their roof lines form rectilinear structures with sides running parallel to two dominant, orthogonal directions.

Figure 4.5: Radius model board image J8.

These two sets of parallel lines are typically quite easy to find using the vanishing point clustering algorithm described in Section 4.1.1. When internal camera parameters are known, each recovered vanishing point provides an estimate of the 3D orientation of each set of parallel lines with respect to the camera, and to thereby determine the orientation the camera with respect to the underlying local site coordinate grid. In the model board images the local building coordinate grid is also aligned with the primary compass directions, and thus the orientation of the camera is determined in a global frame of reference as well.

Accurate ground control measurements are available for about 110 points scattered throughout the model. These ground control points were used by Lynn Quam at SRI International to perform a multi-image, block adjustment procedure [ASP80] to compute the camera pose parameters, effective camera focal length, and principle point (image center) for all eight images. For example, the camera pose parameters, principle point and focal length for the image shown in Figure 4.5 are

$$\boldsymbol{R} = \begin{bmatrix} -0.04679 & -0.7617 & -0.6462 \\ 0.9800 & 0.0904 & -0.1775 \\ 0.1936 & -0.6416 & 0.7422 \end{bmatrix} \quad \boldsymbol{T} = \begin{bmatrix} -6280.8250 \\ -996.3453 \\ 7609.7137 \end{bmatrix} \tag{4.7}$$

$$(u_0, v_0) = (\,651.5\,,\,516.0\,)\ \text{pixels} \tag{4.8}$$

$$f = 5700\ \ \text{pixels/microns} \tag{4.9}$$

Based on these computations, the ground truth orientations of the north-south, east-west building horizontals in the coordinate system of the camera are

$$\text{vp1} = (-0.04679, -0.7617, -0.6462)$$

$$\text{vp2} = (0.98, 0.0904, -0.1775)$$

and the affine transformation for mapping pinhole camera coordinates into observed image plane coordinates is

$$\text{Cam}(u_0, v_0, f) = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad . \tag{4.10}$$

Computing Vanishing Point Orientations

Figure 4.6 shows a set of line segments produced from the image in Figure 4.5 by the Burns straight line extraction algorithm [Burns86], as implemented in the KBVision image understanding development system [AAI91]. Line extraction was run on each of the eight images J1–J8. Due to the large size of these images, they were first reduced in resolution to half their original size before line extraction. After line extraction, the segments found were rescaled back into original image coordinates, then filtered so that each line segment in the final set has a length of at least 5 pixels long and a contrast (difference in average grey level across the line) of at least 15 grey levels. This procedure produced roughly 5000 line segments per image.

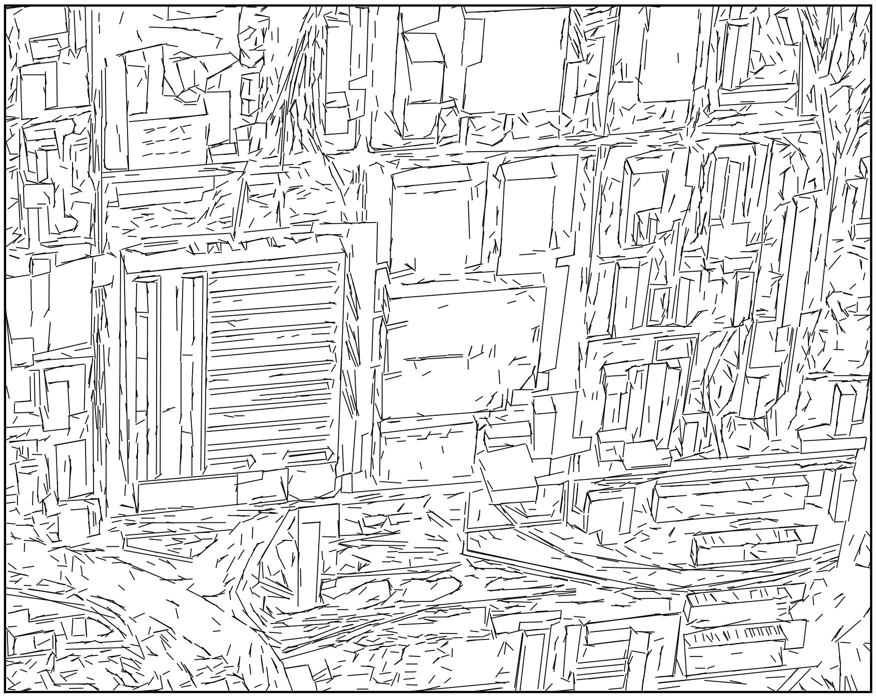


Figure 4.6: Line segments extracted for Radius model board image J8. These lines were found using the Burns straight line extraction algorithm [Burns86].

Theoretically, both vanishing point detection and estimation are invariant to the camera calibration parameters. To test this is practice, both vanishing point detection and estimation were carried out using the following default camera model

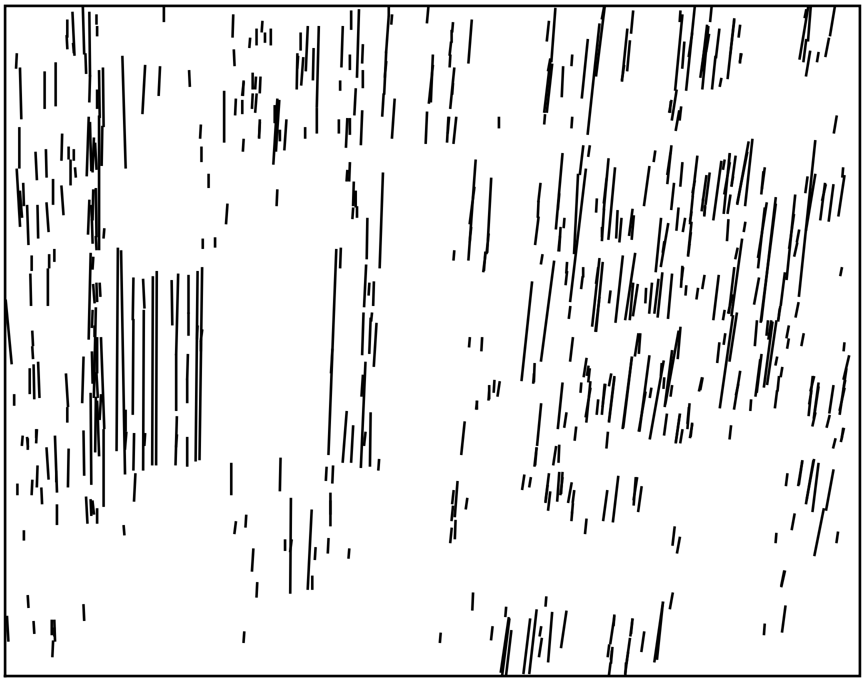
$$u_0* = \text{ImageWidth} \ / \ 2 \quad (4.11)$$

$$v_0* = \text{ImageHeight} \ / \ 2 \quad (4.12)$$

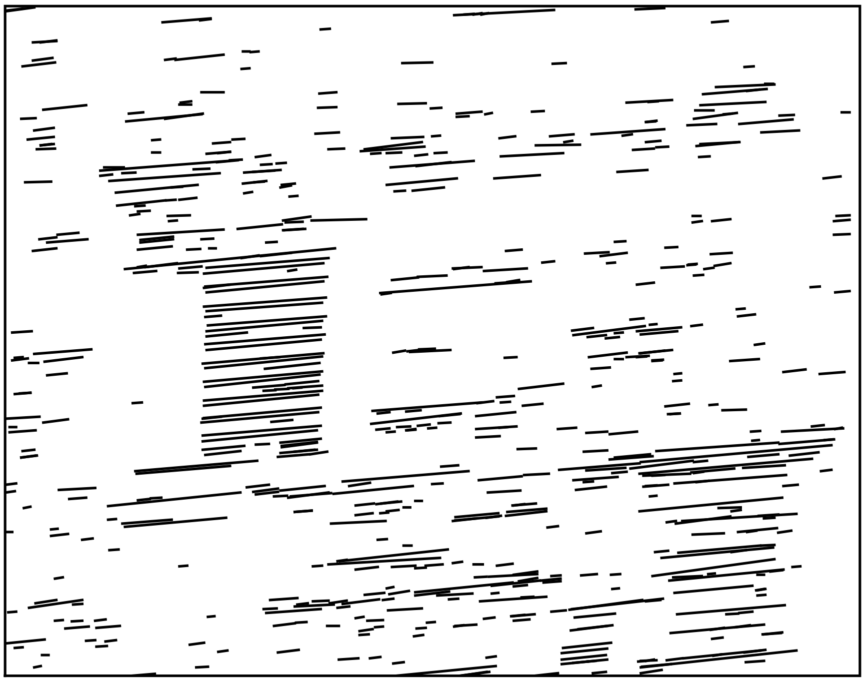
$$f* = (\text{ImageWidth} + \text{ImageHeight}) \ / \ 4 \quad . \quad (4.13)$$

This model places the principle point in the numeric center of an image, and yields a focal length of approximately 90 degrees. For the image in Figure 4.5 the computed focal length was 584, a factor of 10 smaller than the ground truth focal length of 5700.

Figure 4.7 shows the two largest line pencils found by the vanishing point clustering algorithm. As expected, each cluster corresponds to one of the two dominant, orthogonal orientations in the scene. There are roughly 900 line segments in each cluster. Two estimators of vanishing point orientation were tried, corresponding to the Bingham structural model and the Bingham functional model of Section 4.1.2. Both estimators were run on the two dominant line pencils found for each image to produce homogeneous coordinate estimates for both the north-south and east-west vanishing points. The estimators were applied using the default (and incorrect) camera model described above. In order to compare the resulting vanishing point estimates to ground truth data, it is necessary to first convert from default



(a)



(b)

Figure 4.7: The two largest line pencils found for image J8. Each cluster corresponds to one of the two dominant orthogonal orientations in the scene.

coordinates to ground truth coordinates. The proper conversion is performed by the following projective transformation

$$\boldsymbol{v}_{\mathrm{gt}} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} f* & 0 & u_0* \\ 0 & f* & v_0* \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{v}_{\mathrm{def}} \tag{4.14}$$

where $\boldsymbol{v}_{\mathrm{def}}$ is the estimated homogeneous coordinate vector of a vanishing point in the default pinhole coordinates, and $\boldsymbol{v}_{\mathrm{gt}}$ is that same estimate after conversion to ground truth pinhole coordinates. Homogeneous coordinate vector $\boldsymbol{v}_{\mathrm{gt}}$ can also be interpreted as a 3D orientation estimate for one set of parallel lines in the scene.

Comparison with Ground Truth Orientations

The goal of this experiment was to compare estimated vanishing point orientations with ground truth orientations provided with the data set. Table 4.1 summarizes the results for both Bingham structural (S) and Bingham functional (F) vanishing point estimators.

For each image, absolute errors between estimated and ground truth line orientations for the two dominant orthogonal vanishing points in the scene are compared, along with the computed relative angle between the estimated orientations. The functional estimator almost always outperforms the structural estimator. Sample means for the absolute errors in the two vanishing point orientations are 3.5 and 4.4 degrees for the structural model, as compared to 2.7 and 2.3 degrees for the functional model. The reason is that the structural estimator is based on the assumption that all line segments in the image are independent and identically distributed, while the functional model takes into account the fact that longer lines tend to be more stable than shorter lines, and they thus receive higher weights in the final estimate.

Table 4.1: RADIUS Model Board experimental results. Comparison of absolute orientation errors in estimated vanishing point orientations. S = Bingham structural estimator, F = Bingham functional estimator. Also compared is the relative angle between the estimated orientations.

Image	Abs Error in VP1 (degrees)		Abs Error in VP2 (degrees)		Relative Angle (degrees)	
	S	F	S	F	S	F
J1	4.5	3.7	3.8	1.0	90.5	91.2
J2	4.2	3.7	4.7	2.8	91.7	91.7
J3	1.7	1.5	2.5	1.1	89.6	89.5
J4	5.4	4.7	3.0	0.5	86.5	87.3
J5	2.7	1.9	2.2	0.5	91.5	91.1
J6	0.9	2.4	5.0	2.3	89.5	90.3
J7	5.3	1.9	10.2	6.8	86.2	87.7
J8	3.1	2.0	3.8	3.5	92.3	91.9
avg	3.5	2.7	4.4	2.3	89.7	90.1
std	1.6	1.1	2.5	2.1	2.3	1.8

Interestingly, even though the functional estimator outperforms the structural case in terms of absolute error, in this experiment they both perform equally well at estimating the 90 degree relative angle between the two dominant scene orientations. The average relative angle computed was 89.7 for the structural model, and 90.1 for the functional case, both quite accurate.

Although the model board scenes are ideal for vanishing point detection, in the sense that there are many parallel lines in two distinct orientations, they are among the harder cases for accurate vanishing point estimation since the convergence angle of line segments across the image is never very large. In these images each vanishing point is significantly off the image plane, and in many cases is far off, near infinity. The benefits of computing line intersections on the sphere rather than the extended image plane are very apparent in these cases.

4.1.5 Studies using Simulated Data

This experiment was run to determine the accuracy of the confidence regions generated by each of the three vanishing point orientation estimators being considered here. For example, all confidence regions for all three estimators are based on asymptotic distributions of the eigenvectors of the second moment matrix, and are thus only valid for large sample sizes. How do these constructions behave under small sample sizes? As an initial step toward answering this question, all three methods of confidence region construction were tested on simulated data. Line orientation estimates and confidence regions were computed for linesets of size n = 5, 10, 20, 40, 60, 80 and 100. One thousand trials were run for each size n. For each trial, the unit hemisphere was sampled uniformly to choose a random 3D line orientation, and a sample of n lines was generated in a 512×512 image, assuming a camera with a 45 deg field of view. For each line, a uniform random point in the image was chosen, and a length between 5 and 50 pixels long was selected. A line of the selected length was placed with its center on the chosen image point, and oriented towards the vanishing point defined by the given 3D line orientation. After placement, each endpoint of the line segment was perturbed independently by a bivariate normal noise process centered at the endpoint, having standard deviations of 1.5 pixels along the line segment, and 0.5 pixels perpendicular to the line.

For each of the 1000 trials of size n, a 3D line orientation was estimated from the sample line segments. Both Bingham and Prentice structural confidence regions were constructed for 6 χ^2 values, corresponding to 90%, 92.5%, 95%, 97.5%, 99% and 99.5% levels of confidence.

Table 4.2: Bingham structural confidence region counts

number of lines	expected number of members					
	900	925	950	975	990	995
100	827	862	900	938	969	986
80	845	876	917	953	970	985
60	856	885	914	952	981	990
40	874	904	932	955	979	987
20	835	863	899	934	971	982
10	797	837	878	906	937	959
5	650	693	737	783	832	862

Table 4.3: Prentice structural confidence region counts

number of lines	expected number of members					
	900	925	950	975	990	995
100	817	852	896	932	974	987
80	840	874	910	935	982	988
60	845	886	913	952	984	992
40	876	907	930	957	981	988
20	826	860	896	942	977	987
10	765	808	854	893	931	942
5	608	639	681	734	785	811

Bingham functional confidence regions were computed for F distribution values at the same levels of confidence. The original 3D line orientation was then tested for inclusion in each of these regions. The resulting counts are an indication of the actual level of accuracy of each confidence region, for varying sample sizes (Tables 4.2, 4.3 and 4.4).

The derived confidence regions for the structural models appear to be on the small side; the true orientation was contained in a confidence region fewer times than expected, almost

Table 4.4: Bingham functional confidence region counts

number of lines	expected number of members					
	900	925	950	975	990	995
100	904	921	952	974	982	992
80	906	924	953	981	994	996
60	887	914	935	970	986	993
40	908	928	948	967	983	990
20	907	929	947	973	989	992
10	871	900	928	959	980	987
5	887	905	937	969	985	990

uniformly across all confidence levels and lineset sizes. This suggests that the simulated sample distributions did not exactly conform to either distributional assumption. Furthermore, accuracy of the confidence regions in the Prentice structural table do not differ markedly from those in the Bingham structural table, a feature already noted in the numerical example of the last section. Finally, the accuracy levels of all columns stays roughly the same as the number of lines in the sample drops from 100 down to 20. Below 20 the accuracy degrades, dropping notably for small sample sizes of 5. This suggests that the asymptotic arguments involved in the construction of the confidence regions hold for clusters of at least 20 lines.

In contrast, the confidence regions for the Bingham functional model are markedly better. This is evidence that the assumptions underlying the Bingham functional model are closer to the truth. In particular, the functional model dictates that longer lines should be weighted more heavily that shorter lines, since they are expected to be more resistant to noise in the imaging process.

4.2 Line-Correspondence Stereo

This section presents an application of determining line and surface orientation from stereo line segment correspondences. The orientation of 3D lines can be computed directly from stereo line correspondences without first computing point depths, in contrast to methods that compute depth in order to obtain the orientation of line segments. After computing line directions, it is possible to efficiently discover coplanar lines and thereby recover the orientation and distance of planar surfaces in the scene.

A persuasive case is made by Wolff and Boult [Wolff89] that line and plane orientations are more accurately derived from line-based stereo than from point-based stereo. In particular, they show that line orientation computed by intersecting the projection planes of corresponding line segments in a stereo pair is relatively invariant to the 3D locations of each camera in space. Two advantages result from this observation. First, errors in the derived line and plane orientations are relatively independent of errors in the stereo baseline, i.e. knowledge of the absolute translation between the two cameras is not necessary. Secondly, the rate of error in derived orientations grows fairly slowly with increasing distance from the object to the baseline, as compared to the steeper error growth curve of point-based reconstructions. Wolff and Boult verify both of these hypothesized advantages by Monte-Carlo simulation. Note, however, that the authors leave the determination of planar patches from line correspondence stereo as an open problem. One approach to solving this problem is presented in this section.

4.2.1 Line Orientations from Stereo

This section only considers the case of pure translational stereo, that is, the relative poses of the two stereo cameras differ only by a pure translation. If the two cameras are oriented differently with respect to each other, and the relative orientation between them is known, the two images can be unwarped with respect to each other to present a pure translational stereo pair via a transformation process known as image rectification [Ayache91].

Consider a 3D line segment with unit orientation vector $\boldsymbol{U}$, projecting onto the image plane of a single camera. The focal point of the camera together with the 3D line defines a plane called the *projection plane* of the line. The image projection of the 3D line lies on the intersection of the projection plane and the image plane, thus the projection plane can be computed given a line segment in the image and the focal point. Since a 3D line lies in its projection plane, the plane normal $\boldsymbol{\phi}$ is perpendicular to the orientation $\boldsymbol{U}$. If the same 3D line is imaged from a second camera, oriented the same as the first, but translated by a vector $\boldsymbol{T}$, a second projection plane that is still perpendicular to $\boldsymbol{U}$ will be measured. The 3D line orientation can thus be recovered as a unit vector parallel to the cross product of the two projection plane normals, except when the line image lies along an epipolar line for the two images.

Looking at this another way, translating the coordinate system by $\boldsymbol{T}$ is equivalent to translating lines by $-\boldsymbol{T}$. Pure translation does not change line orientations, so the new line remains parallel to the original. Under perspective projection, parallel 3D lines of orientation $\boldsymbol{U}$ project to converging image lines which intersect at the vanishing point associated with $\boldsymbol{U}$. Therefore the image of a line in one camera coordinate system intersects the image of the same line in a translated coordinate system at a vanishing point, from which the line direction can be derived.

Unfortunately, the input line data is always imperfect due to noise in its imaging and extraction, and the effects of noise on the computed 3D orientation must be taken into account. When working with real data, any computed quantity should be treated as an estimate only, with an associated measure of uncertainty.

4.2.2 Recovering Planar Surfaces

Having first computed 3D line directions, it is possible to discover coplanar lines and thereby recover the orientation and distance of the planar surfaces that contain them. This is done in two stages. First the lines are broken into groups consistent with a family of parallel planes, then distances are finally computed to partition the lines into sets consistent with individual plane equations.

The normal vector to a planar surface is perpendicular to the orientations of all lines on that surface. Conversely, given a line of orientation $\boldsymbol{U}$ lying on a planar surface, the set of possible surface normals is the set of unit vectors perpendicular to $\boldsymbol{U}$. On the unit sphere, the heads of this set of possible normals trace out a great circle with polar axis $\boldsymbol{U}$. This geometric constraint leads to an efficient Hough transform technique for finding possible surface normals on the sphere, a transform first employed by Barnard in the context of locating potential vanishing points [Barnard83]. Each line orientation is mapped onto a 2D histogram representing the surface of the unit sphere, partitioned by azimuth and elevation. Each orientation casts a vote in all buckets along the great circle representing vectors perpendicular to it. Potential surface normals are detected as peaks in the histogram, corresponding to areas where several great circles intersect.

Once potential plane orientations have been identified, the unit normal can be estimated with more accuracy. Although the true line orientations lie on a great circle around the true plane normal, this relationship will not be exact due to errors in the derived line orientations. Instead, line orientations lie scattered around a great circle. The statistical estimation techniques previously described are used in Section 4.2.3, to estimate the polar axis of this great circle, providing a vector estimate plus an uncertainty region for the 3D surface orientation.

Finally, for lines within a family of parallel planes, a 1D histogram of plane distances is formed by computing for each line a hypothesized distance. The distance is computed as $d_i = \hat{\boldsymbol{n}} \cdot \boldsymbol{p}_i$, where $\boldsymbol{p}_i$ is a 3D point on line i and $\hat{\boldsymbol{n}}$ is the estimated plane normal. Peaks in this 1D histogram represent sets of lines consistent with a single plane equation.

Figure 4.8 shows an example of the partition created for a stereo hallway image. The algorithm forms hypotheses of all three visible wall planes, and correctly identifies that one plane orientation is shared by two parallel planes at different depths.

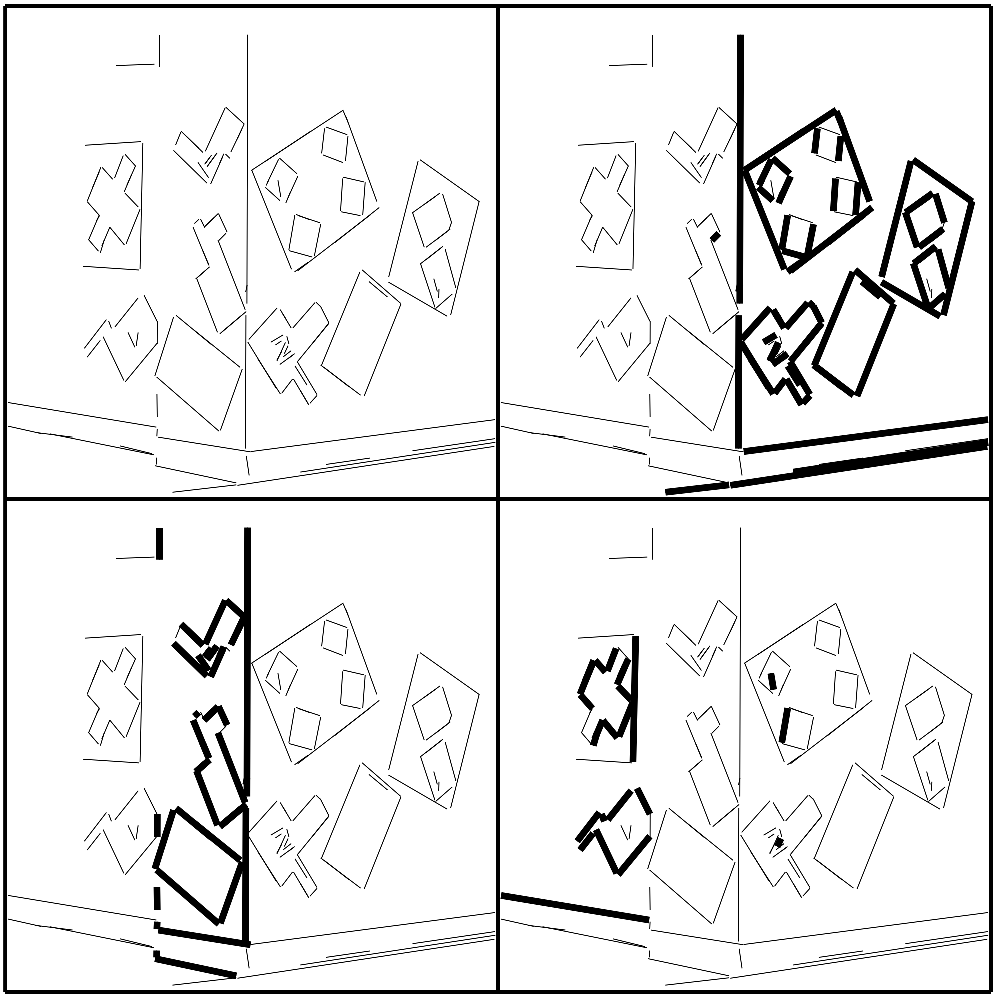


Figure 4.8: Three hypothesized planar surface patches. Line segments extracted from the left image of a translational stereo pair are shown in the upper left. Also shown are three sets of lines consistent with individual planar surface patches.

The above method for hypothesizing planes from stereo line correspondences has some unique features. The method typically employed to solve the same problem is to compute 3D line segments or points, then cluster them into planar patches. The approach presented here has the following advantages:

- The computation of plane distance is decoupled from computing plane orientation, thus the Hough transform for detecting surface normals is 2-dimensional instead of 3, and the entire process is roughly $\mathcal{O}(N)$ in the number of lines [Collins89].
- Parallel planes are immediately identified, and their shared orientation is computed from all of the lines on them.
- When the depth to a point on a line is finally computed, the line is assumed to lie in a plane with a given orientation that has been estimated from several lines. The depth computed under this constraint is presumably more accurate (see [Wolff89]).

4.2.3 An Example

The stereo line correspondence example from the last section was taken with a two-camera setup with parallel focal axes and a baseline of 20 inches. The distance from the left camera to the corner was approximately 18 feet. As described in Section 4.2.1, each line orientation was computed as the cross product of two corresponding projection plane normals. The variance in each input normal was taken to be circular, and inversely proportional to the average length of the matched lines in the image. From these two uncertain projection plane normals, the line orientation was computed as their cross product, and an associated confidence region was computed. Although this generally gave an elliptical confidence region, uncertainty in the line direction was taken to be a circular distribution circumscribing the elliptical one on the sphere, summarizing the region description with a single variance.

Two plane orientations were discovered from the Hough transform array (two planes of the three share the same orientation). To compute plane orientations from each line group found, the approximate MAP estimator developed under the Bingham functional model was used (Section 4.1.2). The mean orientation vector for the right wall was computed to be $(.665, -.061, -.744)$, as depicted in Figure 4.9a. A 95% confidence region around the right wall orientation is approximately an ellipse, with half-lengths on the sphere of 2.48 and 2.76 degrees. The shared orientation of the two left walls was estimated to be $(-.739, -.021 - .673)$, with a 95% confidence region of half-lengths 2.53 and 2.0 degrees. The relative orientation between the two estimated orientations is 89.4 degrees, and the actual walls are in fact perpendicular within the usual limits of construction accuracy.

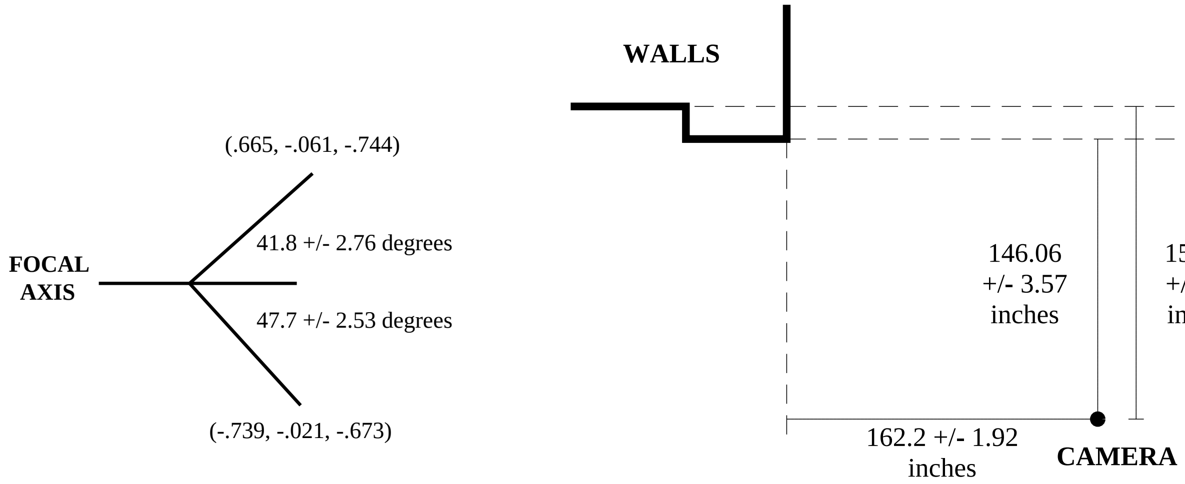


Figure 4.9: Numerical results for planar surface hypotheses. (a) depicts the orientation and uncertainty in surface normals. (b) shows perpendicular distance estimates to each plane.

In partitioning the lines into individual plane equations, the two parallel left walls were separated. The perpendicular distance from the camera to the extended plane of the far left wall was computed as 155.16 ± 4.8 inches, at the 95% confidence level, and that of the closer left wall as 146.06 ± 3.57 inches (see Figure 4.9b). This gives a nominal difference of 9.1 inches, whereas the faces of these two walls are in fact 8.5 inches apart, or a 7% error in depth for the estimated value. A 95% confidence region for the depth of the wall on the right hand side of the image was estimated to be 162.2 ± 1.92 inches.

4.3 Planar Model Extension

Many useful plane to plane mappings can be represented by linear, invertible projective transformations called homographies. A planar homography can be estimated from a small number (four or more) point or line correspondences between the two planes, using the least-squares procedure described in Section A.1.4. 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.

4.3.1 Image Plane to Image Plane

Figures 4.10 and 4.11 present a concrete example of an image to image homographic transformation, in the context of aerial surveying. Figure 4.10a shows a near-nadir aerial photograph of Fort Hood, Texas, taken at a high enough elevation that the whole scene can be considered approximately planar. Extracting line segment features from the image (Figure 4.10c) 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 to form a mosaic.

Figures 4.10b and 4.10d 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. Figure 4.11 shows the final, registered mosaic of the extracted image line segments. Note how the rectangular boundary surrounding the initial line data of Figure 4.10d has been transformed into a quadrilateral as a result of removing the effects of perspective foreshortening.

One real-world issue becomes immediately apparent from this example. If ideal coplanar 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 fusing multiple observations of the same feature. Each point or line feature in homogeneous

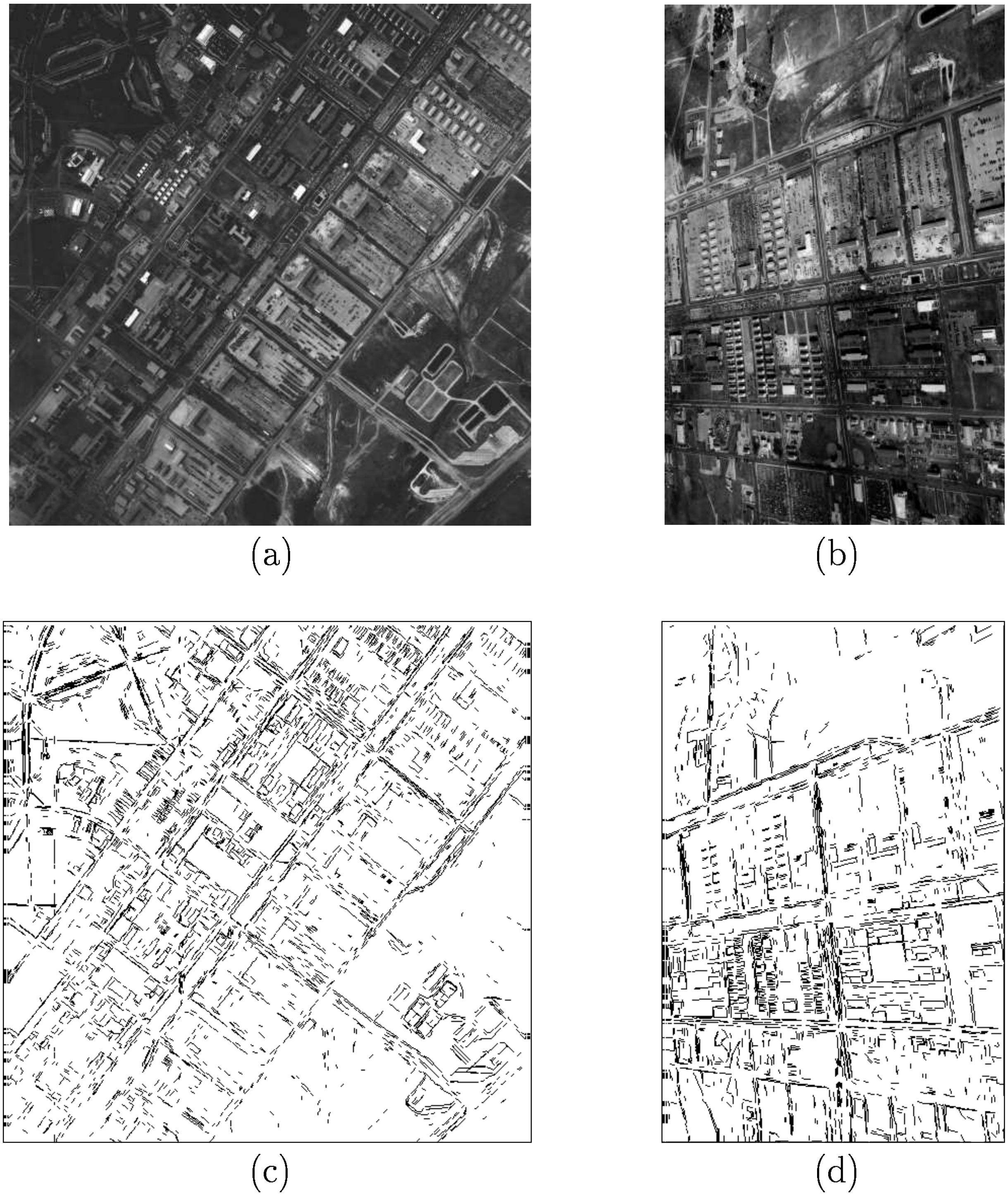


Figure 4.10: Two views of Fort Hood, Texas. (a) A nadir view, suitable for site mapping. (b) An oblique view with strong perspective distortion. (c) Line segment features extracted for the nadir view. (d) Line segments for the oblique view.

Figure 4.11: A final registered mosaic for the two images in Figure 4.10. Line segments from the oblique view have been projectively unwarped into the coordinate system of the nadir view.

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 a point representing the homogeneous coordinates of the true object feature location. This is an example of Rank I incidence, to use the terminology of Section 3.4, and an appropriate framework for estimating the homogeneous coordinates of the true feature location was presented in Section 3.4.2. An application of this merging procedure is presented in the next section, in the context of planar object model extension.

4.3.2 Image Plane to Object Plane

In this section, model extension for planar objects is described. Given at least four object to image correspondences, the homography that maps image points into a local object plane coordinate system can be computed. This transformation allows new observed features in the image to be transferred back into their proper locations on the planar surface of the object, thereby extending the modeled portion of the object.

This approach is similar to one used by [Mohr90], where points are located on an object surface using pairs of cross ratios between an object point and four known object locations. Since the cross-ratio is invariant under homographies, the values in a cross-ratio pair can be computed directly from the image. It is possible to rewrite the mapping effected by Mohr's cross-ratio algorithm as a homography matrix. When exactly four point correspondences are used, the homography estimated using the least squares approach of Section A.1.4 reduces to that used by Mohr. When more than four point or line correspondences are known, direct least squares estimation of the homography matrix will be more accurate.

The PUMA Sequence

An image sequence from the pose estimation literature was chosen to illustrate model extension [Kumar90]. Figure 4.12 shows a typical image from a sequence of 20 images taken by mounting a CCD camera on a PUMA robot arm and rotating the arm 4 degrees between consecutive views. Ground truth data locations of the labeled points (crosses and circles) were measured in a room coordinate system to an accuracy of approximately 0.2 feet along each axis. Kumar used known 3D to 2D correspondences of the 12 ground truth reference points labeled with crosses to compute the pose of the camera for each frame. The positions of the 20 points marked with circles were then estimated via multi-frame triangulation.

The reference points used by Kumar occur in clusters of 4 points on 3 different planar surfaces. In the following experiment, each of these 3 surfaces was treated as a separate object plane containing 4 reference points (crosses) and 4 test points (numbered circles). The far wall (containing points 5–8) will be called "Plane 1", the floor (containing points 9–12) will be labeled "Plane 2", and the side wall (containing points 13–16) will be called "Plane 3". Circled points 1–4 and 17–20 were not used in this experiment. The goal of this experiment was to test the accuracy with which new object points can be located using projective model extension, and to compare these results with those obtained by a more traditional 3D model reconstruction method based on triangulation [Kumar90]. Data fusion of the reconstructed points to derive more accurate location estimates was also performed.

The 3D ground truth coordinate system for the PUMA sequence was aligned with the natural axes of the room. For this experiment, a local 2D coordinate system for each object plane was formed by dropping the coordinate that didn't change for the points in that plane. In this way, each measured 3D object point was converted to a 2D point in a local object

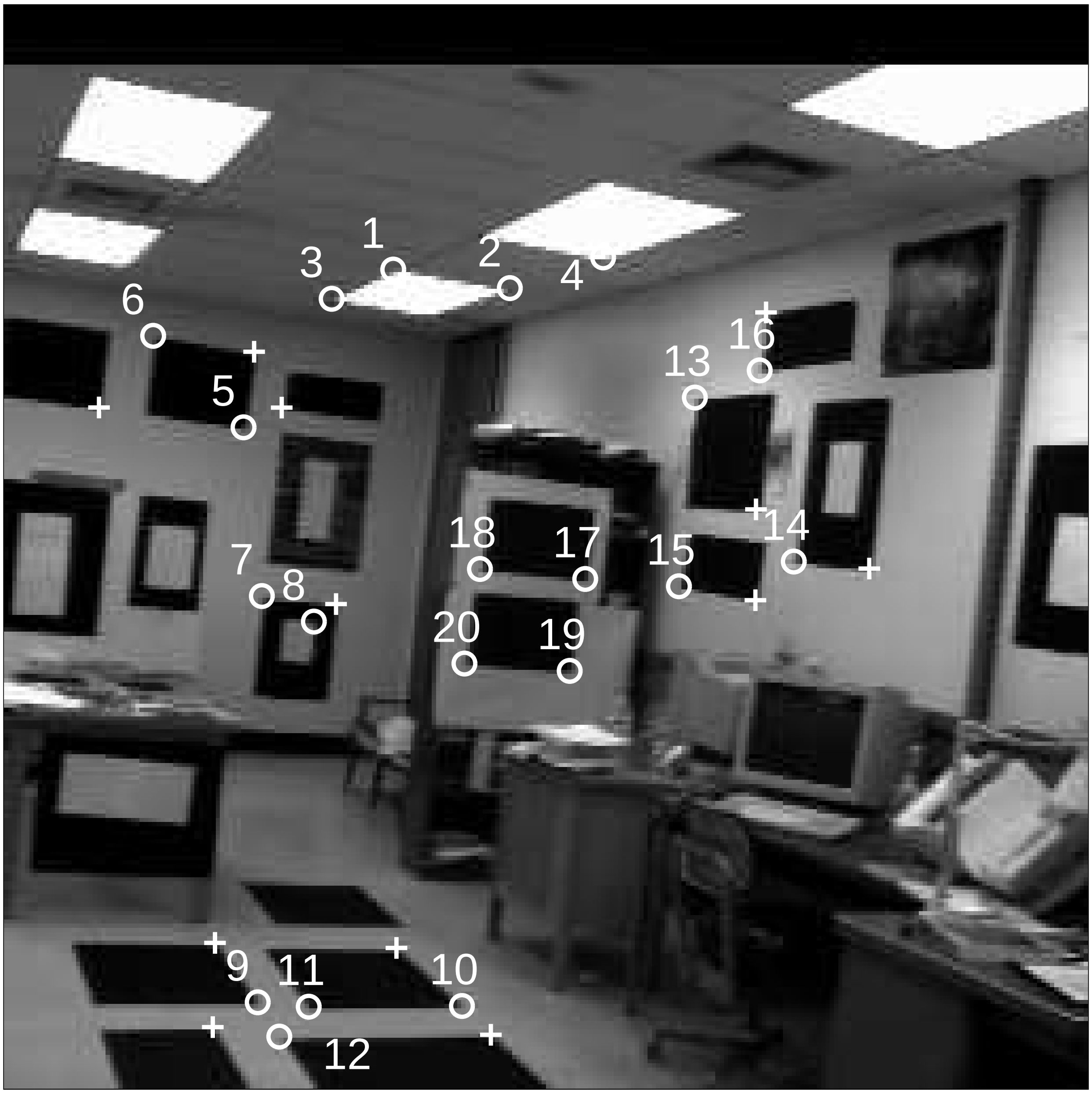


Figure 4.12: A typical image from Kumar's PUMA sequence.

plane coordinate system. It should be noted that there is considerable freedom in choosing local coordinates on both the object plane and the image plane when projective invariant methods are used. For example, any affine transformation of the coordinate system of either object or image plane can be tolerated, since the affine transformation matrix gets absorbed into the estimated homographic transformation matrix. In this experiment, model points for each wall were represented in the local 2D coordinate system described above, while points in the image plane were represented in row-column pixel coordinates. In contrast, Kumar's pose determination procedure required image coordinates to be specified in pinhole camera coordinates, which in turn required knowledge of the principle point (image center) and effective focal length. Indeed, the main concern of [Kumar90] was to analyze how errors in the specified principle point and focal length affect the final results of the model extension process. Such a discussion is moot when model extension is based on projective invariants.

For each of the 3 object planes used in this experiment, the 4 reference points on that plane and their corresponding image points (this is data supplied by the Kumar experiment) were used to estimate the homography that maps image plane points into object plane points by solving the linear system of equations in Section A.1.4. The estimated homography was then used to transfer the images of circled test points in the same plane back into the object plane, where their estimated positions were compared with the known ground truth locations. This process was repeated for each of the 20 images in the sequence.

Accuracy of the Reconstruction

The average absolute distance between reconstructed points on Plane 1 (points 5–8) and their ground truth locations on that plane was about 0.13 feet,[1] around the level of accuracy

[1] All reported distance measurements lie entirely within the object plane.

of the ground truth measurements themselves. Distance errors in points on Plane 2 (points 9–12) were consistently low, around 0.03 feet over all 20 images, and never more than 0.04 feet. This is because the chosen reference points are well spread out in that object plane, completely surrounding the estimated test points. On the other hand, errors in Plane 3 (points 13–16) were quite high: on average around 1.6 feet, and in some cases as much as 13 feet! This occurs because the reference points on that object plane do not adequately span the space of points to be estimated. In fact three of the reference points are very nearly colinear, a poor configuration for homography estimation.

To illustrate that the poor selection of reference points in Plane 3 is the cause of the large errors found, a new set of four reference points was chosen from the 8 points on that planar surface, namely points 13, 15, the point directly above 16, and the point directly to the right of 14. The process of model extension was rerun, using the other four points as test points. This time, the selected set of reference points is one of the better possible configurations, since the four points are well spread out and nearly surround the test points. The absolute distance errors reflect this fact – the average absolute error between predicted and actual object point locations is around 0.03 feet, as low as the errors for Plane 2. This new configuration of reference and test points will be referred to below as “Plane 3b”.

A rough comparison of the accuracy of model extension using planar homography estimation vs. full 3D model extension via pose determination and triangulation is possible for this sequence. Kumar reports average distance errors for each reconstructed point over 20 frames in both absolute terms and as a percentage of depth [Kumar90]. Table 4.5 condenses and summarizes his results for each of the three object planes considered in this experiment. For each plane, the average ground truth depth of points in that plane is shown (this stayed roughly constant across all 20 images), the average absolute error between reconstructed and

Table 4.5: Comparison of traditional vs. invariant model extension.

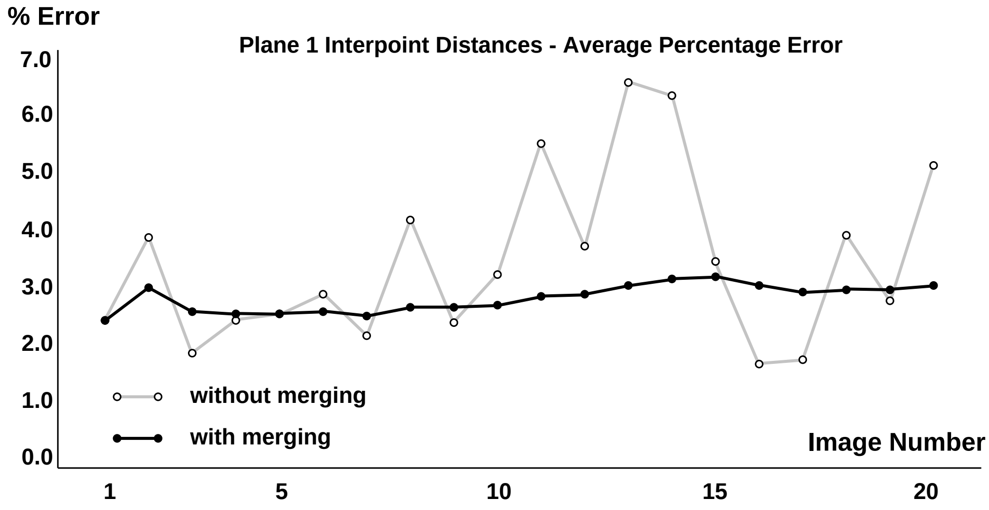
Plane	Avg Depth (feet)	**Pose + Triangulation** Avg Abs Error (feet)	Avg Pct Error	**Planar Invariance** Avg Abs Error (feet)	Avg Pct Error
plane 1	30.75	0.25	0.8 %	0.13	0.4 %
plane 2	14.41	0.13	0.9 %	0.03	0.2 %
plane 3	20.64	0.05	0.2 %	1.60	7.7 %
plane 3b	20.64			0.03	0.1 %

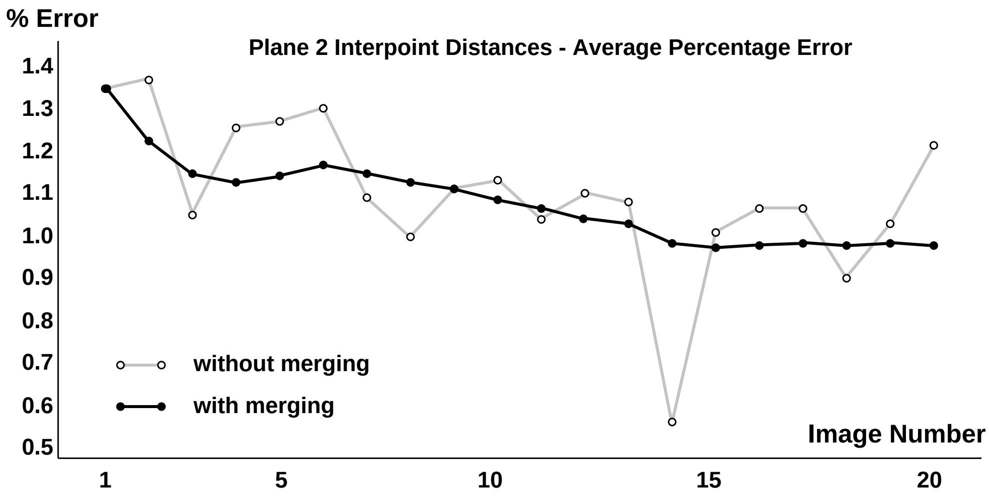
ground truth points, and the absolute percentage error of the reconstruction with respect to depth. Compared with this are the average absolute and percentage errors for points using planar invariant model extension. Accuracy results for both Plane 3 and Plane 3b using planar invariance are to be compared against the single result for Plane 3 using pose determination and triangulation. Aside from the results for Plane 3, where the reference points used to compute the projective image to object transformation were nearly singular, model extension based on planar invariance compares favorably with the 3D pose and triangulation algorithm.

It should be cautioned that such direct comparisions between two vastly different approaches is much like comparing apples and oranges. For instance, absolute errors in reconstructed points via planar invariance are distances measured solely within the given object plane, since the reconstructed point is constrained to lie in that plane, while residuals in 3D triangulated points can be arbitrary situated. Furthermore, reporting percentage error with respect to depth of scene points from the camera makes sense for reconstruction methods based on triangulation, which are notably less accurate as the depth increases. It is not clear that percentage error in depth makes any sense for methods based on projective invariants, which theoretically should be insensitive to depth. In practice, such methods will also be less accurate for planes located very far from the camera due to the effects of digitization on the possible resolution with which point features can be extracted, but difficulties of that kind should not be a problem in the relatively narrow range of plane depths exhibited here. Caveats aside, some tentative conclusions can be drawn from this case study. In places where planar model extension is applicable it yields reconstructions comparable to state-of-the-art triangulation methods. Care must be taken to select a well-conditioned set of reference points, however.

To explore the relative accuracy of the estimated point positions in a manner more appropriate to planar reconstructions, all distances between estimated points in a single object plane were compared to the ground truth distances, and percentage errors were computed. Figure 4.13 shows the results. The light grey curve for each plane shows the average percentage error taken over all point to point distances in that object plane for each image (the meaning of the dark grey curve will be explained shortly). Plane 2 (containing points 9–12) shows the most accurate results; the average percentage error over all images is 1.1%, around the level of noise in the ground truth measurements. Plane 1 (points 5–8) shows slightly worse errors, the average percentage error over all images is 3.5%. Originally plane 3 (points 13–16) shows markedly bad errors due to the ill-suited set of reference points chosen on that plane – the average error over all images is 51%, but for some images the average error is as high as 200%. However, using the new configuration of reference points devised for Plane 3b, the average percentage error drops to 0.9%.

Even in planes where points are estimated with high accuracy, the level of accuracy varies unpredictably from image to image. This is typical for a system that makes no use of previous estimates. In contrast, the smoother dark curve overlaid on each graph shows average percentage point to point distance errors when position estimates for previous images





Continued, next page

Figure 4.13: Average percentage point to point distance. These graphs show, for each object plane, the average percentage distance error between estimated interpoint distances and ground truth interpoint distances in that plane. The light grey curve shows percentage errors for each image plane computed separately. The dark grey curve shows the results of incremental data fusion performed up to and including image n, with $n = 1, \ldots, n$.

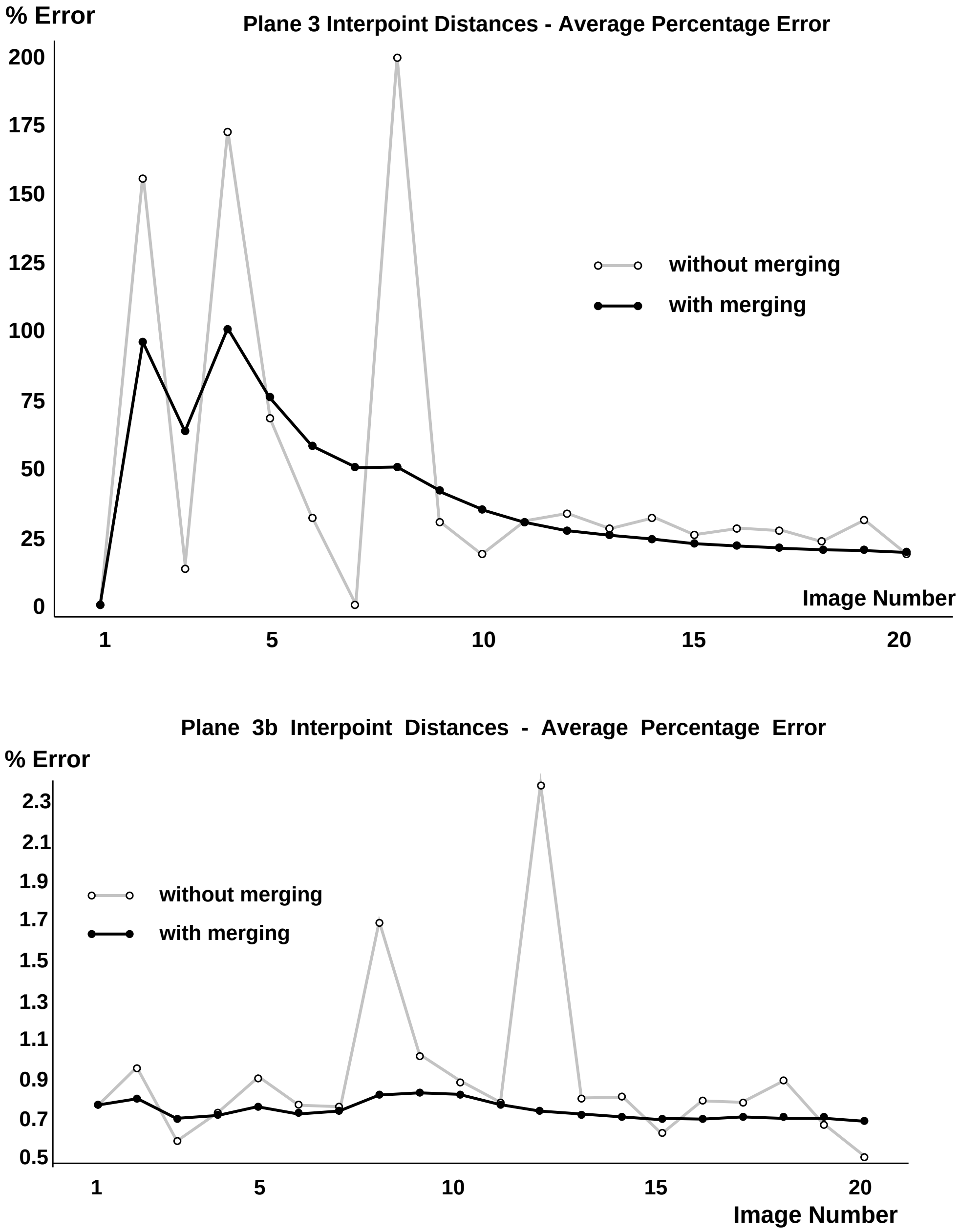


Figure 4.13: Continued.

are combined with current position estimates using the data fusion technique described in the next section.

Data Fusion

Planar projective model extension permits valid reconstructions to be derived from a single image taken by an uncalibrated camera with an unknown viewpoint. Over multiple images, multiple location estimates are obtained, and it can be reasonably assumed that combining these estimates will lead to a more accurate reconstruction. Each point location estimate in homogeneous coordinates represents a point in the projective plane; multiple location estimates for each object point form a sample of points in the projective plane, clustered around an axis representing the homogeneous coordinates of the true location.

The situation just described is clearly an example of Rank I incidence (Section 3.4), where multiple estimates of a single homogeneous coordinate vector are given and the goal is to estimate that vector. Each homogeneous coordinate vector is represented as a pair of antipodal points on the unit sphere. Due to noise, the given sample estimates form a bipolar cluster about an axis representing the true homogeneous coordinate vector. In Section 3.4.2 a Bayesian solution to this estimation problem was given. In the present case, no estimates of image feature errors are reported [Kumar90]. It was assumed, therefore, that all image points are equally valid, and this assumption was represented by setting the prior density function of each sample unit vector to be an identically weighted, rotationally symmetric Bingham density function with major axis oriented along the given sample vector. The MAP estimate in this case reduces to the eigenvector associated with the largest eigenvalue of the sample second moment matrix (see Section 3.4.2). This is also the maximum likelihood estimate under the Bingham structural model of Section 3.4.1.

The results of applying data fusion to this example are shown as dark curves in Figure 4.13. Although the fusion process smooths over random fluctuations in percentage error from frame to frame, it does not appear to increase the overall reconstruction accuracy by the end of the sequence. We hypothesize that this is because no provisions were made for dealing with errors in each estimated homography. Each homography matrix was assumed to be correct, but in fact was estimated from a minimal number of object to image point correspondences, and both the image point and object point measurements contain errors. These errors propagate into the parameters of the estimated transformation matrix, and provide a potentially large, unmodeled source of error in the estimate homogeneous coordinate vector of each object point. In this case "blind averaging" is not guaranteed to produce better overall results, and to do better it will be necessary to take into account errors in the estimated projective transformation [Thomas92].

CHAPTER 5

Conclusions and Future Work

Projective geometry is currently a topic of significant interest in computer vision. A representative sample of work in the field was presented at the first ARPA-ESPRIT workshop on *Applications of Invariance in Computer Vision,* and later collected into book form [Mundy92a]. In 1993, the prestigious Marr prize in computer vision was awarded for a paper firmly advocating a projective geometric approach to object recognition and structure recovery [Rothwell93]. Projective geometry has finally come out of the appendices and into the titles of important computer vision publications.

Many of the contributions of projective geometry to date have been in theoretical computer vision, answering fundamental questions such as how many points or lines must be observed to reconstruct a scene up to a scale factor, from how many views must they be seen, and are there special configurations of the points or lines that still yield ambiguous reconstructions [Buchanan92, Faugeras90, Tsai82, Maybank90a]. Other theoretical contributions have been made with regards to camera calibration [Beardsley92, Faugeras92b] and the recovery of projective structure from uncalibrated cameras [Faugeras92a, Hartley93, Mohr92, Shashua93]. Applications of projective invariant indexing for object recognition appear promising [Forsyth91, Rothwell93]. However, very few researchers have considered the effects of noise on the values of projective invariants such as the cross-ratio, and the concomitant errors introduced into the object indexing process and into the recovery of projective structure [Coelho92, Kanatani93b, Maybank93]. This thesis is one of the first bodies of work to directly address the combination of probability theory and projective geometry. Such investigations must be made before the mathematical theory of projective geometry can contribute significantly to successful and reliable computer vision applications.

5.1 Thesis Summary

A methodology for scene reconstruction has been presented that is based on the principles of projective geometry, while dealing with uncertainty at a fundamental level. Uncertainty in geometric features is represented and manipulated using probability density functions on projective space, allowing valid geometric constructions to be carried out via statistical inference.

The projective plane is topologically equivalent to the quotient space formed by the surface of the Euclidean unit sphere modulo the set of infinite lines through the sphere's center, which amounts to equating antipodal pairs of points on the surface of the sphere. Associating the projective plane with this quotient space introduces a metric that allows antipodally symmetric probability density functions on the sphere to be reinterpreted as density functions in the projective plane. An exponential family of probability density functions on the sphere was considered, and Bingham's distribution was chosen to represent uncertainty in the projective plane.

The main contribution of this thesis is the development of *stochastic projective geometry*, a formalism for performing uncertain geometric reasoning during the scene reconstruction process. The homogeneous coordinates of points and lines in the projective plane are represented by antipodal pairs of points on the sphere, and geometric uncertainty in their location is represented using Bingham's distribution. Geometric reasoning about homogeneous coordinate vectors reduces to well-defined manipulations on probability density functions. The result is an uncertainty calculus in projective space analogous to the Gaussian uncertainty calculus in affine space. The main strengh of the Gaussian calculus is maintained, namely its uniform treatment of uncertainty in all stages of the geometric reasoning process. At the same time, the limitations of the Gaussian density function as a representation of uncertainty in projective space are removed.

There are two fundamental types of operations that any uncertain geometric reasoning system should address. The first is determining how a given representation of uncertainty transforms under a change of coordinates . The second involves *fusing* multiple estimates or observations to yield more accurate geometric descriptions. The coordinate transformations of interest in plane projective geometry are the nonsingular planar projective transformations, or homographies. Although they are linear in homogeneous coordinates, homographies induce a nonlinear bijective transformation on the surface of the sphere. Using a method called *transformation of variables*, the change in point and line uncertainty due to a projective coordinate transformation was determined by propagating the Bingham probability density function through this nonlinear transformation.

Bingham's distribution does not remain invariant under projective transformations. In order to maintain a uniform representation of uncertainty, methods for approximating the resulting density function with a Bingham density were explored. Three approaches to finding an approximating Bingham density were considered: maximum likelihood approximation, Taylor series expansion around the mode, and propagation of 3D singular covariance matrices. Comparisons on the basis of mean-squared error showed singular covariance propagation to be more accurate over a wider range of transformations. Singular covariance propagation was then applied to the problem of initializing Bingham uncertainty measures by mapping Gaussian representations of uncertain points and lines in the image plane onto the sphere via central projection.

A Bayesian approach to evidence combination on the sphere was presented for fusing noisy homogeneous coordinate observations constrained by known projective incidence relations. Rank I incidence is used when fusing multiple observations of the same point or line in the projective plane. Homogeneous coordinate vectors related by Rank I incidence are represented by the same pair of antipodal points on the sphere. The associated estimation problem involves computing the axis of a bipolar cluster of points. Rank II incidence is relevant for estimating the homogeneous coordinates of a line formed by colinear points, or of a point formed by intersecting lines. Homogeneous coordinates related by Rank II incidence are points separated by an arc distance of $\pi/2$ on the surface of the sphere. The associated estimation problem involves computing the polar axis of a great circle of points clustered on the sphere.

The effectiveness of stochastic projective geometry for dealing with noisy projective relationships was demonstrated on three geometric problems: deriving line and plane orientations using vanishing point analysis, partitioning scene features into planar patches using line correspondence stereo, and extending a partial model of planar surface structure using projective invariants. Vanishing point analysis is a correspondenceless approach for deriving 3D line and plane orientations from a single monocular view of sets of parallel lines in the world. An efficient histogram-based method on the sphere was implemented for clustering line segments from an image into converging line pencils, and three statistical estimators for the homogeneous coordinates of their point of intersection were presented and compared. In an experiment run on aerial photographs, line orientations were derived to with an average absolute error of 2–3 degrees with respect to "ground truth" orientations that were computed by a multi-image camera resection algorithm using over 100 ground control points.

As a second application, a hierarchical method was developed for partitioning line segments extracted from a stereo pair of images into sets that are coplanar in the scene. Given a set of image-to-image line segment correspondences from a translational stereo pair, line orientation can be computed directly without first determining the depth of features in the scene. The same histogramming method used to find convergent line segment pencils for vanishing point analysis can be applied to cluster the resulting 3D line orientations into sets consistent with lying on a family of parallel planes, and a statistical estimate of the shared surface orientation of those planes can be computed. Each set of lines consistent with a family of parallel planes is then partitioned again into sets consistent with a single plane equation, using knowledge of the stereo baseline to compute depth. The feasibility of this approach was demonstrated on a single stereo pair taken from the stereo literature. Three planar structures in the scene were recovered, two of which share the same surface normal but are at different depths. Surface normals were computed with an estimated accuracy of roughly 2.5 degrees (radius of 95% confidence cone around the computed normal). The relative orientation between the two surface normals, which should in fact be perpendicular, was computed as 89.4 degrees.

Finally, a method for planar model extension based on homography estimation was presented that allows valid planar reconstructions to be derived from a single image taken by an uncalibrated camera with an unknown viewpoint. This method was tested on a sequence of images used in the pose literature, and shown to yield reconstructions comparable to state-of-the-art triangulation methods, i.e. reconstruction errors of around 1% with respect to the distance of scene points from the camera. It was also demonstrated that a well-conditioned set of basis points is essential for accurate projective reconstructions. Data fusion over multiple images did not significantly improve the overall accuracy of the results. It was hypothesized that this is due to unmodeled errors in the estimated homography matrix parameters, and that to do better will require representation and propagation of these transformation errors.

5.2 Future Work

The research reported here is only a beginning. Three possible directions in which this work could be extended are discussed below.

5.2.1 Robustness

The analysis of uncertainty on the sphere and the projective plane using Bingham's distribution bears similarity to the least-squares method of error analysis in Euclidean space. The similarity arises due to the form of Bingham's distribution (the exponential of a second-order polynomial), so that, like the Gaussian, sufficient statistics for the distribution are provided by the second-order sample moments. Eigenvectors of the sample second moment matrix turn out to be maximum likelihood estimates for the axes of the distribution. But these eigenvectors are also the orientations of lines and planes through the origin that minimize the sum of squares of orthogonal distances to the observed sample points on the sphere

[Pearson01]. Maximum likelihood estimation of orientations under the Bingham assumption on the sphere is thus closely related to orthogonal least-squares line and plane fitting in R^3. Even the more sophisticated Bayesian analysis of Section 3.4 reduces in special cases to weighted, orthogonal least-squares regression of lines and planes to a set of sample points.

It is well-known that least-squares techniques perform badly in the presence of gross errors or *outliers* [Harter83, Kumar92]. Outliers can be intuitively defined as observations that disagree so much from the bulk of the observed data that they are highly unlikely to have arisen from the same process. For example, when observation errors are assumed to be Gaussian any observation more than three standard deviations away from the sample mean is likely to be an outlier. Robust statistics are designed to be relatively unaffected by the presence of outliers in the data.

To some extent, the vanishing point and stereo applications presented in Chapter 4 are already robust due to the Hough transform method used to detect sets of intersecting great circles on the sphere. This method was described in Section 4.1.1. Recall that each great circle is entered into a set of histogram buckets on the sphere. Intersecting great circles are detected as peaks in the histogram. To retrieve all great circles potentially contributing to the point of intersection, a sample mean intersection $\hat{\boldsymbol{u}}$ is computed from all the great circles contained in the peak bucket. All buckets neighboring the peak are then searched for great circles lying within a predetermined error radius ρ of the sample mean. As a result, all great circles selected as candidates for intersection are known to pass through a circular area of radius ρ on the sphere. Great circles that deviate notably from the proposed intersection site are thus filtered out at the very start.

This approach can be related to robust techniques that attempt to filter out incorrect data points. One such technique is median filtering [Kumar92]. In median filtering, robust estimators of sample mean and standard deviation are first applied to the input data. All sample points lying farther away than three standard deviations from the computed mean are then discarded as outliers. Finally, the remaining data points are treated as an outlier-free sample using ordinary least-squares statistical techniques. The histogram-based method of great circle cluster selection used in Chapter 4 would be virtually identical to median filtering if only the error radius within which surrounding buckets are searched was computed automatically from the data rather than predetermined by the user.

Very little has been written about robust statistical estimators for directional data. Two papers that have appeared only consider angular data – that is, data points on the circle [Lenth81, Stein92]. Lenth presents M-estimators for directional data that minimize error functions that are more resistant to outliers than the sum of squares error function. Stein and Werman consider a number of robust estimates based on the circular median, in the context of robust line and conic fitting. They define the circular median as any point ξ on the unit circle such that half of the data points are on either side of the diameter from ξ to $\xi + \pi$, and the majority of the points are nearer to ξ than to $\xi + \pi$. Unfortunately, it is not readily apparent how this definition of circular median can be generalized for use with directional data on the sphere and hypersphere. We propose here a method for computing the median of a set of directional unit vectors $\boldsymbol{v}_i$ in any dimension. A well-known estimator for the sample mean in this case is provided by $\sum \boldsymbol{v}_i / \| \sum \boldsymbol{v}_i \|$, which is the normalized vector resultant [Watson83]. In three dimensions, for example, the resulting unit vector is

$$\left(\frac{\sum x_i}{R}, \frac{\sum y_i}{R}, \frac{\sum z_i}{R} \right) \quad \text{where} \quad R = \sqrt{(\sum x_i)^2 + (\sum y_i)^2 + (\sum z_i)^2}$$

which can be rewritten as

$$(\bar{X}, \bar{Y}, \bar{Z}) \; / \; \sqrt{\bar{X}^2 + \bar{Y}^2 + \bar{Z}^2} \tag{5.1}$$

where $\bar{X}$, $\bar{Y}$, and $\bar{Z}$ are the sample means of the sets $\{x_i\}$, $\{y_i\}$, and $\{z_i\}$ of vector components. The proposed sample median vector is obtained simply by replacing each sample component mean in equation (5.1) by the component median, i.e. replacing $\bar{X}$ by $median\{x_i\}$, and so on. This definition also has limitations, in that it only works for signed directional vectors. Robust estimation of axial data, particularly under the assumption of Bingham-distributed errors, will ultimately require investigation into techniques for robust eigenvector analysis, or equivalently, robust methods for orthogonal least-squares regression.

Any proposed estimator should first be subjected to rigorous statistical testing to determine its variance, bias, and degree of robustness to outliers. Direct application of routine statistical tests to circular and spherical data should be approached with caution, however. Stein and Werman point out that standard robust estimation concepts such as "breakdown point" do not generalize well to the case of directional data. The breakdown point of an estimator is the percentage of gross errors that can be tolerated before the difference between the estimated value and the true value becomes arbitrarily large – i.e. before the estimator becomes totally unreliable. For example, the mean of a random sample has a 0% breakdown point because only one outlier at infinity can shift the sample mean to infinity, while the median has a breakdown point of 50% it can tolerate up to half of the data points being shifted to infinity before it too becomes infinite. However, outliers in directional data can only be off by 180 degrees for signed directions; and only 90 degrees in the case of unsigned, axial data. Therefore, no amount or severity of outliers can cause an orientation estimate to be "arbitrarily" incorrect. Clearly, the development of new methods for robust estimation of directional data, particularly methods that generalize well to higher dimensions, is an open area for future research.

5.2.2 Extensions to Other Manifolds

Bingham's distribution on the unit sphere S^2 has been used in this thesis as a probability density function over points and lines in the projective plane P^2. Bingham's distribution is readily generalized to P^n. A projective point (and its dual in projective n-space) is represented by an antipodal pair of points on the surface of the unit hypersphere S^n. Bingham's distribution on S^n (and hence P^n) is formed as the intersection of an (n+1)-dimensional zero-mean Gaussian with the surface S^n, suitably renormalized so that the total probability density over S^n integrates to one. Bingham's distribution on the circle S^2 represents a probability density function over points in the projective line P^1. Bingham's distribution on the hypersphere S^3 represents a probability density function over points and planes in projective three-space P^3. A rigid rotation in $SO(3)$, when represented as a unit quaternion, can be visualized as an antipodal pair of points on the unit hypersphere S^3 [Horn86]. Therefore Bingham's distribution on S^3 is also a probability density function over the space of 3D rotations.

Probability distributions on the hypercylinder $S^m \times R^n$ can be formed as the direct product of an m-dimensional Bingham distribution with an n-dimensional Gaussian [Bagchi91]. Some hypercylinders that are useful in computer vision are the Euclidean line manifold

$S^1 \times R^1$, the Euclidan plane manifold $S^2 \times R^1$, and the rigid motion or pose manifold $S^3 \times R^3$. Extensions to the hypertorus $S^m \times S^n$ are formed similarly.

5.2.3 Towards Nonplanar Model Extension

One of the applications considered in Chapter 4 was planar model extension. Given knowledge of at least four coplanar points or lines in the scene and their correspondences in an image, a planar homography is estimated that maps new points and lines from the image into their proper locations on the object plane without first computing pose or calibrating the camera.

When other, non-coplanar scene features are present, 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 at least four coplanar points or lines are available for use as a *reference plane,* and that an image-to-image homography has been computed that transforms the projections of reference plane point features from image one into their corresponding projections in image two. What can be said about the transformed image of a scene point outside of the reference plane? The location predicted for it by the planar homography will not in general coincide with the actual location of the projected scene point in image two. However, the residual difference between a the predicted and actual positions of the projected point in image two are constrained to lie along lines intersecting in a single point. These lines are called epipolar lines, and they intersect at the epipole, which geometrically corresponds to the image of the focal point of the camera when image one was taken. Furthermore, the direction of each residual difference vector either towards or away from the epipole determines whether the corresponding 3D scene point lies either forward or behind the plane of reference. These results are valid regardless of either camera's location, orientation, or calibration parameters, and have been noted previously [Demey92, Luong93, Shashua93].

For illustration, two aerial photographs are shown in Figures 5.1a and 5.1b. Figures 5.1c and 5.1d show 37 corresponding pairs of points that were chosen by hand from these images. 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 5.2 shows residual 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. All remaining residual vectors lie along infinite epipolar lines that intersect at a single epipolar point, which in this case is far off the image. Furthermore, 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 property holds in general, and can be used to qualitatively partition scene points into three categories depending on the orientation of their residuals: those lying closer to the viewer than the reference plane, those lying on the plane (difference vector is zero), and those lying further away.

Accurately recovering the position of the epipole is an important step in recovering scene structure from an uncalibrated stereo pair. Statistical estimation of the epipole should be a straightforward application of the same estimation techniques used in this thesis to precisely locate vanishing points from noisy line segments. A more detailed analysis of

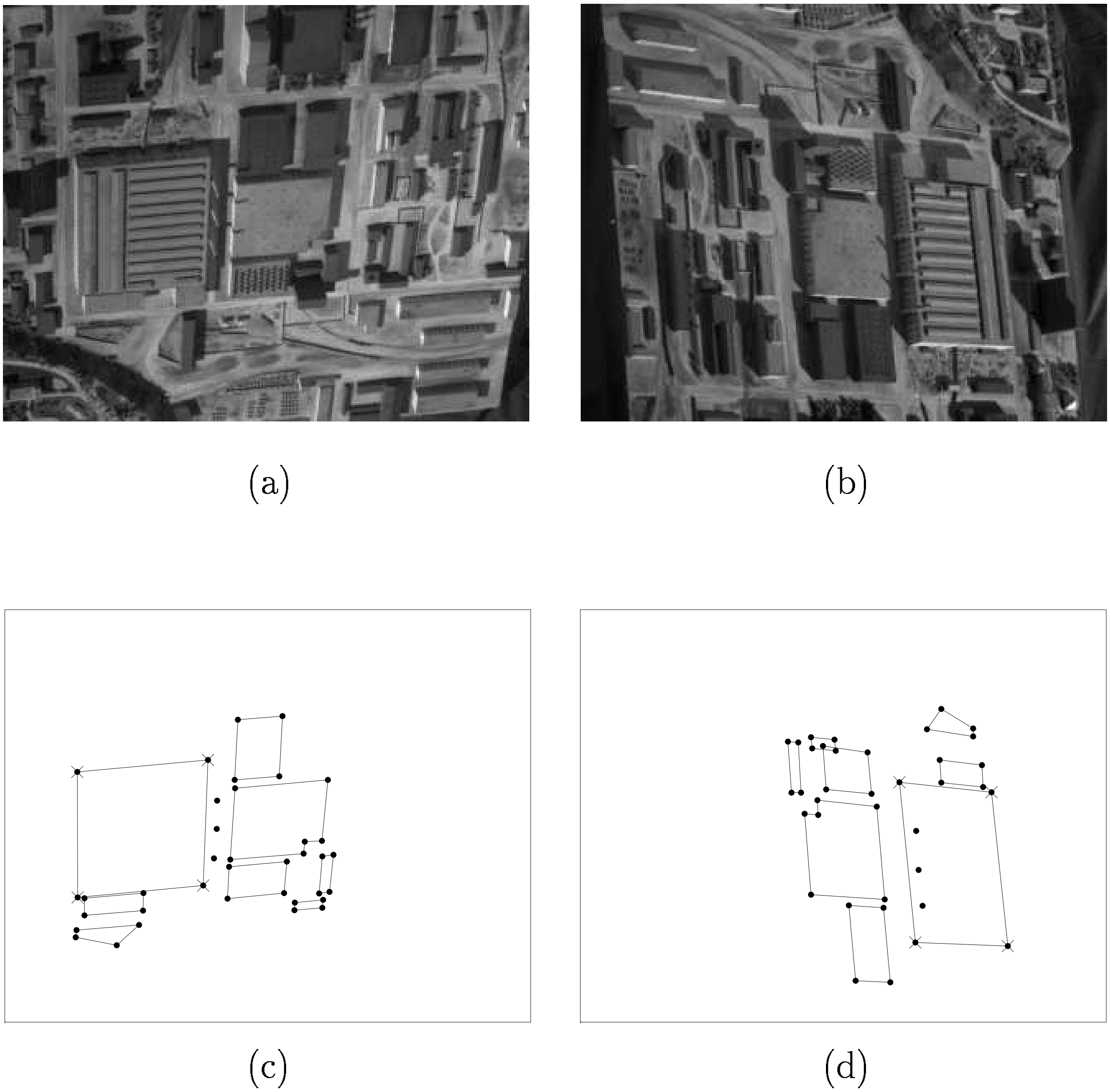


Figure 5.1: Two aerial photographs that are not well-approximated by a single plane. (a) Radius Model Board 1, Image J8. (b) Radius Model Board 1, Image J2. (c) Interesting points extracted by hand from Image J8. (d) Corresponding points extracted by hand from Image J2. Some building boundaries have been added for clarity. Crosses mark points that will be used to estimate a homography between the two images.

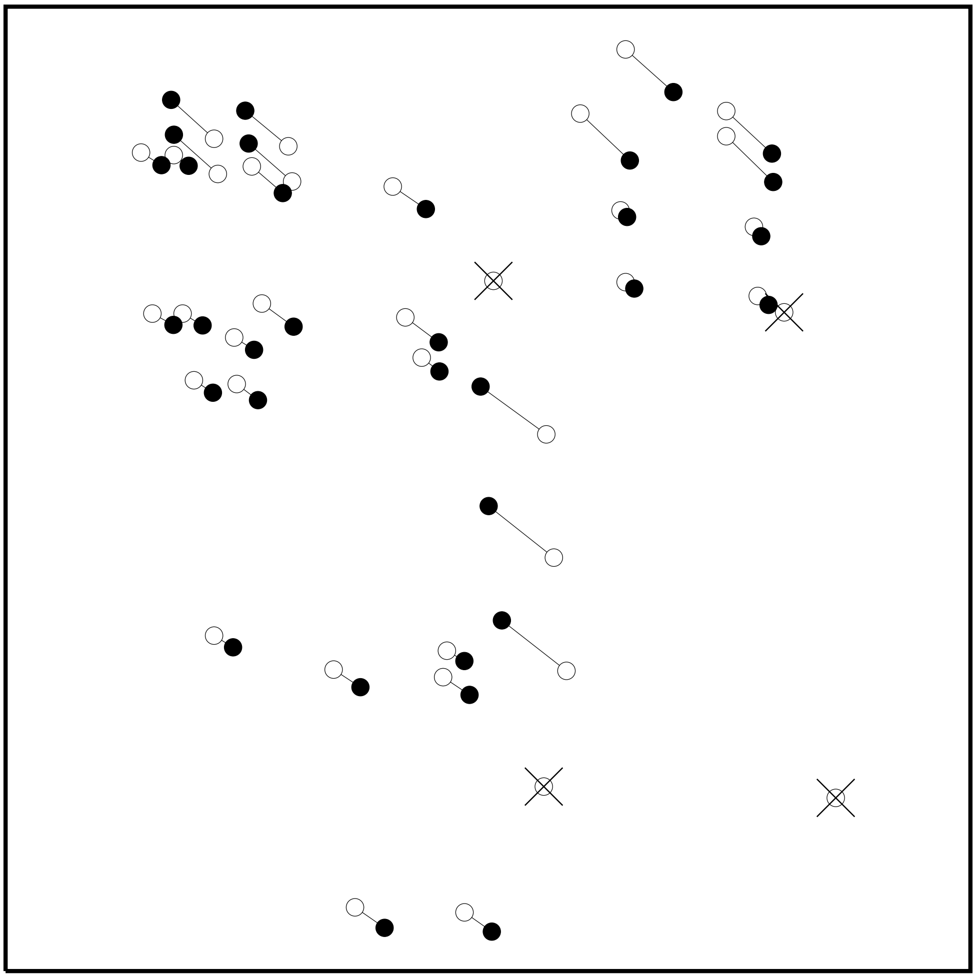


Figure 5.2: Residual difference vectors between predicted and actual locations. Black dots mark the predicted locations of points from image one transformed by a planar homography into image two. White dots mark the actual locations where the points were found . Points marked with crosses were used to define the homography.

error propagation in the recovery of projective structure from uncalibrated stereo will be the subject of future research.

APPENDIX A

Projective Geometry Primer

A.1 Projective Transformations

What is projective geometry? There are several answers, reflecting the many different ways of approaching this vast subject. The classic axiomatic approach postulates the existence of points, lines, planes and a binary incidence relation between them, then systematically explores the logical conclusions of a short list of axioms. The most notable axiom states that two distinct lines in a plane always intersect in a point, therefore there is no notion of parallelism in projective geometry, as there is in affine and Euclidean geometry. The rigorous axiomatic approach has led to the study of projective spaces with a finite number of elements, and to the study of projective spaces with coordinates over arbitrary division rings.

We will be concerned with continuous, real projective spaces of dimension one (the projective line), two (the projective plane), and three (projective 3-space). A feel for these spaces can be gained by comparision with their affine and Euclidean analogs. For example, the projective plane can be thought of as an affine plane with extra points added to represent the intersections of parallel lines. Alternatively, upon the introduction of homogeneous coordinates, points and lines in the projective plane can be identified with lines and planes through the origin of R^3. The analysis of incidence relations between projective elements then boils down to questions involving the dependence or independence of linear subspaces, and the techniques of linear algebra become available.

Perhaps the simplest description of projective geometry was given by Felix Klein in his famous Erlanger Programm. Klein classified geometry as the study of properties of point sets that remain invariant under some group of transformations. At one end of the spectrum is classical Euclidean geometry, studying properties such as length and angle that remain invariant under the Euclidean group of rotations, translations and reflections. At the opposite end is topology, the study of properties such as connectedness that remain invariant under homeomorphisms, or "rubber-sheet" deformations. Projective geometry, then, can be characterized as the study of properties that remain invariant under projective transformations.

A.1.1 Perspectivities

The canonical example of a projective transformation is central projection, also known as a *perspectivity.* Figure A.1 shows a perspectivity between two distinct lines L and M in the affine plane. (The affine plane can be thought of as the familiar Euclidean plane as long as metric notions such as length and angle are ignored. Parallelism is still defined, however.) Under a perspectivity, corresponding pairs of points are joined by *projection lines* that all pass through a single, distinct point C called the *perspective center.* Viewed as a mapping from line L to line M, the *image* of point l is the point m where the projection line C-l intersects M, and likewise the *preimage* of m is the intersection of projection line C-m with line L. This mapping is one-to-one except for one point on L that has no image because its projection line is parallel to M. Likewise, the mapping is onto except for one point on M that has no preimage because its projection line is parallel to L.

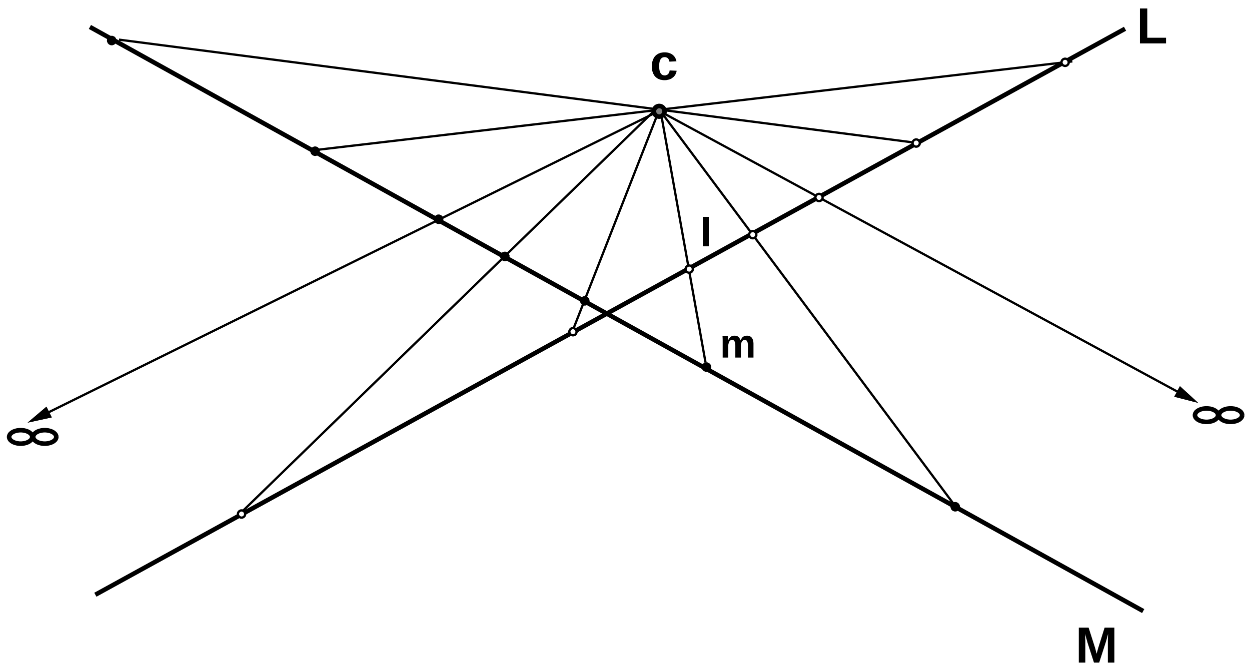


Figure A.1: A perspectivity between two lines in a plane.

It is easier mathematically to deal with mappings that are both one-to-one and onto, called *bijections*. In order to make a perspectivity between two lines bijective, points at infinity are explicitly added to lines L and M. The image of the point on L having a projection line parallel to M is defined to be the point at infinity on line M. Similarly, the preimage of the point on M whose projection line is parallel to L is said to be the point at infinity on line L. With these additions, every point on L has an image on M and every point of M has a preimage on L, and thus the mapping is bijective. When an affine line is augmented in this way with a single point at infinity, a new geometric entity called the *projective line* is created. The projective line has a different global topology than the affine line, and this has profound implications for the representation of observed points and their uncertainty, as discussed in Section 1.2 of this thesis.

A perspectivity between two distinct planes in space is defined in a manner analogous to that for lines. Given two distinct planes P and Q in affine space, and a point C not on either, point p from plane P is mapped to the point q where projection line C-p intersects plane Q (Figure A.2). As in the one dimensional case, some points on P don't have any corresponding images in q, because their projection lines are parallel to plane Q. In fact, the locus of points in P having no image is a line, and likewise a whole line of points in plane Q have no preimage. Once again, a bijective mapping can be arranged by adding points at infinity to the configuration. One infinite point is added to plane Q to correspond to each point in P having no image, and similarly one point at infinity is added to P for each point in Q having no preimage. The locus of infinite points added to each plane forms a line at infinity. When an affine plane is augmented with a line of points at infinity, it becomes a new geometric entity called the *projective plane.* Parallel lines in the affine plane intersect in a point at infinity in the projective plane, so the notion of parallelism no longer holds any special significance.

A perspectivity between two 3D-spaces in affine 4-space can also be defined, although it is easier to visualize this as a mapping from projective 3-space into itself. *Projective space* is formed by adding points at infinity to affine space. For each family of parallel lines in affine space, a point at infinity is added to represent their intersection. Each family of parallel planes in affine space intersect in a line of points at infinity. The totality of points at infinity added to affine space forms a plane at infinity in projective space.

A.1.2 Projectivities

Perspectivity is just one form of projective transformation. The most general type of projective mapping, called a *projectivity*, is formed as a sequence of perspectivities. An example appears in Figure A.3, which depicts a perspective mapping from line L to a second

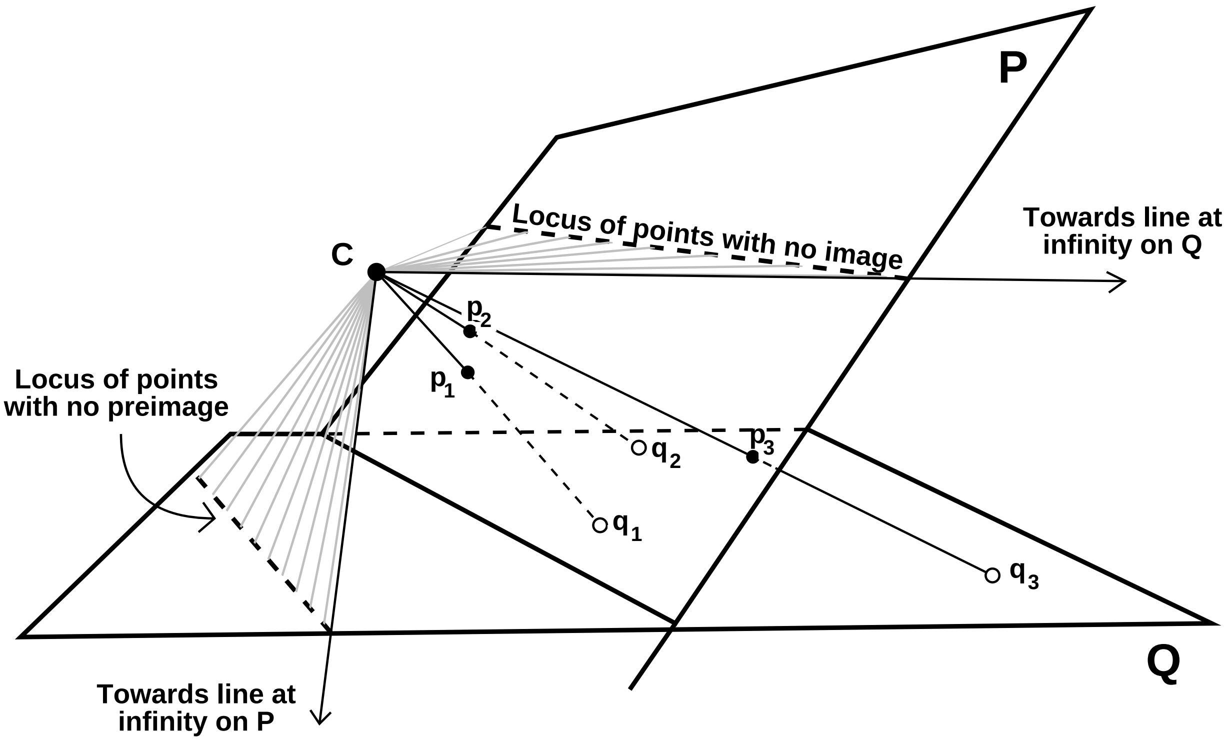


Figure A.2: A perspectivity between two planes.

line M, followed by a perspective mapping from line M to a third line N. The resulting map from line L to line N is not in general a perspectivity, because lines through corresponding points no longer converge to a perspective center.

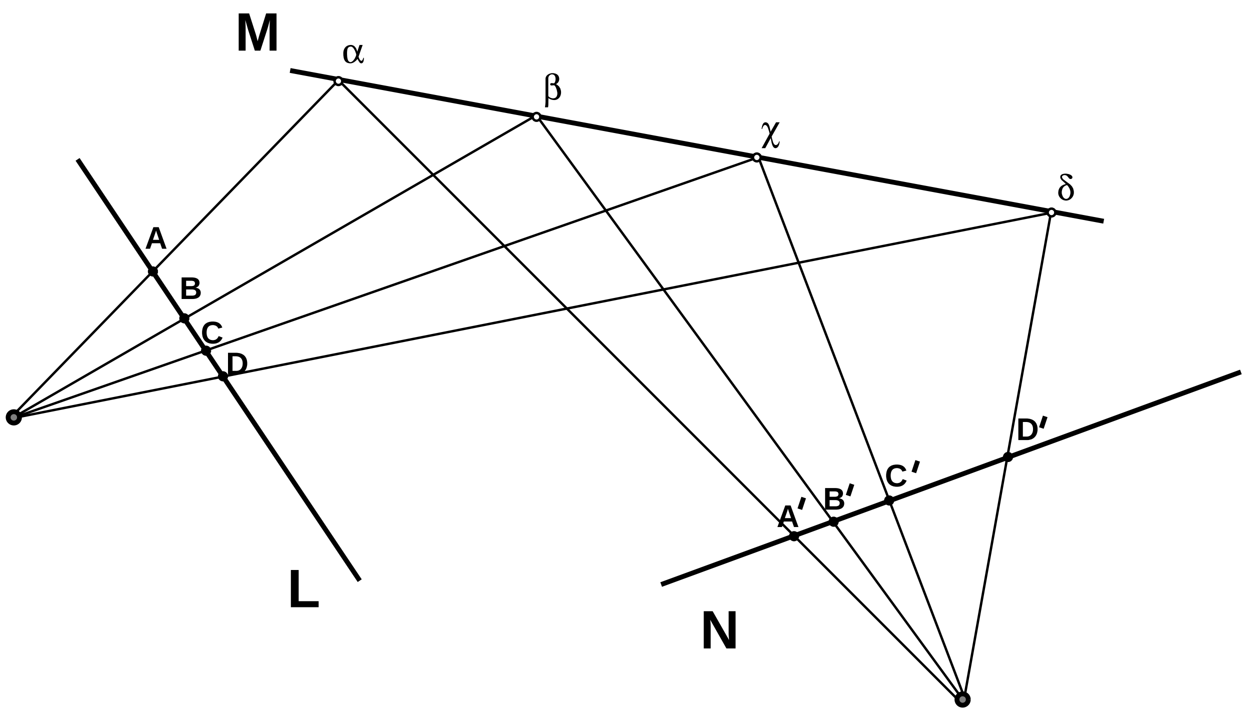


Figure A.3: A projectivity is a sequence of two or more perspectivities.

General projectivities are defined algebraically as linear transformations in homogeneous coordinates. Homogeneous coordinates represent a point in projective n-space by an $(n+1)$-dimensional coordinate vector, the extra component being used to represent points at infinity. In the projective plane, for instance, the homogeneous coordinate vector (X, Y, S) represents affine plane coordinates $(X/S\ ,\ Y/S)$ when $S \neq 0$, but represents a point at infinity when $S = 0$. The homogeneous coordinates of a point can be scaled by any nonzero number and still represent the same point. More formally, coordinate vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ are equivalent, written $\boldsymbol{p} \sim \boldsymbol{q}$, if and only if $\boldsymbol{p} = k\boldsymbol{q}$ for $k \neq 0$.

In homogeneous coordinates, a projectivity between two lines is defined as

$$\begin{bmatrix} X' \\ S' \end{bmatrix} \sim \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} X \\ S \end{bmatrix}, \tag{A.1}$$

and between two planes as

$$\begin{bmatrix} X' \\ Y' \\ S' \end{bmatrix} \sim \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} X \\ Y \\ S \end{bmatrix}. \tag{A.2}$$

Since the equivalence relation $\sim$ is used rather than strict equality, these equations are statements about equivalence classes rather than specific vector results. Any such matrix relation $\boldsymbol{p}' \sim \boldsymbol{H}\boldsymbol{p}$ can be rewritten using equality by introducing a nonzero scalar variable k and writing $\boldsymbol{p}' = k\boldsymbol{H}\boldsymbol{p}$. Since homogeneous coordinates are equivalent up to scalar multiples, each matrix of transformation parameters can be multiplied by any nonzero constant and still represent the same mapping. Matrices representing projective transformations are therefore equivalent up to scalar multiples also, so that a projectivity between lines has only 3 independent parameters, a projectivity between planes has 8, and a projective transformation of 3-space has 15.

A projectivity is nonsingular, and therefore invertible, if and only if the matrix of parameters representing it is square and has non-zero determinant. Only projective transformations between spaces of the same dimension can fill these requirements. An important example of a projectivity between spaces of different dimension is

$$\begin{bmatrix} X' \\ Y' \\ S' \end{bmatrix} \sim \begin{bmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ S \end{bmatrix}, \tag{A.3}$$

which projects points in 3D space onto a plane. Since a whole line of points in space is mapped onto each point in the plane, this mapping is obviously non-invertible.

A.1.3 Homographies

A nonsingular projectivity, represented by a nonsingular square matrix in homogeneous coordinates, is called a *homography*. Functional composition of two homographies is performed by multiplying their corresponding matrices. The determinant of this product will also be nonzero, thus the set of all homographies is closed under function composition. The composition of any homography with a scalar multiple of the identity matrix performs the same projective transformation as the original homography. The inverse of a homography always exists, and can be found by inverting the nonsingular matrix representing the homography. These conditions verify that the matrices representing homographies form a mathematical structure called a *group*, under the operations of matrix multiplication and matrix inverse. Group theory is a powerful branch of mathematics with a rich set of results that can be applied to the study of transformations and their invariants [Tsai82, Forsyth91].

A set of matrices form a subgroup of the homography group if they all have some identifying characteristic that remains unchanged under matrix multiplication. The projective transformations they represent then form a functionally closed subgroup of transformations, and according to Klein define a subgeometry. Two important subgroups of the homography group are the affine group and the proper Euclidean group. An n-dimensional affine transformation can be written as an $n \times n$ matrix with a 1 in the bottom, rightmost position, and zeros everywhere else in the bottom row, e.g.

$$\begin{bmatrix} X' \\ Y' \\ S' \end{bmatrix} \sim \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ S \end{bmatrix} . \tag{A.4}$$

represents an affine transformation in the plane. It is easily verified that the composition of two affine matrices yields a third. Under an affine transformation, points at infinity ($S = 0$) only map to other points at infinity, and finite points ($S \neq 0$) stay finite. Parallel lines, defined as lines having an intersection point at infinity, therefore remain parallel under affine transformations. Affine geometry is concerned with properties of figures, such as parallelism, that remain invariant under affine transformations.

The proper Euclidean group is a subgroup of the affine group consisting only of rigid motions (rotations and translations, no reflections). A rigid motion in 3-space can be written as

$$\begin{bmatrix} X' \\ Y' \\ Z' \\ S' \end{bmatrix} \sim \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ S \end{bmatrix} , \tag{A.5}$$

where the upper left submatrix $\{r_{ij}\}$ is an orthonormal rotation matrix, and $\{t_i\}$ is an arbitrary vector of translation parameters. Euclidean motions preserve distances, and therefore preserve angles and parallelism as well.

Figure A.4 illustrates the effects of the eight independent plane to plane homography parameters in isolation. Given an initial image of a unit square, centered about the origin, the three proper Euclidean group parameters rotate the square and translate it about the plane. The four-parameter affine or *similarity* group adds an extra scale factor that allows the figure to shrink or expand. These mappings all preserve the shape of the image, that is, the figure always remains square. The general affine group includes two more parameters that can stretch and skew the square into a rectangle or parallelogram. Although the shape of the square is not preserved under general affine transformations, opposite sides of the figure remain parallel. The final two parameters of the planar homography group allow the line at infinity to be mapped into the finite portion of the plane. The intersection points of opposite sides of the resulting figure lie on this line, thus opposite sides of the figure are no longer parallel and the square image is transformed into a quadrilateral.

A.1.4 Estimating Projective Transformations

Since the parameters of an arbitrary plane to plane homography can be characterized in terms of their effects on the image of a unit square (Figure A.4), it seems reasonable to expect that these parameters could be recovered from a careful examination of the before and after images of a square or quadrilateral. The fundamental theorem of projective geometry makes this intuitive notion more precise. It states that a plane to plane homography is uniquely determined by the correspondences of four points, no three of which are colinear. That four is the minimum number of independent points required to estimate the parameters of a planar homography can also be arrived at by a simple analysis of degrees of freedom. A planar homography has 8 unknown parameters. The position of a point in the plane has two degrees of freedom, therefore each point correspondence provides two equations in the eight unknowns. As a result, four points are needed to provide eight equations in eight unknowns. The condition that no three of the four points can lie in a line is necessary to ensure that all eight equations are in fact independent. This simple analysis can be used to extend the fundamental theorem to projective spaces of other dimension. It can then be shown that at least three distinct point correspondences are necessary to estimate a line

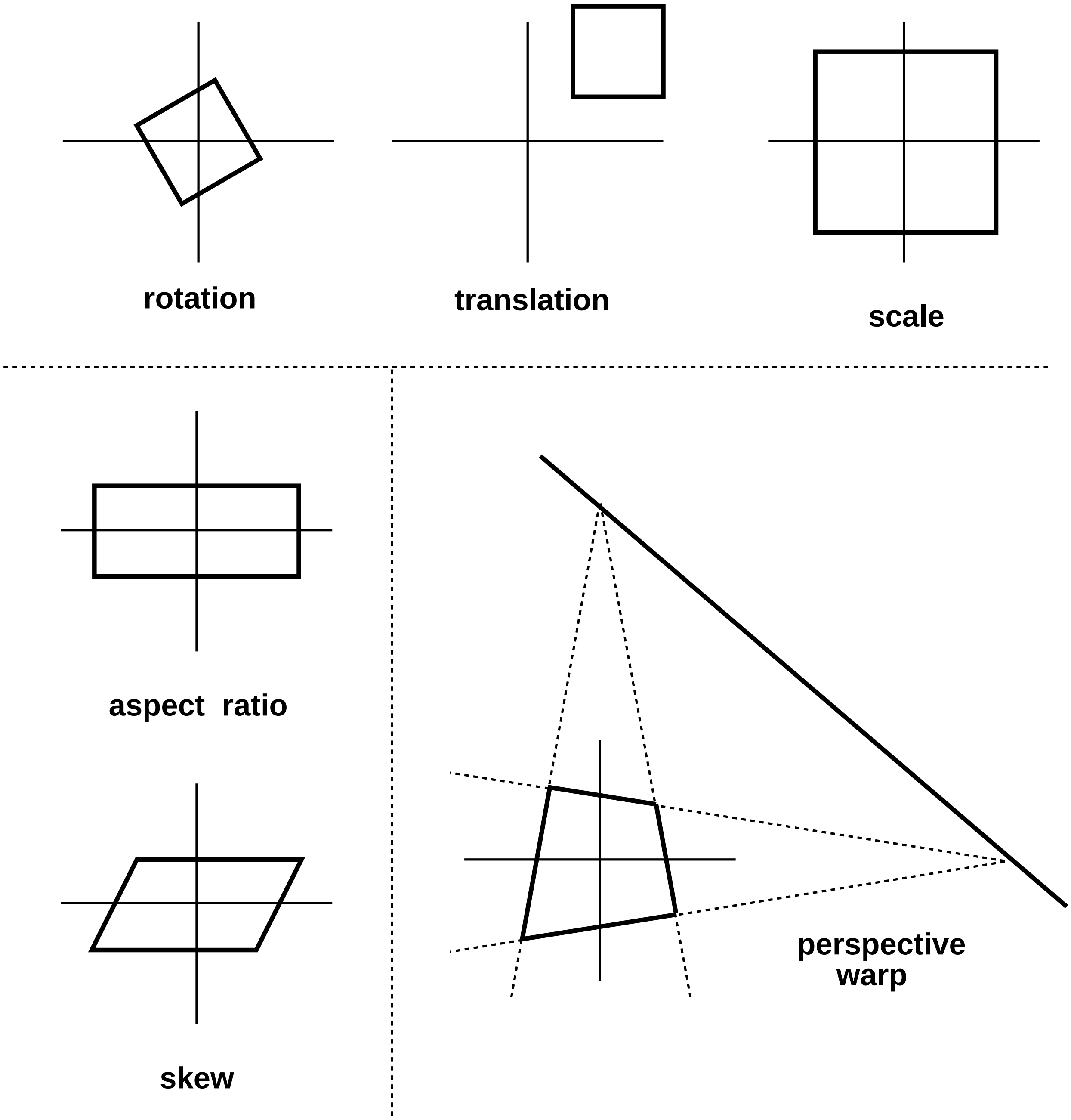


Figure A.4: Effects of the eight independent planar homography parameters on the image of a unit square.

to line homography, and at least five points, no four of which are coplanar, are needed to uniquely estimate a projective transformation over three-space.

The fundamental theorem of projective geometry determines the theoretical minimum number of points required to estimate the parameters of a homography, but in practice it is better to use as many point 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 estimating a plane to plane homography from four or more finite point correspondences [Faugeras88]. In this case equation (A.2) can be written as the fractional linear transformation

$$X' = \frac{aX + bY + c}{gX + hY + i} \quad , \quad Y' = \frac{dX + eY + f}{gX + hY + i} \; . \tag{A.6}$$

Each point to point correspondence provides two constraints on the transformation parameters, namely

$$\begin{aligned} aX + bY + c - X'(gX + hY + i) &= 0 \\ dX + eY + f - Y'(gX + hY + i) &= 0. \end{aligned} \tag{A.7}$$

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.

A.2 Duality

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.

There is a simple relationship between a homography matrix estimated from points and the corresponding dual homography matrix estimated from lines. Under a point homograpy $\boldsymbol{H}$, the line passing through points $\boldsymbol{p}$ and $\boldsymbol{q}$ now passes through $\boldsymbol{Hp}$ and $\boldsymbol{Hq}$, and its homogeneous coordinates therefore become $\boldsymbol{Hp} \times \boldsymbol{Hq}$. It is straightforward, though slightly tedious, to show that for any nonsingular 3×3 matrix $\boldsymbol{H}$,

$$\boldsymbol{Hp} \times \boldsymbol{Hq} \;=\; [\boldsymbol{H}^{-1}]^{T}(\boldsymbol{p} \times \boldsymbol{q})\,. \tag{A.8}$$

This means that the homography matrix that maps line coordinates to line coordinates under a given projective mapping is the transpose of the inverse of the matrix mapping points to points, and vice versa.

Using the above relationship a homography estimated from line correspondences can always be converted to a homography that maps points. Since line segment extraction is a more global process than point extraction, it seems reasonable to expect a homography derived by transforming an estimated line homography to be more accurate than one estimated directly from points.

A.3 Incidence Relations

Once the existence of two or more geometric entities is postulated, one can begin to ask how they relate to each other geometrically. The fundamental relationships defined between projective geometric elements are the incidence relations. For example, given two points in the projective plane, they can either coincide, or be distinct. Given three points, they can either coincide, form a *colinear* set all lying on a single line, or else form the vertices of a nondegenerate triangle. A similar situation holds for lines in the projective plane, due to the principle of duality. Three lines in the plane can either overlap exactly, form a *line pencil* intersecting in a single point, or intersect in pairs to form three distinct points.

Staying with the projective plane for concreteness, the homogeneous coordinate representation of points and lines in the plane allows questions about incidence to be reformulated in terms of linear subspace dimension. Recall that a point or line in the projective plane is represented by a nonzero, 3D homogeneous coordinate vector that can be scaled arbitrarily. The set of equivalent homogeneous coordinate vectors that all represent the same point or line can therefore be visualized as a set of vectors lying along a single infinite line through the origin of R^3, and therefore spanning a linear subspace of dimension one. The dimension of the subspace spanned by a set of vectors can be represented by the rank of a matrix with columns formed by the point coordinates. Given n homogeneous coordinate vectors, arranged to form the columns of a $3 \times n$ matrix, those homogeneous coordinates represent coincident points, *i.e.* represent the same point, iff the matrix has rank one.

Using the notion of matrix rank, the incidence relations between sets of points (and by duality, lines) in the projective plane can be enumerated. The rank of a $3 \times n$ matrix can be at most three, and since a homogeneous coordinate vector cannot consist of all zeros, the rank of a matrix formed by homogeneous coordinate vectors must be at least one. If the rank is one, all points (lines) are coincident, and therefore equivalent. In the text, this is called Rank I incidence. If the rank of the homogeneous coordinate matrix is two, all the points lie on a single line (all the lines intersect in a single point). This is called Rank II incidence. Finally, if the matrix has full rank then at least three points (lines) in the set form a nondegenerate triangle. Questions about whether a configuration of four points is arranged so that no three of them are colinear, a necessary condition for estimating the parameters of a projective transformation, can be verified by requiring all 3×3 submatrices formed by choosing three columns from the 3×4 homogeneous coordinate matrix describing the geometric configuration to have full rank.

A.4 Projective Invariants

Geometric properties can be classified into two types : descriptive and metric [Mihalek72]. Descriptive properties represent qualitative positional relationships between structures in the scene, including colinearity, coplanarity and tangency. Metric properties, on the other hand, are measured quantities like distance, orientation and area. The classification of a geometric property as either descriptive or metric is made precise by the study of projective geometry. A descriptive property remains invariant under projective transformations.

All of the incidence relations between projective geometric elements are invariant. For example, consider n colinear points in a plane. Their colinearity relation is expressed by the matrix equation

$$rank\,[\boldsymbol{p}_1\,\boldsymbol{p}_2\,\cdots\,\boldsymbol{p}_n] \;=\; 2, \tag{A.9}$$

where each column $\boldsymbol{p}_i$ is the 3×1 homogeneous coordinate vector of point i. An arbitrary homography $\boldsymbol{H}$ transforms point $\boldsymbol{p}_i$ to a new point $\boldsymbol{p}_i{}' = k_i \boldsymbol{H} \boldsymbol{p}_i$. This is expressed by

$$[\boldsymbol{p}_1{}'\,\boldsymbol{p}_2{}'\,\cdots\,\boldsymbol{p}_n{}'] \;=\; \boldsymbol{K}\boldsymbol{H}\,[\boldsymbol{p}_1\,\boldsymbol{p}_2\,\cdots\,\boldsymbol{p}_n]\,, \tag{A.10}$$

where $\boldsymbol{K}$ is a diagonal matrix composed of the nonzero scale factors k_i. $\boldsymbol{K}$ is obviously nonsingular, and $\boldsymbol{H}$ is also nonsingular due to the fact that it represents a homography. It is a well-known result from matrix algebra that when a matrix $\boldsymbol{A}$ is nonsingular, the rank of $\boldsymbol{AB}$ is that of $\boldsymbol{B}$ [Ayres62]. Therefore

$$rank\,[\boldsymbol{p}_1{}'\,\boldsymbol{p}_2{}'\,\cdots\,\boldsymbol{p}_n{}'] \;=\; rank\,[\boldsymbol{p}_1\,\boldsymbol{p}_2\,\cdots\,\boldsymbol{p}_n] \;=\; 2. \tag{A.11}$$

Therefore, when a set of colinear points in a plane is subjected to a planar homography, their resulting positions are still colinear. Colinearity is thus invariant under a homography, and is a descriptive rather than a metric geometric property.

An important example of a numeric projective invariant on a line is the *cross-ratio* of four points. More general projective invariants in the plane and in 3-space can be constructed from the cross-ratio. Letting $\boldsymbol{p_1}$, $\boldsymbol{p_2}$, $\boldsymbol{p_3}$ and $\boldsymbol{p_4}$ be the 2×1 homogeneous coordinate vectors of four points on a line, their cross-ratio is defined as

$$CR(\boldsymbol{p_1}, \boldsymbol{p_2}, \boldsymbol{p_3}, \boldsymbol{p_4}) = \frac{det\,[\boldsymbol{p_1}\;\boldsymbol{p_3}]}{det\,[\boldsymbol{p_1}\;\boldsymbol{p_4}]} \times \frac{det\,[\boldsymbol{p_2}\;\boldsymbol{p_4}]}{det\,[\boldsymbol{p_2}\;\boldsymbol{p_3}]}\,. \tag{A.12}$$

To prove invariance one proceeds as before, letting $\boldsymbol{p_i}' = k_i \boldsymbol{H} \boldsymbol{p_i}$ for some nonsingular matrix $\boldsymbol{H}$. The cross ratio of the transformed points is

$$\begin{aligned} CR(\boldsymbol{p_1}', \boldsymbol{p_2}', \boldsymbol{p_3}', \boldsymbol{p_4}') &= \frac{k_1 k_3 (det\,\boldsymbol{H})(det\,[\boldsymbol{p_1}\;\boldsymbol{p_3}])}{k_1 k_4 (det\,\boldsymbol{H})(det\,[\boldsymbol{p_1}\;\boldsymbol{p_4}])} \times \frac{k_2 k_4 (det\,\boldsymbol{H})(det\,[\boldsymbol{p_2}\;\boldsymbol{p_4}])}{k_2 k_3 (det\,\boldsymbol{H})(det\,[\boldsymbol{p_2}\;\boldsymbol{p_3}])} \\ &= \frac{det\,[\boldsymbol{p_1}\;\boldsymbol{p_3}]}{det\,[\boldsymbol{p_1}\;\boldsymbol{p_4}]} \times \frac{det\,[\boldsymbol{p_2}\;\boldsymbol{p_4}]}{det\,[\boldsymbol{p_2}\;\boldsymbol{p_3}]}\,. \end{aligned} \tag{A.13}$$

This verifies that the value of the cross-ratio of four points on a line remains unchanged by a homography, and it is therefore a projective invariant.

APPENDIX B

Probability on the Sphere

Readers of this thesis are most likely unfamiliar with applications of probability theory and statistics on the sphere. To help make this work more accessible, several relevant mathematical preliminaries and statistical results are collected here. Among the background topics covered are

- definitions of the sphere S^n and projective space P^n,
- spherical vs. cartesian coordinates,
- integration and spherical surface area,
- probability density functions on the sphere,
- polynomial approximations and spherical harmonic expansions,

Although for generality's sake equations are written in terms of an arbitrary dimension n, frequent examples are provided that specialize these general equations to the familiar one, two and three dimensional cases.

B.1 The Unit Sphere and Projective Space

The unit sphere S^n is defined as the set of all points unit distance away from some origin in Euclidean space E^{n+1}. Let $\boldsymbol{x} = (x_1, \ldots, x_{n+1}) \in R^{n+1}$ be the rectangular coordinates of a point in E^{n+1}. Then the unit sphere S^n is the set of all points $\boldsymbol{x}$ such that $x_1^2 + \cdots + x_n^2 = \|\boldsymbol{x}\|^2 = 1$. We will be concerned primarily with the unit circle S^1, the unit sphere S^2, and the unit hypersphere S^3. Note that S^n corresponds to the n-dimensional *surface* of a $(n+1)$-dimensional unit ball centered at the origin.

A curvilinear coordinate system on S^n is provided by the spherical angles $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$, which are related to the rectangular coordinates of a point on the sphere by

$$\begin{array}{rcccl} x_1 & = & x_1(\boldsymbol{\theta}) & = & \sin\theta_1 \sin\theta_2 \ldots \sin\theta_{n-1} \sin\theta_n \\ x_2 & = & x_2(\boldsymbol{\theta}) & = & \sin\theta_1 \sin\theta_2 \ldots \sin\theta_{n-1} \cos\theta_n \\ x_3 & = & x_3(\boldsymbol{\theta}) & = & \sin\theta_1 \sin\theta_2 \ldots \cos\theta_{n-1} \\ \vdots & & & & \\ x_n & = & x_n(\boldsymbol{\theta}) & = & \sin\theta_1 \cos\theta_2 \\ x_{n+1} & = & x_{n+1}(\boldsymbol{\theta}) & = & \cos\theta_1 \end{array}$$

with $0 \leq \theta_n \leq 2\pi$ and $0 \leq \theta_i \leq \pi$ for $i = 1, \ldots, n-1$ (see [Muirhead82]). Rectangular and spherical coordinate representations will be used interchangeably in this section. In most cases rectangular coordinates provide the most concise equations, however integration over the sphere is expressed more naturally in spherical coordinates.

Real projective space P^n can be defined as the set of all lines through the origin of Euclidean space E^{n+1}. Again identify R^n with E^n, and let $(R^n - \{0\})$ denote Euclidean space with the origin removed. We define an equivalence relation $\sim$ on points $\boldsymbol{x}, \boldsymbol{y} \in (R^n - \{0\})$ such that $\boldsymbol{x} \sim \boldsymbol{y}$ iff there exists some nonzero scalar k such that $\boldsymbol{x} = k\,\boldsymbol{y}$. The projective plane P^n is then formally defined as the quotient space $(R^{n+1} - \{0\}) \;/ \sim$. Geometrically speaking, 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^n, and form the quotient space $S^n / \sim$. Each equivalence class now contains one pair of diametrically opposite points, or *antipodes.* Equating these equivalence classes with those of the projective plane in the obvious way shows that the surface of the unit sphere with antipodal points equated is isomorphic to the projective plane. The unit circle S^1, the unit sphere S^2, and the unit hypersphere S^3, when combined with the antipodal symmetry constraint $\boldsymbol{x} = -\boldsymbol{x}$, represent the projective line P^1, the projective plane P^2, and projective 3-space P^3, respectively.

B.2 Density Functions on the Sphere

Any real-valued function $f(\boldsymbol{x})$ defined over R^{n+1} can be interpreted as a function $f(\boldsymbol{x}(\boldsymbol{\theta}))$ on the unit sphere, by restricting its domain of operation to points in S^n. If the function satifies the relation $f(\boldsymbol{x}) = f(-\boldsymbol{x})$ for $x \in S^n$, then it is said to be *antipodally symmetric* over the sphere, and thus additionally represents a function in real projective space P^n.

The integral of a real-valued function $f(\boldsymbol{x})$ over the surface of the unit sphere in any dimension will be written

$$\int f(\boldsymbol{x})\, dS^n$$

where dS^n is the surface area element of the sphere S^n and the single integral sign denotes the appropriate n-dimensional multiple integral over the sphere. In spherical coordinates, the surface area element for S^n is

$$dS^n = dS^n(\boldsymbol{\theta}) = \sin^{n-1}\theta_1 \sin^{n-2}\theta_2 \cdots \sin\theta_{n-1}\, d\theta_1 d\theta_2 \cdots d\theta_n$$

and the multiple integral consists of n integrals, one for each θ_i.

Any non-negative continuous function can be interpreted as a probability density function on the sphere with the addition of normalization constant such that the function integrates to unity over the sphere. Thus if $f(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in S^n$, then

$$\left[\int f(\boldsymbol{x})\, dS^n\right]^{-1} f(\boldsymbol{x})\, dS^n \;=\; C\, f(\boldsymbol{x})\, dS^n$$

is a probability density on the sphere, with dS^n now playing the role of Lebesgue measure on S^n. The total surface area of the sphere is

$$\omega_n = \int dS^n = \frac{2(\pi)^{\frac{n+1}{2}}}{\Gamma(\frac{n+1}{2})} .$$

Therefore a simple probability density function on the sphere is

$$\omega_n^{-1}\, dS^n .$$

This function has constant value (the reciprocal of the total surface area) over the whole surface of the sphere , and therefore represents the *uniform* probability density function on the sphere. It is obviously antipodally symmetric, so it also serves as a uniform density over projective space.

Example B.4 *A point on the unit circle S^1 is indexed by angle $\theta = \theta_1$, and simultaneously by rectangular coordinates $(x, y) = (x_1, x_2) = (\sin\theta, \cos\theta)$. The surface area element dS^1 is merely $d\theta$. Integration of a function $f(x, y)$ over S^1 is performed by*

$$\int_0^{2\pi} f(\sin\theta, \cos\theta) d\theta .$$

Total surface area (circumference) is $\int d\theta = 2\pi$. The uniform distribution on the circle and the projective line therefore has density function $(2\pi)^{-1}\, d\theta$.

■

Example B.5 *The unit sphere S^2 is indexed by spherical angles $(\theta, \phi) = (\theta_1, \theta_2)$ and by rectangular coordinates*

$$(x, y, z) = (x_1, x_2, x_3) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)\ .$$

The surface area element dS^2 is $\sin\theta\, d\theta d\phi$. Integration of a function $f(x, y, z)$ over S^2 is handled by the double integral

$$\int_0^{2\pi}\int_0^{\pi} f(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)\ \sin\theta\, d\theta d\phi\ .$$

Total surface area of the unit sphere is $\int\int\ \sin\theta\, d\theta d\phi = 4\pi$. The uniform density function on the unit sphere and the projective plane is $(4\pi)^{-1}\ \sin\theta\, d\theta d\phi$.

■

Example B.6 *The unit hypersphere S^3 is indexed by spherical angles $(\theta, \phi, \xi) = (\theta_1, \theta_2, \theta_3)$ and by rectangular coordinates*

$$(x, y, z, w) = (x_1, x_2, x_3, x_4) = (\sin\theta\sin\phi\cos\xi, \sin\theta\sin\phi\sin\xi, \sin\theta\cos\phi, \cos\theta)\ .$$

The surface area element dS^3 is $\sin^2\theta\sin\phi\, d\theta d\phi d\xi$. Integration of a function $f(x, y, z, w)$ over S^3 corresponds to the multiple integral

$$\int_0^{2\pi}\int_0^{\pi}\int_0^{\pi} f(x(\boldsymbol{\theta}), y(\boldsymbol{\theta}), z(\boldsymbol{\theta}), w(\boldsymbol{\theta}))\ \sin^2\theta\sin\phi\, d\theta d\phi d\xi\ .$$

Total surface area (actually a volume) of the hypersphere is $\int\int\int\ \sin^2\theta\sin\phi\, d\theta d\phi d\xi = 2\pi^2$. The uniform density function on the unit hypersphere and projective 3-space is thus

$$(2\pi^2)^{-1}\ \sin^2\theta\sin\phi\, d\theta d\phi d\xi.$$

■

B.3 Polynomial Approximations and Spherical Harmonics

A polynomial in n variables $x_1, \ldots, x_n$ consists of a sum of terms, called *monomials,* each having the form

$$cx_1^{q_1} x_2^{q_2} \cdots x_n^{q_n}.$$

The *degree* or *order* of a monomial is the sum of its exponents $q_1 + q_2 + \cdots + q_n$. The degree of a polynomial is the highest degree attained by any monomial appearing in it. Thus a polynomial of degree k consists of monomials of degree less than or equal to k. A polynomial consisting only of monomials having exactly degree k is said to be a *homogeneous* polynomial of that degree. To avoid confusion, a general polynomial of degree k will be written as P_k, while H_k will denote a homogeneous polynomial of degree k.

By the Weierstrass approximation theorem, any continuous function on the sphere can be approximated uniformly by a polynomial [Dennery69]. This means a polynomial with finite degree can always be found that agrees arbitrarily closely with a given continuous function at points on the surface of the sphere. A nested family of polynomial approximations can be constructed by considering polynomials of degree k, for increasing k. Absolute error in the approximation monotonically decreases as terms of higher degree are added, since polynomials of higher degree have more degrees of freedom available to fit a desired function. At some point, however, there is a tradeoff between accuracy and the number of terms that must be handled. To find the polynomial of degree k that "best" approximates a given continuous function, we will choose polynomial coefficients that minimize the least-squares fit error.

The space of continuous functions on the sphere is a linear vector space, albeit one of infinite dimension. The set of all possible polynomials of degree k on the sphere constitutes a linear vector space of finite dimension. A basis for this space is provided by the set of all monomials of degree $\leq k$. The number of monomials of exactly degree k in $n+1$ variables is $\begin{pmatrix} k+n \\ k \end{pmatrix}$. the number of ways of choosing k unordered elements from a $(k+n)$ element set. However, when restricted to the sphere, not all of these monomials are linearly independent. Let $N(k,n)$ be the number of linearly independent monomials of degree k over S^n. Then

$$N(k,n) \;=\; \frac{(2k+n-1)}{(k+n-1)} \begin{pmatrix} k+n-1 \\ k \end{pmatrix}.$$

The dimension $D(k,n)$ of the space of general polynomial functions of degree k over S^n therefore has dimension

$$D(k,n) \;=\; \sum_{j=0}^{k} N(j,n).$$

Each polynomial "point" in this space is a linear combination

$$\sum_{i=0}^{D(k,n)} a_i \, M_i(\boldsymbol{x})$$

with a_i an arbitrary real coefficient and M_i one of the $D(k,n)$ linearly independent basis monomials of degree $\leq k$. In addition, letting M_k be an arbitrary monomial of degree k, we have $M_k(-\boldsymbol{x}) = (-1)^k M_k(\boldsymbol{x})$, so monomials of even degree form a basis for the antipodally symmetric functions on the sphere, and thus for functions in projective space.

Example B.7 *The monomials of degree three or less in the variables x, y and z are:*

degree 0:	1
1:	$x,\ y,\ z$
2:	$x^2,\ xy,\ y^2,\ xz,\ yz,\ z^2$
3:	$x^3,\ x^2y,\ xy^2,\ y^3,\ x^2z,\ xyz,\ y^2z,\ xz^2,\ yz^2,\ z^3.$

On the unit sphere S^2 *the constraint* $x^2 + y^2 + z^2 = 1$ *is in effect. Therefore, one of the monomials of degree two can be written in terms of other monomials with the same, or lesser degree, namely* $z^2 = 1 - x^2 - y^2$*. Of the 10 monomials of degree three, only 7 are linearly independent, since*

$$\begin{aligned} xz^2 &= x - x^3 - xy^2 \\ yz^2 &= y - x^2y - y^3 \\ z^3 &= z - x^2z - y^2z. \end{aligned}$$

Altogether, there are

$$D(3,2) = N(0,2) + N(1,2) + N(2,2) + N(3,2) = 1 + 3 + 5 + 7 = 16$$

linearly independent monomials forming a basis for the 16-dimensional space of third-order polynomials on the sphere S^2*. Of these, only six monomials* 1*,* x^2*,* xy*,* y^2*,* xz *and* yz *are antipodally symmetric functions, and thus able to provide a basis for second-order polynomials in the real projective plane* P^2*.*

■

Although theoretically any linearly independent basis set is as good as any other, when fitting polynomials there are benefits to choosing an orthonormal basis set, both in terms of numerical stability and ease of computation [Rivlin69]. To define orthonormality on a vector space, metric definitions of distance and angle must be introduced by choosing an inner product on the space. For the space of functions on the sphere, the inner product is defined as

$$\langle f(\boldsymbol{x}), g(\boldsymbol{x}) \rangle \;=\; \int f(\boldsymbol{x}) g(\boldsymbol{x})\, dS^n \;.$$

A set of functions $\{\, Y_i(\boldsymbol{x}) \,\}$ is said to be orthonormal on the sphere when

$$\int Y_i(\boldsymbol{x}) Y_j(\boldsymbol{x})\, dS^n \;=\; \begin{cases} 1 & i = j \\ 0 & \text{otherwise.} \end{cases}$$

An orthonormal basis for polynomials on S^n is provided by the set of *spherical harmonics* in n dimensions [Erdélyi53, Müller66]. Since they are merely a different basis for the same polynomial space, there are $N(k,n)$ spherical harmonics having exactly degree k and $D(k,n)$ harmonics of degree k or less. Spherical harmonics form a complete set of orthonormal functions, so any continuous function f on the sphere can be written as a series

$$f(\boldsymbol{x}) \;=\; \sum_{i=0}^{\infty} a_i\, Y_i(\boldsymbol{x})$$

of spherical harmonic polynomials. Because they are orthonormal, we have

$$\int Y_j(\boldsymbol{x}) f(\boldsymbol{x})\, dS^n \;=\; \sum_{i=0}^{\infty} a_i \int Y_i(\boldsymbol{x}) Y_j(\boldsymbol{x})\, dS^n \;=\; a_i \;,$$

therefore each coefficient in the series may be computed relatively easily. Furthermore, the spherical harmonic polynomial approximation of degree k

$$\sum_{i=0}^{D(k,n)} a_i\, Y_i(\boldsymbol{x}) \;=\; \sum_{i=0}^{D(k,n)} \left(\int Y_i(\boldsymbol{x}) f(\boldsymbol{x})\, dS^n \right) Y_i(\boldsymbol{x})$$

is the best kth-order approximation to $f(\boldsymbol{x})$ in the least-squares sense, i.e. this kth-order approximating polynomial is the orthogonal projection of $f(\boldsymbol{x})$ onto the space of all possible polynomials of degree k.

APPENDIX C

Transformation of Random Variables

The transformation of variables technique provides a general method for determining how a probability density function changes when its underlying coordinate system is changed. In the context of this thesis, the method is used to determine how the uncertainty in a geometric object is transformed due to a change in object representation. The parameter vectors describing a geometric object before and after a change in representation are related by the vector function $\boldsymbol{y} = T(\boldsymbol{x})$. If parameter vector $\boldsymbol{x}$ is a vector of random variables, representing an uncertain object description, then the resulting parameter vector $\boldsymbol{y}$ is also a random vector, with a density function determined both by the density of $\boldsymbol{x}$ and the transformation T.

Recall that a probability density function $g(\boldsymbol{x})$ is a function such that $\int_V g(\boldsymbol{x})\, d\boldsymbol{x}$ yields the probability that the value of random vector $\boldsymbol{x}$ lies within a given volume V of parameter space. The general idea behind the transformation of variables technique is to rewrite the function $g(\boldsymbol{x})$ in terms of the new variables $\boldsymbol{y}$, and to determine the change in differential volume $d\boldsymbol{x}$ that occurs due to the transformation. This latter manipulation is simplified considerably by the fact that, at an infinitesimal scale, any well-behaved nonlinear function can be approximated locally by an affine function, and likewise, any nonlinear parameter space (manifold) can be approximated locally by an affine tangent space.

There is a similarity between this approach and the function linearization approach to approximate covariance propagation as described in Section 2.2.1. In that approach, a given nonlinear transformation is approximated by the first order terms of its Taylor series expansion, evaluated about the current mean value of $\boldsymbol{x}$. The mean and covariance of $\boldsymbol{x}$ are then propagated through the resulting affine approximation to determine the approximate mean and covariance of the transformed variables $\boldsymbol{y}$. The important difference in this section is that instead of computing a single, global affine approximation to the nonlinear transformation, a seperate affine approximation is being implicitly computed at *every* point in the parameter space of $\boldsymbol{x}$ to determine how the probability density changes in an infinitesimal neighborhood of that point. By simultaneously considering how the density changes locally at every point, an exact formula for the transformed probability density function over the whole parameter space is determined.

C.1 Mapping R^n to R^n

Consider a random vector $\boldsymbol{x} \in R^n$ distributed according to the density function $g(\boldsymbol{x})$, and the vector transformation $T : R^n \mapsto R^n$. For the transformed random vector $\boldsymbol{y} = T(\boldsymbol{x})$ we would like to compute a density function $h(\boldsymbol{y})$ such that for any volume V of R^n

$$\int_V g(\boldsymbol{x})\, d\boldsymbol{x} \quad = \quad \int_{T(V)} h(\boldsymbol{y})\, d\boldsymbol{y} \tag{C.1}$$

where $T(V)$ denotes the image of volume V under transformation T. This constraint insures that the probability computed over a given volume of parameter space will be the same no matter what coordinate system is used to do the computation.

If transformation T is one-to-one, onto, and continuously differentiable, computation of the resulting density function $h(\boldsymbol{y})$ is a straightforward application of the change of variables formula for multiple integrals [Williamson68]. Since the inverse transformation T^{-1} is well-defined in this case, we can make the substitutions

$$\boldsymbol{x} = T^{-1}(\boldsymbol{y}) \tag{C.2}$$

$$d\boldsymbol{x} \quad = \quad \frac{d\boldsymbol{y}}{J_T(T^{-1}(\boldsymbol{y}))} \quad = \quad J_{T^{-1}}(\boldsymbol{y})\, d\boldsymbol{y} \tag{C.3}$$

where the Jacobians J_T and $J_{T^{-1}}$ specify how the infinitesimal volume element at a point is scaled under the transformation T and its inverse, respectively (see Figure C.1). The Jacobian of a transformation from R^n to R^n is the absolute value of the determinant of the $n \times n$ *Jacobian matrix* containing the first partial derivatives of each transformed coordinate with respect to each original coordinate, evaluated at a point in the domain of the transformation. That is,

$$J_T(\boldsymbol{x}_0) \quad \equiv \quad |\, det\, \frac{\partial T}{\partial x}(\boldsymbol{x}_0)| \quad \equiv \quad |\, det \left\{ \frac{\partial y_i}{\partial x_j}(\boldsymbol{x}_0) \right\} | \ , \ \boldsymbol{x}_0 \in V \tag{C.4}$$

$$J_{T^{-1}}(\boldsymbol{y}_0) \quad \equiv \quad |\, det\, \frac{\partial T^{-1}}{\partial y}(\boldsymbol{y}_0)| \quad \equiv \quad |\, det \left\{ \frac{\partial x_i}{\partial y_j}(\boldsymbol{y}_0) \right\} | \ , \ \boldsymbol{y}_0 \in T(V) \tag{C.5}$$

The two alternative substitutions of equation (C.3) reflect the fact that the reciprocal of the Jacobian of a transformation is the Jacobian of the inverse transformation.

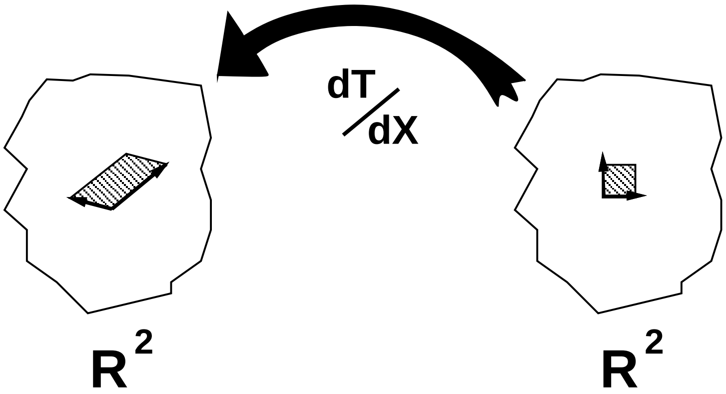


Figure C.1: The Jacobian of a transformation T from R^2 to R^2. The Jacobian matrix $\partial T/\partial X$ maps a unit square at point x_0 in the domain of T into a parallelogram at point $T(x_0)$ in the range. The change in area between the unit square and the parallelogram is the Jacobian of transformation T at point x_0.

Using substitutions (C.2) and (C.3) to change the variables of integration, we write

$$\int_V g(\boldsymbol{x})\, d\boldsymbol{x} \;=\; \int_{T(V)} \frac{g(T^{-1}(\boldsymbol{y}))}{J_T(T^{-1}(\boldsymbol{y}))}\, d\boldsymbol{y} \;=\; \int_{T(V)} g(T^{-1}(\boldsymbol{y}))J_{T^{-1}}(\boldsymbol{y})\, d\boldsymbol{y} \qquad \text{(C.6)}$$

Comparing (C.6) with equation (C.1) we find that the probability density function resulting from the transformation $\boldsymbol{y} = T(\boldsymbol{x})$ is given by the formula

$$h(\boldsymbol{y}) \;=\; \frac{g(T^{-1}(\boldsymbol{y}))}{J_T(T^{-1}(\boldsymbol{y}))} \;=\; g(T^{-1}(\boldsymbol{y}))J_{T^{-1}}(\boldsymbol{y}) \qquad \text{(C.7)}$$

Which Jacobian is easier to compute, J_T or $J_{T^{-1}}$, will in practice determine which of the two equivalent forms of this equation to use.

Example C.8 *Consider the affine transformation* $\boldsymbol{y} = A\boldsymbol{x} + b$*, where* $\boldsymbol{A}$ *is a nonsingular* $n \times n$ *matrix and* b *is an* $n \times 1$ *vector of constants. This transformation, its inverse, and their corresponding Jacobians are listed as follows*

$$
\begin{aligned}
T(x) &= A\boldsymbol{x} + b & \qquad T^{-1}(y) &= A^{-1}(\boldsymbol{y} - b) && \text{(C.8)}\\
\frac{\partial T}{\partial x}(x) &= A & \qquad \frac{\partial T^{-1}}{\partial y}(y) &= A^{-1} && \text{(C.9)}\\
J_T(x) &= |\, det\, A\,| & \qquad J_{T^{-1}}(y) &= |\, det\, A^{-1}| && \text{(C.10)}
\end{aligned}
$$

Let $\boldsymbol{x}$ *be an n-dimensional multivariate Gaussian random variable with mean vector* μ *and covariance matrix* Σ*. The probability density function is written*

$$
g(\boldsymbol{x}) = \frac{|\, det\, \Sigma^{-1}|^{\frac{1}{2}}}{(\sqrt{2\pi})^n}\; exp\{ -\frac{1}{2}\, (x - \mu)^t\; \Sigma^{-1}\, (x - \mu)\, \} \tag{C.11}
$$

Plugging into the transformation of variables formula (C.7)

$$
\begin{aligned}
h(\boldsymbol{y}) &= |\, det\, A^{-1}|\, g(A^{-1}(\boldsymbol{y} - b))\\
&= \frac{|\, det\, A^{-1}||\, det\, \Sigma^{-1}|^{\frac{1}{2}}}{(\sqrt{2\pi})^n}\; exp\{ -\frac{1}{2}\, \left[A^{-1}(\boldsymbol{y} - b) - \mu\right]^t\; \Sigma^{-1}\, \left[A^{-1}(\boldsymbol{y} - b) - \mu\right] \}\\
&= \frac{|\, det(A\Sigma A^t)^{-1}|^{\frac{1}{2}}}{(\sqrt{2\pi})^n}\; exp\{ -\frac{1}{2}\, [\boldsymbol{y} - (A\mu + b)]^t\; (A\Sigma A^t)^{-1}\, [\boldsymbol{y} - (A\mu + b)]\, \}
\end{aligned}
$$

The transformed vector $\boldsymbol{y}$ *of random variables is therefore also distributed as a multivariate Gaussian density, with new mean vector* $A\mu + b$ *and covariance matrix* $A\Sigma A^t$*.*

■

In many cases, the standard restrictions placed on the transformation of variables technique, namely that transformation T should be one-to-one, onto and nonsingular, can be relaxed. If T is not onto, then the domain of the resulting density function simply needs to be restricted to the range of T. If T is not one-to-one, the inverse transformation T^{-1} is not uniquely defined. However, each solution $\boldsymbol{x}_i$ such that $\boldsymbol{y} = T(\boldsymbol{x}_i)$ contributes its share of probability mass to the resulting density function about $\boldsymbol{y}$, so that the resulting density can be written as [Papoulis65]

$$h(\boldsymbol{y}) \;=\; \frac{g(\boldsymbol{x}_1)}{J_T(\boldsymbol{x}_1)} + \cdots + \frac{g(\boldsymbol{x}_i)}{J_T(\boldsymbol{x}_i)} + \cdots \tag{C.12}$$

Finally, transformation T is permitted to be singular, i.e. to have a Jacobian of either 0 or infinity, provided it does so only on a set of points having zero measure [Williamson68]. Intuitively, an m-dimensional set of points has measure zero with respect to an n-dimensional space when $m < n$.

Example C.9 *Let P be a projective transformation in the plane, represented in homogeneous coordinates by a nonsingular 3×3 matrix A. To see how P affects probability distributions in the plane, first write it in terms of inhomogeneous plane coordinates, namely as the fractional linear form*

$$y_1 \equiv \frac{u}{w} = \frac{a_{11}x_1 + a_{12}x_2 + a_{13}}{a_{31}x_1 + a_{32}x_2 + a_{33}} \qquad y_2 \equiv \frac{v}{w} = \frac{a_{21}x_1 + a_{22}x_2 + a_{23}}{a_{31}x_1 + a_{32}x_2 + a_{33}}$$

This representation is singular when the denominator w of the fractional linear form is zero. The denominator is only zero, however, for the set of points (x_1, x_2) lying on the line $a_{31}x_1 + a_{32}x_2 + a_{33} = 0$. Since a line of points has measure zero in the plane, the transformation of variables procedure is still applicable.

The Jacobian matrix of first partial derivatives can be written as

$$\frac{\partial P}{\partial x} = \frac{1}{w}\begin{bmatrix} 1 & 0 & -u/w \\ 0 & 1 & -v/w \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \tag{C.13}$$

$$= \frac{1}{w}\left\{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} - \begin{bmatrix} u \\ v \end{bmatrix}\begin{bmatrix} 1/w \end{bmatrix}\begin{bmatrix} a_{31} & a_{32} \end{bmatrix}\right\}$$

To evaluate the determinant of this matrix, we first recall a result from the theory of matrix algebra. Let M be an $n \times n$ matrix, partitioned into four pieces as

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

where M_{11} is $l \times l$, M_{12} is $l \times m$, M_{21} is $m \times l$ and M_{22} is $m \times m$, for some $l + m = n$. Then the determinant of M is

$$det\, M = det\, M_{22}\ det\,(M_{11} - M_{12}M_{22}^{-1}M_{21})$$

assuming M_{22} is invertible [Eaton83, proposition 1.34]. Using this identity,

$$det\,\frac{\partial P}{\partial x} = \frac{1}{w^2}\ det\left\{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} - \begin{bmatrix} u \\ v \end{bmatrix}\begin{bmatrix} 1/w \end{bmatrix}\begin{bmatrix} a_{31} & a_{32} \end{bmatrix}\right\}$$

$$= \frac{1}{w^2}\left(\frac{1}{w}\right) det \begin{bmatrix} a_{11} & a_{12} & u \\ a_{21} & a_{22} & v \\ a_{31} & a_{32} & w \end{bmatrix}$$

$$= \frac{1}{w^3}\ det \begin{bmatrix} a_{11} & a_{12} & (a_{11}x_1 + a_{12}x_2 + a_{13}) \\ a_{21} & a_{22} & (a_{21}x_1 + a_{22}x_2 + a_{23}) \\ a_{31} & a_{32} & (a_{31}x_1 + a_{32}x_2 + a_{33}) \end{bmatrix}$$

where the values of u, v and w have been substituted in the last column of the matrix. To simplify further, recall that a basic property of the determinant is that it remains unchanged *when a scalar multiple of any column is added to any other column. Adding $-x_1$ times the first column, and $-x_2$ times the second column, to the third column of this matrix shows that its determinant is the same as the determinant of A. Therefore*

$$J_P(x) \;=\; \left|\frac{1}{w^3}\,det\,A\right| \;=\; \frac{|\,det\,A\,|}{|\,a_{31}x_1 + a_{32}x_2 + a_{33}\,|^3} \tag{C.14}$$

Notice that the Jacobian of a projective transformation may depend on the position of the point at which it is evaluated. Note further that a necessary and sufficient condition for the Jacobian of a projective transformation to remain constant is that $a_{31} = a_{32} = 0$, precisely the conditions under which a projective transformation reduces to an affine transformation.

It is straightforward to see that the Jacobian of the inverse projective transformation, represented in homogeneous coordinates by matrix $B \equiv A^{-1}$ is

$$J_{P^{-1}}(y) \;=\; \frac{|\,det\,B\,|}{|\,b_{31}y_1 + b_{32}y_2 + b_{33}\,|^3} \tag{C.15}$$

and therefore a probability density function $g(x_1, x_2)$ in the plane is transformed into the new density function

$$h(y_1, y_2) \;=\; g\Big(\frac{b_{11}y_1 + b_{12}y_2 + b_{13}}{b_{31}y_1 + b_{32}y_2 + b_{33}}, \frac{b_{21}y_1 + b_{22}y_2 + b_{23}}{b_{31}y_1 + b_{32}y_2 + b_{33}}\Big)\; \frac{|\,det\,B\,|}{|\,b_{31}y_1 + b_{32}y_2 + b_{33}\,|^3}$$

For the measure zero set of points $b_{31}y_1 + b_{32}y_2 + b_{33} = 0$ where this formula will fail, the density function can be set arbitrarily to zero.

∎

C.2 Mapping between R^n and S^n

This section considers mappings between a 2D image plane or tangent plane and the surface of the unit sphere in 3D. Mappings from the image to the sphere determine how probability distributions in the image plane induce distributions on the sphere. Mappings from the sphere to a tangent plane are useful for display purposes. Some of the standard mappings that will be considered are orthographic projection, central projection, and equal-area projection. In order for these mappings to be well-defined when going from a plane to the sphere, their range will be restricted to the surface of a hemisphere. Restriction to a hemisphere also makes these mappings one-to-one, and hence invertible.

The standard method for computing Jacobian values assumes a mapping from R^n to R^n, since a square Jacobian matrix is needed to compute the determinant. The determinant of the Jacobian matrix indicates how differential areas and volumes are scaled as a result of the mapping. However, a mapping from plane coordinates (u, v) to hemisphere coordinates (x, y, z) seems to be a mapping from R^2 to R^3. The resulting Jacobian matrix is not square, and computation of the determinant can not be carried out. This difficulty can be resolved by keeping in mind the underlying definition of the Jacobian as a magnification factor for differential areas and volumes. The surface of the hemisphere is, of course, just a 2D surface after all, reflected by the constraint $x^2 + y^2 + z^2 = 1$ between Cartesian point coordinates. A standard result from vector calculus states that the independent column vectors of the 3×2 Jacobian matrix of a mapping from a plane to a 2D surface embedded in 3D, span the tangent plane to the surface at the point where the Jacobian is being evaluated [Williamson68]. A differential unit square around the evaluation point in R^2 is thus mapped to a differential parallelogram in the tangent plane to the surface in R^3. At the infinitesimal scale of differentials, the structure of a surface at a point is represented faithfully by the tangent plane at that point, therefore the area of the tangent parallelogram represents the exact change of differential area under the nonlinear mapping at that point (Figure C.2).

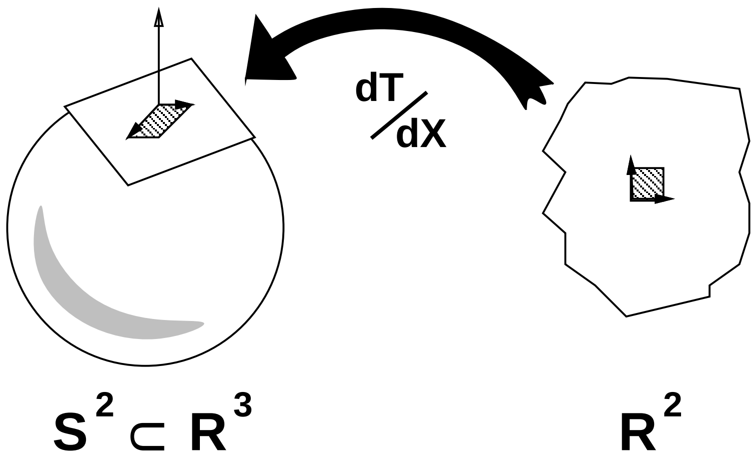


Figure C.2: The Jacobian of a transformation T from R^2 to S^2. The Jacobian matrix $\partial T/\partial X$ maps a unit square at point x_0 into a parallelogram in the tangent plane at point $T(x_0)$ on S^2. The change in area between the unit square and the tangent parallelogram is the Jacobian of transformation T at point x_0.

The area of a parallelogram is a multiple of the lengths of both sides and the sine of the angle between them. Each column vector of the 3×2 Jacobian matrix represents one side of the parallelogram (translated to the origin). But the length of the cross product of two vectors is equal to the length of both vectors multiplied by the sine of the angle between them. Thus the length of the cross product of the columns of the Jacobian matrix represents the change in differential area under the mapping at the point about which the Jacobian is evaluated.

In summary, when mapping from R^2 to a 2D surface in R^3, the columns of the Jacobian matrix span a parallelogram in the tangent plane to the surface. The area of this parallelogram, computed as the length of the cross product of the two column vectors, is the Jacobian. The Jacobian value computed in this way can be substituted directly into the transformation of variables formula C.7. In higher dimensions, the mapping from R^n to an n-dimensional surface in $R^{(n+1)}$ will generate an $(n+1) \times n$ Jacobian matrix with columns spanning a parallelopiped in the tangent space to the surface. The volume of this parallelopiped is the Jacobian of the transformation. The volume can be computed using a generalization of the cross product, called the wedge product or exterior product.

Example C.10 *The conversion that maps between spherical coordinates $\Theta = (\theta, \phi)$ and Cartesian coordinates $X = (x, y, z)$ on the hemisphere $z > 0$ is a mapping from a subset of R^2 to a surface in R^3. The forward and inverse mappings can be written as*

$$X(\Theta) = \begin{cases} x = \sin\theta\cos\phi \\ y = \sin\theta\sin\phi \\ z = \cos\theta \end{cases} \qquad \Theta(X) = \begin{cases} \theta = \mathit{arctan}(\sqrt{x^2+y^2}\,/z) \\ \phi = \mathit{arctan2}(y, x) \end{cases}$$

For notational clarity the transformations have been labeled by the coordinate systems they convert from and to. The label $X(\Theta)$ denotes the transformation that converts spherical (θ, ϕ) coordinates into Cartesian (x, y, z) coordinates.

After some simple computations, the corresponding matrices of first partial derivatives are found to be

$$\frac{\partial X(\Theta)}{\partial \Theta} = \begin{bmatrix} \cos\theta\cos\phi & -\sin\theta\sin\phi \\ \cos\theta\sin\phi & \sin\theta\cos\phi \\ -\sin\theta & 0 \end{bmatrix}$$

and

$$\frac{\partial \Theta(X)}{\partial X} = \begin{bmatrix} \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\csc\theta\sin\phi & \csc\theta\cos\phi & 0 \end{bmatrix} = \begin{bmatrix} \frac{xz}{\sqrt{x^2+y^2}} & \frac{yz}{\sqrt{x^2+y^2}} & -\sqrt{x^2+y^2} \\ \frac{-y}{x^2+y^2} & \frac{x}{x^2+y^2} & 0 \end{bmatrix}$$

The two columns of the Jacobian matrix $\partial X(\Theta)/\partial\Theta$ *define a parallelogram in the tangent plane to the hemisphere at point* (θ, ϕ)*. The Jacobian value is found as the length of their cross product.*

$$J_{X(\Theta)} = \sin\theta$$

The usual absolute value sign has been left off because $\sin\theta$ *is never negative on the positive hemisphere where* $0 \le \theta \le \pi/2$*. The Jacobian is zero, and hence the transformation is singular, only when* $\theta = 0$*. This corresponds to a single point at the pole of the hemisphere, and thus the transformation is only singular on a set of measure zero with respect to the whole hemispherical surface. The inverse Jacobian is easily computed as the reciprocal of the Jacobian, and rewritten in terms of Cartesian variables* x*,* y *and* z *to yield*

$$J_{\Theta(X)} = \frac{1}{J_{X(\Theta)}} = \frac{1}{\sin\theta} = \frac{1}{\sqrt{x^2+y^2}}$$

■

The simplest method for computing the Jacobian of a projection between the tangent plane $z = 1$ and the positive hemisphere is to break the transformation into two stages:

projecting from plane coordinates $U = (u, v)$ to spherical coordinates $\Theta = (\theta, \phi)$, then converting from spherical coordinates to Cartesian coordinates $X = (x, y, z)$. The first stage is a mapping from R^2 to R^2, so the Jacobian is easily computed. The Jacobian from spherical to Cartesian coordinates was worked out in the last example. By the chain rule, the Jacobian matrix of a two-stage transformation is found by composing the Jacobian matrices for each stage computed separately [Williamson68], i.e.

$$\frac{\partial X(U)}{\partial U} = \frac{\partial X(\Theta)}{\partial \Theta} \frac{\partial \Theta(U)}{\partial U} \quad \text{and} \quad \frac{\partial U(X)}{\partial X} = \frac{\partial U(\Theta)}{\partial \Theta} \frac{\partial \Theta(X)}{\partial X} \tag{C.16}$$

Similarly, the Jacobian value for a composition is computed by multiplying together the Jacobians from each stage, so that

$$J_{X(U)} = J_{X(\Theta)} \; J_{\Theta(U)} \quad \text{and} \quad J_{U(X)} = J_{U(\Theta)} \; J_{\Theta(X)} \tag{C.17}$$

What follows is a summary of the Jacobian computations for three common projections between the tangent plane $z = 1$ and the positive hemisphere. The three transformations are orthographic projection, central projection, and equal-area projection. Each transformation maps coordinates U of the plane $z = 1$ into spherical coordinates Θ on the positive hemisphere. The Jacobians for mapping between plane coordinates U and Cartesian coordinates X on the hemisphere are found by substituting these results, along with the results from Example C.10, into equations C.16 and C.17.

C.2.1 Orthographic Projection

Under orthographic projection, a point on the hemisphere is related to a point on the plane $z = 1$ by a line parallel to the z-axis. The set of points in the plane related to some point on the hemisphere lie within a circle of radius 1. Orthographic projection is singular for the (measure zero) set of points on the boundary of this circle, corresponding to the great circle of points $\theta = \pi/2$ on the boundary of the hemisphere.

$$U(\Theta) = \begin{cases} u = \sin\theta\cos\phi \\ v = \sin\theta\sin\phi \end{cases} \qquad J_{U(\Theta)} = \sin\theta\cos\theta$$

$$\Theta(U) = \begin{cases} \theta = \arcsin\sqrt{u^2+v^2} \\ \phi = \operatorname{arctan2}(v,u) \end{cases} \qquad J_{\Theta(U)} = \frac{1}{\sin\theta\cos\theta}$$

$$\frac{\partial U(\Theta)}{\partial\Theta} = \begin{bmatrix} \cos\theta\cos\phi & -\sin\theta\sin\phi \\ \cos\theta\sin\phi & \sin\theta\cos\phi \end{bmatrix} \qquad J_{U(X)} = \cos\theta = z$$

$$\frac{\partial\Theta(U)}{\partial U} = \begin{bmatrix} \sec\theta\cos\phi & \sec\theta\sin\phi \\ -\csc\theta\sin\phi & \csc\theta\cos\phi \end{bmatrix} \qquad J_{X(U)} = \frac{1}{\cos\theta} = 1/z$$

C.2.2 Central Projection

Under central projection, a point on the hemisphere is related to a point on the plane $z = 1$ by a line (projection ray) through the center of the sphere. Unlike orthographic projection, every point on the plane corresponds to some point on the hemisphere. The projection is singular only for the measure zero set of points on the boundary $\theta = \pi/2$ of the hemisphere, since the projection rays for those points are parallel to the plane.

$$U(\Theta) = \begin{cases} u = \tan\theta\cos\phi \\ v = \tan\theta\sin\phi \end{cases} \qquad J_{U(\Theta)} = \sin\theta / \cos^3\theta$$

$$\Theta(U) = \begin{cases} \theta = \arctan(\sqrt{u^2+v^2}) \\ \phi = \arctan2(v,u) \end{cases} \qquad J_{\Theta(U)} = \cos^3\theta / \sin\theta$$

$$\frac{\partial U(\Theta)}{\partial \Theta} = \begin{bmatrix} \sec^2\theta\cos\phi & -\tan\theta\sin\phi \\ \sec^2\theta\sin\phi & \tan\theta\cos\phi \end{bmatrix} \qquad J_{U(X)} = \frac{1}{\cos^3\theta} = \frac{1}{z^3}$$

$$\frac{\partial \Theta(U)}{\partial U} = \begin{bmatrix} \cos^2\theta\cos\phi & \cos^2\theta\sin\phi \\ -\cot\theta\sin\phi & \cot\theta\cos\phi \end{bmatrix} \qquad J_{X(U)} = \cos^3\theta = z^3$$

C.2.3 Equal-Area Projection

Equal-area projection is a one-to-one mapping of points on the hemisphere into a circle of radius $\sqrt{2}$ in the plane $z = 1$. Unlike orthographic and central projection, equal-area projection has no simple geometric interpretation. It is instead a projection method designed specifically to have a Jacobian value of 1 at every point, when considered as a projection between Cartesian hemisphere coordinates X and plane coordinates U. This property ensures that the area of surface patches on the hemisphere is preserved through the projection process. Note, for instance, that the area of a circle of radius $\sqrt{2}$ is 2π, precisely the surface area of the unit hemisphere itself. The mapping has no singularities.

Equal-area projection is ideal for displaying likelihood contours of a probability density function on the sphere since the area contained within the contour is preserved through the projection process. Equal-area projection is also used in Geology for making probabilistic inferences about random samples of points on the sphere, since the density of point clusters per unit area is preserved. For example, a set of points uniformly distributed on the hemisphere will map to a set of points uniformly distributed over a circular area of radius $\sqrt{2}$ [Mardia72].

$$U(\Theta) = \begin{cases} u = 2\sin\frac{\theta}{2}\cos\phi \\ v = 2\sin\frac{\theta}{2}\sin\phi \end{cases} \qquad J_{U(\Theta)} = \sin\theta$$

$$\Theta(U) = \begin{cases} \theta = 2\arcsin(\sqrt{u^2+v^2}/2) \\ \phi = \mathrm{arctan2}(v,u) \end{cases} \qquad J_{\Theta(U)} = \frac{1}{\sin\theta}$$

$$\frac{\partial U(\Theta)}{\partial\Theta} = \begin{bmatrix} \cos\frac{\theta}{2}\cos\phi & -2\sin\frac{\theta}{2}\sin\phi \\ \cos\frac{\theta}{2}\sin\phi & 2\sin\frac{\theta}{2}\cos\phi \end{bmatrix} \qquad J_{U(X)} = 1$$

$$\frac{\partial\Theta(U)}{\partial U} = \begin{bmatrix} \sec\frac{\theta}{2}\cos\phi & \sec\frac{\theta}{2}\sin\phi \\ -\frac{1}{2}\csc\frac{\theta}{2}\sin\phi & \frac{1}{2}\csc\frac{\theta}{2}\cos\phi \end{bmatrix} \qquad J_{X(U)} = 1$$

C.3 Mapping S^n to S^n

In Cartesian coordinates, the Jacobian matrix of a mapping from the surface of the sphere S^2 to itself is a 3×3 matrix. Unfortunately this matrix is always singular, since it is essentially a 3D representation of a 2D transformation. As it is not possible to compute a Jacobian value by taking the determinant of a singular Jacobian matrix, once again an appeal must be made to the underlying definition of the Jacobian as a magnification factor. The Jacobian matrix of a mapping that takes x_0 to y_0 on the surface of the sphere transforms a unit square in the tangent plane at point x_0 into a parallelogram in the tangent plane at point y_0. The area of the tangent parallelogram represents the change in differential area under the mapping at that point (Figure C.3).

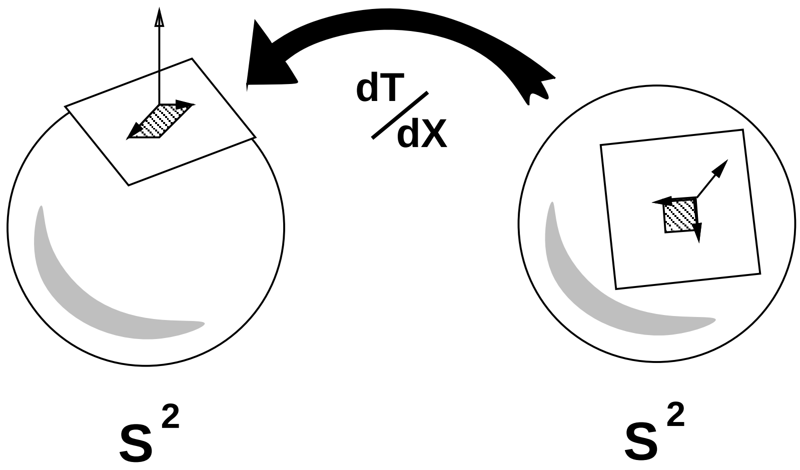


Figure C.3: The Jacobian of a transformation T from S^2 to S^2. The Jacobian matrix $\partial T/\partial X$ maps a unit square in the tangent plane at point x_0 on S^2 into a parallelogram in the tangent plane at point $T(x_0)$ on S^2. The change in area between the unit square and the tangent parallelogram is the Jacobian of transformation T at point x_0.

To compute the Jacobian value at point x_0 on the sphere, it suffices to choose two orthonormal vectors u_1 and u_2 (the unit square) in the tangent plane at point x_0, multiply them on the left by the Jacobian matrix to produce vectors a and b (the parallelogram) in the tangent plane at point y_0, then compute the length of the cross product $a \times b$ (the area of the parallelogram). Note further than if v_1 and v_2 are an orthonormal basis for the tangent plane at y_0, then

$$\begin{aligned} a &= \alpha_1 v_1 + \alpha_2 v_2 \\ b &= \beta_1 v_1 + \beta_2 v_2 \end{aligned} \quad \text{and} \quad ||a \times b|| = \left| det \begin{bmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \beta_2 \end{bmatrix} \right| \tag{C.18}$$

The Jacobian of transformation T at point x_0 on the sphere can therefore be computed as the absolute value of the determinant of a 2×2 matrix as follows

$$J_T(x_0) = \left| det \left\{ \begin{array}{ccc} 2 \times 3 & 3 \times 3 & 3 \times 2 \\ \begin{bmatrix} v_1^t \\ \\ v_2^t \end{bmatrix} & \frac{\partial T}{\partial X}(x_0) & \begin{bmatrix} u_1 & u_2 \end{bmatrix} \end{array} \right\} \right| \tag{C.19}$$

where v_1 and v_2 are any two orthogonal unit vectors perpendicular to $y_0 = T(x_0)$, $\frac{\partial T}{\partial X}(x_0)$ is the Jacobian matrix of transformation T, evaluated at point x_0, and u_1 and u_2 are any two orthogonal unit vectors perpendicular to x_0.

Example C.11 *The Jacobian of a projective transformation in the plane was computed in Example C.9 in order to determine how probability distributions in the plane were affected. This example considers the effect a projective transformation has over probability density functions on the sphere. An equivalent result for this problem has been presented previously, without derivation, in [Watson83].*

Let A be a nonsingular 3×3 matrix representing a plane-to-plane projective transformation in homogeneous coordinates. Identify each equivalence class of homogeneous coordinates with a line through the origin of R^3. This line intersects the unit sphere S^2 in a pair of antipodal points. Under a projective transformation, each infinite line through the origin is mapped to a new line, which intersects the unit sphere in another pair of antipodal points. The operation of A upon the sphere is said to be the transformation P that maps a pair *of antipodal points on the sphere before transformation A to the corresponding pair after transformation A.*

Transformation A is a linear transformation on R^3, mapping points on the surface of the sphere to points on an ellipsoid. This ellipsoid is mapped back onto the sphere by renormalizing each point vector to have unit length. Projective transformation A thus induces the following nonlinear mapping over points on the sphere

$$y \;=\; P(x) \;=\; \frac{Ax}{\|Ax\|}. \tag{C.20}$$

Notice that $P(-x) = -P(x)$ so that antipodal pairs are indeed mapped to antipodal pairs. Taking the derivatives of $P(x)$ with respect to vector x, the Jacobian matrix is found to be

$$\frac{\partial P}{\partial X} \;=\; \left[I_3 - \frac{(Ax)(Ax)^t}{\|Ax\|^2}\right] \frac{A}{\|Ax\|} \;=\; \left[I_3 - yy^t\right] \frac{A}{\|Ax\|}. \tag{C.21}$$

Plugging into equation (C.19)

$$J_P(x) \;=\; \left| det \left\{ \begin{bmatrix} v_1^t \\ v_2^t \end{bmatrix} \left[I_3 - yy^t\right] \frac{A}{\|Ax\|} \begin{bmatrix} u_1 & u_2 \end{bmatrix} \right\} \right| \tag{C.22}$$

where the u_i are arbitary orthonormal vectors perpendicular to x, and the v_i are arbitrary orthonormal vectors perpendicular to y. To simplify, note that

$$v_i^t \left[I_3 - yy^t\right] \;=\; v_i^t \tag{C.23}$$

since both vectors v_i are perpendicular to y by construction. Therefore

$$J_P(x) \;=\; \left| det \left\{ \begin{bmatrix} v_1^t \\ v_2^t \end{bmatrix} \frac{A}{\|Ax\|} \begin{bmatrix} u_1 & u_2 \end{bmatrix} \right\} \right| \;\equiv\; |\, det\, M \,| \quad . \tag{C.24}$$

It does not seem right that the value of the Jacobian should depend on the arbitrary choice of tangent plane basis vectors u_i and v_i. In fact, it does not. Form the orthonormal rotation matrices $R_x = [u_1\ u_2\ x]$ and $R_y = [v_1\ v_2\ y]$, and note that

$$R_y{}^t\ \frac{A}{\|Ax\|}\ R_x \quad = \quad \begin{bmatrix} \begin{bmatrix} M \end{bmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \text{---}\ \ \text{---} & 1 \end{bmatrix} . \tag{C.25}$$

Because of its special form, the determinant of this 3×3 matrix is same as M, therefore

$$J_P(x) \quad = \quad |\, det\, M\, | \quad = \quad \left| det \left\{ R_y{}^t\ \frac{A}{\|Ax\|}\ R_x \right\} \right| \quad = \quad \frac{|\, det\, A\, |}{\|Ax\|^3} . \tag{C.26}$$

Since A is nonsingular, the length of Ax can never be zero, and hence there are no singularities in a projective transformation on the sphere.

By similar reasoning, the Jacobian of the inverse projective transformation, specified in homogeneous coordinates by the matrix $B \equiv A^{-1}$, is

$$J_{P^{-1}}(y) \quad = \quad \frac{|\, det\, B\, |}{\|B\,y\|^3} . \tag{C.27}$$

Combining this with equation (C.7) shows that a probability density function $g(x)$ on the surface of the sphere is transformed into the new density function

$$h(y) \quad = \quad g\left(\frac{B\,y}{\|B\,y\|} \right) \frac{|\, det\, B\, |}{\|B\,y\|^3} . \tag{C.28}$$

■

BIBLIOGRAPHY

[AAI91] Amerinex Artificial Intelligence, *The KBVision System User's Guide,* Amerinex Artificial Intelligence, Amherst, MA, 1991.

[ASP80] American Society of Photogrammetry, *Manual of Photogrammetry,* Fourth Edition, American Society of Photogrammetry, Falls Church, VA, 1980.

[Anderson84] Anderson, T.W. "Estimating Linear Statistical Relationships," *The Annals of Statistics,* Vol. 12, No. 1, 1984, pp. 1–45.

[Ayache91] Ayache, N. *Artificial Vision for Mobile Robots,* MIT Press, Cambridge, MA, 1991.

[Ayres62] Ayres, F. *Matrices,* Shaum's Outline Series in Mathematics, McGraw-Hill, New York, 1962.

[Baddeley82] Baddeley, A. "Stochastic Geometry: An Introduction and Reading List," *International Statistical Review,* Vol. 50, 1982, pp. 179–193.

[Bagchi91] Bagchi, P. and Kadane, J.B. "Laplace Approximations to Posterior Moments and Marginal Distributions on Circles, Spheres, and Cylinders," *The Canadian Journal of Statistics,* Vol. 19, No. 1, 1991, pp. 67–77.

[Barnard83] Barnard, S.T. "Interpreting Perspective Images," *AI Journal,* Vol. 21, No. 4, November 1983, pp. 435–462.

[Bates88] Bates, D.M. and Watts, D.G. *Nonlinear Regression Analysis and its Applications,* John Wiley & Sons, New York, 1988.

[Beardsley92] Beardsley, P., Murray, D. and Zisserman, A. "Camera Calibration Using Multiple Images," *Proceedings of the Second European Conference on Computer Vision,* Santa Margherita Ligure, Italy, May 1992, pp. 312–320.

[Bellutta89] Bellutta, P., Collini, G., Verri, A. and Torre, V. "3D Visual Information from Vanishing Points," *IEEE Workshop on Interpretation of Three-Dimensional Scenes,* 1989, pp. 41–49.

[Benichou89] Benichou, J. and Gail, M.H. "A Delta Method for Implicitly Defined Random Variables," *The American Statistician,* Vol. 43, No. 1, 1989, pp. 41–44.

[Beran79] Beran, R. "Exponential Models for Directional Data," *The Annals of Statistics,* Vol. 7, No. 6, 1979, pp. 1162–1178.

[Bevington69] Bevington, P.R. *Data Reduction and Error Analysis for the Physical Sciences,* McGraw-Hill, New York, 1969.

[Bingham74] Bingham, C. "An Antipodally Symmetric Distribution on the Sphere," *The Annals of Statistics,* Vol. 2, 1974, pp. 1201–1225.

[Boldt89] Boldt, M., Weiss, R.S. and Riseman, E.M. "Token-Based Extraction of Straight Lines," *IEEE Transactions on Systems, Man and Cybernetics,* Vol. 19, No. 6, 1989, pp. 1581–1594.

[Boothby86] Boothby, W.M. *An Introduction to Differentiable Manifolds and Riemannian Geometry,* Second Edition, Academic Press, New York, 1986.

[Box73] Box, G.E. and Tiao, G.C. *Bayesian Inference in Statistical Analysis,* Addison-Wesley, Reading, MA, 1973.

[Breitenberger63] Breitenberger, E. "Analogues of the Normal Distribution on the Circle and the Sphere," *Biometrika,* Vol. 50, 1963, pp. 81–88.

[Brillault91] Brillault-O'Mahony, B. "A New Method for Vanishing Point Detection," *Computer Vision, Graphics, and Image Processing: Image Understanding,* Vol. 54, 1991, pp. 289–300.

[Britt73] Britt, H.I. and Lueke, R.H. "The Estimation of Parameters in Nonlinear, Implicit Models," *Technometrics,* Vol. 15, No. 2, 1973, pp. 233–246.

[Brooks84] Brooks, R. "Aspects of Mobile Visual Map Making," *Proceedings of the Second International Symposium on Robotics Research,* Kyoto, Japan, 1984.

[Buchanan92] Buchanan, T. "Critical Sets for 3D Reconstruction using Lines," *Proceedings of the Second European Conference on Computer Vision,* Santa Margherita Ligure, Italy, May 1992.

[Burns86] J Burns, J.B., Hanson, A.R. and Riseman, E.M. "Extracting Straight Lines," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 8, No. 4, July 1986, pp. 425-456.

[Burns93] Burns, J.B., Weiss, R.S. and Riseman, E.M. "The Non-Existence of General-Case View-Invariants," in *Geometric Invariance in Computer Vision,* Edited by J.L. Mundy and A. Zisserman, MIT Press, Cambridge, MA, pp. 120–131.

[Caprile90] Caprile, B. and Torre, V. "Using Vanishing Points for Camera Calibration," *International Journal of Computer Vision,* Vol. 4, 1990, pp. 127–140.

[Clutton-Brock90] Clutton-Brock, M. "Density Estimation using Exponentials of Orthogonal Series," *Journal of the American Statistical Association,* Vol. 85, 1990, pp. 760–764.

[Coelho92] Coelho, C., Heller, A., Mundy, J.L., Forsyth, D.A. and Zisserman, A. "An Experimental Evaluation of Projective Invariants," in *Geometric Invariance in Computer Vision,* Edited by J.L. Mundy and A. Zisserman, MIT Press, Cambridge, MA, 1992.

[Collins89] Collins, R.T. and Weiss, R.S. "An Efficient and Accurate Method for Computing Vanishing Points," Topical Meeting on Image Understanding and Machine Vision, 1989 Technical Digest Series, Vol. 14, Optical Society of America, Washington, D.C., pp. 92–94.

[Collins90] Collins, R.T. and Weiss, R.S. "Vanishing Point Calculation as a Statistical Inference on the Unit Sphere," *Proceedings of the Third International Conference on Computer Vision,* Osaka, Japan, December 1990, pp. 400–403.

[Collins92] Collins, R.T. "Single Plane Model Extension using Projective Transformations," *Proceedings of the Darpa Image Understanding Workshop,* San Diego, CA, January 1992, pp. 917–923.

[Crowley87] Crowley, J.L. and Ramparany, F. "Mathematical Tools for Representing Uncertainty in Perception", *Proceedings of the Spatial Reasoning and Multi-Sensor Fusion Workshop,* Edited by Avi Kak and Su-shing Chen, 1987, pp. 293–302.

[Demey92] Demey, S., Zisserman, A. and Beardsley, P. "Affine and Projective Structure from Motion," *Proceedings of the British Machine Vision Conference,* 1992, pp. 49–58.

[Dennery69] Dennery, P. and Krzywicki, A. *Mathematics for Physicists,* Harper & Row, New York, 1969.

[Dolby76] Dolby, G.R. "The Connection Between Methods of Estimation in Implicit and Explicit Nonlinear Models," *Applied Statistics,* Vol. 25, No. 2, 1976, pp. 157–162.

[Downs72] Downs, T.D. "Orientation Statistics," *Biometrika,* Vol. 59, No. 3, 1972, pp. 665-676.

[Durrant-Whyte88] Durrant-Whyte, H.F. "Uncertain Geometry," *IEEE Journal of Robotics and Computer Vision,* Vol. 4, 1988. Reprinted in *Geometric Reasoning,* Edited by J.L. Mundy and D. Kapur, MIT Press, Cambridge, MA, 1989, pp. 447–481.

[Eaton83] Eaton, M.L. *Multivariate Statistics – A Vector Space Approach,* John Wiley & Sons, New York, 1983.

[Efron86] Efron, B. "Why Isn't Everyone a Bayesian?," *The American Statistician,* Vol. 40, No. 1, 1986. Reprinted in *Readings in Uncertain Reasoning,* edited by G. Shafer and J. Pearl, Morgan Kaufmann, San Mateo, CA, 1990, pp. 21–31.

[Erdélyi53] Erdélyi, A. *et.al. Higher Transcendental Functions,* The Bateman Manuscript Project, Vol. 2, McGraw-Hill, New York, 1953.

[Faugeras88] Faugeras, O.D. and Lustman, F. "Motion and Structure from Motion in a Piecewise Planar Environment," *International Journal of Pattern Recognition and Artificial Intelligence,* Vol. 2, 1988, pp. 485–508.

[Faugeras90] Faugeras, O.D. and Maybank, S.J. "Motion from Point Matches: Multiplicity of Solutions," International Journal of Computer Vision, Vol. 4, 1990, pp. 225-246.

[Faugeras92a] Faugeras, O.D. "What Can be Seen in Three Dimensions with an Uncalibrated Stereo Rig?," *Proceedings of the Second European Conference on Computer Vision,* Santa Margherita Ligure, Italy, May 1992, pp. 563–578.

[Faugeras92b] Faugeras, O.D., Luong, Q.T. and Maybank, S.J. "Camera Self-Calibration: Theory and Experiments," *Proceedings of the Second European Conference on Computer Vision,* Santa Margherita Ligure, Italy, May 1992, pp. 321–334.

[Forsyth91] Forsyth, D., Mundy, J., Zisserman, A., Coelho, C. and Heller, A. "Invariant Descriptors for 3D Object Recognition and Pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 13, 1991, pp. 971–991.

[Gee93] Gee, S.J. and Newman, A.M. "RADIUS : Automating Image Analysis Through Model-Supported Exploitation," *Proceedings of the Darpa Image Understanding Workshop,* Washington, DC, April 1993, pp. 185–196.

[Gelb86] Gelb, A. *et.al., Applied Optimal Estimation,* MIT Press, Cambridge, MA, 1986.

[Gerson92] Gerson, D.J. "RADIUS : The Government Viewpoint," *Proceedings of the Darpa Image Understanding Workshop,* San Diego, CA, January 1992, pp. 173–175.

[Grimson88] Grimson, W.E.L. and Huttenlocher, D.P. "On the Sensitivity of the Hough Transform for Object Recognition," *Proceedings of the Second International Conference on Computer Vision,* 1988, pp. 700–706.

[Grimson90] Grimson, W.E.L. *Object Recognition by Computer: The Role of Geometric Constraints,* MIT Press, Cambridge, MA, 1990.

[Grosky90] Grosky, W.I. and Tamburino, L.A. "A Unified Approach to the Linear Camera Calibration Problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 12, 1990, pp. 663–671.

[Haralick80] Haralick, R.M. "Using Perspective Transformations in Scene Analysis," *Computer Graphics and Image Processing,* Vol. 13, 1980, pp. 191–221.

[Harter83] Harter, H.L. "Least Squares," *Encyclopedia of Statistical Sciences,* Vol. 4, Edited by S. Kotz and N.L. Johnson, John Wiley and Sons, NY, 1983, pp. 593–598.

[Hartley93] Hartley, R., Gupta, R. and Chang, T. "Stereo from Uncalibrated Cameras," *Proceedings IEEE Computer Vision and Pattern Recognition,* Champaign, IL, June 1992, pp. 761–764.

[Hel-Or92] Hel-Or, Y. and Werman, M. "Absolute Orientation from Uncertain Point Data: A Unified Approach," *Proceedings IEEE Computer Vision and Pattern Recognition,* Champaign, IL, June 1992, pp. 77–82.

[Horaud87] Horaud, R. "New Methods for Matching 3D Objects with Single Perspective Views," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 9, 1987, pp. 401–412.

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

[Huffman71] Huffman, D.A. "Impossible Objects as Nonsense Sentences," in *Machine Intelligence 6,* Edited by B. Metzler and D. Michie, American Elsevier, New York, 1971, pp. 295–323.

[Jupp79] Jupp, P.E. and Mardia, K.V. "Maximum Likelihood Estimators for the Matrix Von Mises-Fisher and Bingham Distributions," *The Annals of Statistics,* Vol. 7, No. 3, pp. 599–606.

[Jupp91] Jupp, P.E. and Mardia, K.V. "A Unified View of the Theory of Directional Statistics : 1975–1988," *International Statistical Review,* Vol. 57, pp. 261–294.

[Kanatani91] Kanatani, K. "Computational Projective Geometry," *Computer Vision, Graphics, and Image Processing : Image Understanding,* Vol. 54, 1991, pp. 333–348.

[Kanatani92] Kanatani, K. "Statistical Analysis of Focal Length Calibration using Vanishing Points," *IEEE Transactions on Robotics and Automation,* Vol. 8, No. 6, 1992.

[Kanatani93a] Kanatani, K. "Renormalization for Unbiased Estimation," *Proceedings of the Fourth International Conference on Computer Vision,* Berlin, Germany, May 1993, pp. 599–606.

[Kanatani93b] Kanatani, K. "Statistical Analysis of Geometric Computation," *Computer Vision, Graphics, and Image Processing : Image Understanding,* to appear.

[Kender80] Kender, J.R. *Shape from Texture,* Ph.D. Thesis, Department of Computer Science, Carnegie-Mellon University, Schenley Park, Pittsburgh, PA, 1980.

[Kent87] Kent, J.T. "Asymptotic Expansions for the Bingham Distribution," *Applied Statistics,* Vol. 36, No. 2, 1987, pp. 139–144.

[Kumar90] Kumar, R. and Hanson, A.R. "Pose Refinement: Application to Model Extension and Sensitivity to Camera Parameters," *Proceedings of the Darpa Image Understanding Workshop,* Pittsburgh, PA, September 1990, pp. 660–669.

[Kumar92] Kumar, R. *Model Dependent Inference of 3D Information from a Sequence of 2D Images,* Ph.D. Thesis, Department of Computer Science, University of Massachusetts, Amherst, MA, February 1992.

[Lenth81] Lenth, R.V. "Robust Measures of Location for Directional Data," *Technometrics,* Vol. 23, 1981, pp. 77–81.

[Li90] Li, S., Tsuji, S. and Imai, M. "Determining Camera Rotation from Vanishing Points of Lines on Horizontal Planes," *Proceedings of the Third International Conference on Computer Vision,* Osaka, Japan, December 1990, pp. 499–502.

[Lindley65] Lindley, D.V. *Introduction to Probability and Statistics from a Bayesian Viewpoint, Part2 – Inference* Cambridge University Press, Cambridge, 1965.

[Luong93] Luong, Q.T. and Faugeras, O.D. "Determining the Fundamental Matrix with Planes: Instability and New Algorithms," *Proceedings IEEE Computer Vision and Pattern Recognition,* New York City, 1993, pp. 489–494.

[Mackworth73] Mackworth, A.K. "Interpreting Pictures of Polyhedral Scenes," *Artificial Intelligence,* Vol. 4, No. 2, 1973, pp. 121–137.

[Magee84] Magee, M.J. and Aggarwal, J.K. "Determining Vanishing Points from Perspective Images," *Computer Vision, Graphics, and Image Processing,* Vol. 26, 1984, pp. 256-267.

[Mardia72] Mardia, K.V. *Statistics of Directional Data,* Academic Press, New York, 1972.

[Mardia77] K Mardia, K.V. and Zemroch, P.J. "Table of Maximum Likelihood Estimates for the Bingham Distribution," *Journal of Statistical Computation and Simulation,* Vol. 6, 1977, pp. 29–34.

[MM93] Martin Marietta and SRI International, *RCDE User's Guide,* Martin Marietta, Management and Data Systems, Philadelphia, PA, 1993.

[Matthies87a] Matthies, L. and Kanade, T. "The Cycle of Uncertainty and Constraint in Robot Perception," *Proceedings of the International Symposium on Robotics Research,* MIT Press, Cambridge, MA, 1987.

[Matthies87b] Matthies, L. and Shafer, S.A. "Error Modeling in Stereo Navigation," *Journal of Robotics and Automation,* Vol. 3, 1987, pp. 239–248.

[Maybank90a] Maybank, S.J. "The Projective Geometry of Ambiguous Surfaces," *Philosophical Transactions of the Royal Society of London, Series A,* Vol. 332, 1990, pp. 1–47.

[Maybank90b] Maybank, S.J. "Filter Based Estimates of Depth," *Proceedings of the British Machine Vision Conference,* University of Oxford, September 1990.

[Maybank93] Maybank, S.J. "Error Trade Offs for the Cross Ratio in Model Based Vision," to appear, *Workshop on Computer Vision for Space Applications,* Antibes, France, September 1993.

[McCulloch88] McCulloch, R.E. "Information and the Likelihood Function in Exponential Families," *The American Statistician,* Vol. 42, No. 1, 1988, pp. 73–75.

[Mendel87] Mendel, J.M. *Lessons in Digital Estimation Theory,* Prentice-Hall Signal Processing Series, Prentice-Hall, Inc., NJ, 1987.

[Mihalek72] Mihalek, R.J. *Projective Geometry and Algebraic Structures,* Academic Press, New York, 1972.

[Minsky67] Minsky, M. "Stereo and Perspective Calculations," Project Mac Technical Report, AI Memo No. 143, Massachusetts Institute of Technology, Cambridge, MA, September 1967.

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

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

[Mohr92] Mohr, R., Quan, L., Veillon, F. and Boufama, B. *Relative 3D Reconstruction using Multiple Uncalibrated Images,* Technical Report, IRIMAG-LIFIA, 1992.

[Muirhead82] Muirhead, R.J. *Aspects of Multivariate Statistical Theory,* John Wiley & Sons, New York, 1982.

[Müller66] Müller, C. *Spherical Harmonics,* Lecture Notes in Mathematics Series, No. 17, Springer-Verlag, New York, 1966.

[Mundy92a] Mundy, J.L. and Zisserman, A. editors, *Geometric Invariance in Computer Vision,* MIT Press, Cambridge, MA, 1992.

[Mundy92b] Mundy, J.L. and Zisserman, A. "Projective Geometry for Machine Vision," in *Geometric Invariance in Computer Vision,* Edited by J.L. Mundy and A. Zisserman, MIT Press, Cambridge, MA, 1992.

[Naeve87] Naeve, A. and Eklundh, J. "On Projective Geometry and the Recovery of 3-D Structure," *Proceedings of the First International Conference on Computer Vision,* London, England, June 1987, pp. 128–135.

[Papoulis65] Papoulis, A. *Probability, Random Variables, and Stochastic Processes,* McGraw-Hill, New York, 1965.

[Pearson01] Pearson, K. "On Lines and Planes of Closest Fit to Systems of Points in Space," *Philosophy Magazine,* Vol. 2, 1901, pp. 559–572.

[Pickert83] Pickert, G., Stendor, R. and Hellwich, M. "From Projective to Euclidean Geometry," in *Fundamentals of Mathematics,* Edited by H. Behnke, *el.al.,* Translated by S.H. Gould, MIT Press, Cambridge, MA, 1983.

[Prentice82] Prentice, M.J. "Antipodally Symmetric Distributions for Orientation Statistics," *Journal of Statistical Planning and Inference,* Vol. 6, 1982, pp. 205–214.

[Prentice84] Prentice, M.J. "A Distribution-Free Method of Interval Estimation for Unsigned Directional Data," *Biometrika,* Vol. 71, 1984, pp. 147–154.

[Press90] W.H. Press, Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. *Numerical Recipes in C – The Art of Scientific Computing,* Cambridge University Press, New York, 1990.

[Quan89] Quan, L. and Mohr, R. "Determining Perspective Structures Using Hierarchical Hough Transform," *Pattern Recognition Letters,* Vol. 9, 1989, pp. 279–286.

[Reilly81] Reilly, P.M. and Patino-Leal, H. "A Bayesian Study of the Errors-in-Variables Model," *Technometrics,* Vol. 23, No. 3, 1981, pp. 221–231.

[Rivlin69] Rivlin, T.J. *An Introduction to the Approximation of Functions,* Blaisdell Publishing Company, Waltham, MA, 1969.

[Roberts65] Roberts, L.G. "Machine Perception of Three-Dimensional Solids," *Optical and Electro-Optical Information Processing,* Edited by J.T. Tippett *et.al.,* MIT Press, Cambridge, MA, 1965.

[Rothwell92] Rothwell, C.A., Zisserman, A., Mundy, J.L. and Forsythe, D.A. "Efficient Model Library Access by Projectively Invariant Indexing Functions," *Proceedings IEEE Computer Vision and Pattern Recognition,* Champaign, IL, June 1992, pp. 109–114.

[Rothwell93] Rothwell, C.A., Forsythe, D.A., Zisserman, A. and Mundy, J.L. "Extracting Projective Structure from Single Perspective Views of 3D Point Sets," *Proceedings of the Fourth International Conference on Computer Vision,* Berlin, Germany, May 1993, pp. 573–582.

[Schwetlick85] Schwetlick, H. and Tiller, V. "Numerical Methods for Estimating Parameters in Nonlinear Models with Errors in the Variables," *Technometrics,* Vol. 27, No. 1, 1985, pp. 17–24.

[Seber89] Seber, G.A.F. and Wild, C.J. *Nonlinear Regression,* Wiley & Sons, New York, 1989.

[Shafer83] Shafer, S.A., Kanade, T. and Kender, J.R. "Gradient Space under Orthography and Perspective," *Computer Vision, Graphics, and Image Processing,* Vol. 24, No. 2, 1983, pp.182–199.

[Shakunaga92] Shakunaga, T. "3D Corridor Scene Modeling from a Single View under Natural Lighting Conditions," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 14, 1992, pp. 293–298.

[Shashua93] Shashua, A. "Projective Depth: A Geometric Invariant for 3D Reconstruction from Two Perspective/Orthographic Views and for Visual Recognition," *Proceedings of the Fourth International Conference on Computer Vision,* Berlin, Germany, May 1993, pp. 583–590.

[Springer64] Springer, C.E. *Geometry and Analysis of Projective Spaces,* W.H. Freeman and Company, San Francisco, 1964.

[Stein92] Stein, A. and Werman, M. "Robust Statistics in Shape Fitting," *Proceedings IEEE Computer Vision and Pattern Recognition,* Champaign, IL, June 1992, pp. 540–546.

[Stoyan87] Stoyan, D., Kendall, W.S. and Mecke, J. *Stochastic Geometry and it Applications,* John Wiley and Sons, New York, 1987.

[Straforini92] Straforini, M., Coelho, C., Compani, M., and Torre, V. "The Recovery and Understanding of a Line Drawing from Indoor Scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 14, 1992, pp. 298–303.

[Thomas92] Thomas, J.I. and Oliensis, J. "Recursive Multi-Frame Structure from Motion Incorporating Motion Error," *Proceedings of the Darpa Image Understanding Workshop,* San Diego, CA, January 1992, pp. 507–513.

[Tsai82] Tsai, R.Y., Huang, T.S. and Zhu, W. "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.

[Tyler87] Tyler, D.E. "Statistical Analysis for the Angular Central Gaussian Distribution on the Sphere," *Biometrika,* Vol. 74, No. 3, 1987, pp. 579–589.

[Upton89] Upton, G.J.G. and Fingleton, B. *Spatial Data Analysis by Example,* Volume 2, "Categorical and Directional Data," John Wiley and Sons, New York, 1989.

[Waltz72] Waltz, D. "Generating Semantic Descriptions from Drawings of Scenes with Shadows," AI-TR-271, Project MAC, Massachusetts Institute of Technology, Cambridge, MA, 1972. Reprinted in *The Psychology of Computer Vision,* Edited by P. Winston, McGraw-Hill, New York, 1975, pp. 295–323.

[Wang91] Wang, L.L. and Tsai, W.H. "Camera Calibration by Vanishing Lines for 3-D Computer Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 13, 1991, pp. 370–376.

[Watson83] Watson, G.S. "More Significance Tests on the Sphere," *Biometrika,* Vol. 47, 1960, pp. 87–91.

[Watson83] Watson, G.S. *Statistics on Spheres,* John Wiley and Sons, New York, 1983.

[Weiss90] Weiss, R.S., Nakatani, H. and Riseman, E.M. "An Error Analysis for Surface Orientation from Vanishing Points," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 12, 1990, pp. 1179–1185.

[Williamson68] Williamson, R.E. R.H. Crowell, and H.F. Trotter, *Calculus of Vector Functions,* Prentice-Hall Inc., New Jersey, 2nd edition, 1968.

[Wolff89] Wolff, L.B. and Boult, T.E. "Using Line Correspondence Stereo to Measure Surface Orientation," *Proceedings Eleventh International Joint Conference on Artificial Intelligence,* Detroit, Michigan, August 1989, pp. 1655–1660.

[Zacks81] Zacks, S. *Parametric Statistical Inference : Basic Theory and Modern Approaches,* Pergamon Press, New York, 1981.