SRI International

INTERPRETING PERSPECTIVE IMAGES

Technical Note 271

November 1982

By: Stephen T. Barnard, Computer Scientist

Artificial Intelligence Center
Computer Science and Technology Division

SRI Project 1009

The research reported herein was supported at SRI by the Defense Advanced Research Projects Agency under Contract MDA903-79-C-0588.

SRI International®

333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

INTERPRETING PERSPECTIVE IMAGES.

Stephen T. Barnard

Artificial Intelligence Center, SRI International
Menlo Park, California 94025

ABSTRACT

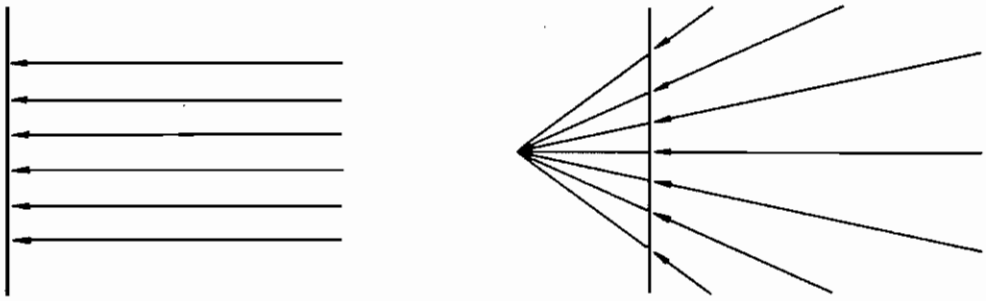
A fundamental problem in computer vision is how to determine the 3-D spatial orientation of curves and surfaces appearing in an image. The problem is generally underconstrained, and is complicated by the fact that metric properties, such as orientation and length, are not invariant under projection. Under perspective projection (the correct model for most real images) the transform is nonlinear, and therefore hard to invert. Two constructive methods are presented. The first finds the orientation of parallel lines and planes by locating vanishing points and vanishing lines. The second determines the orientation of planes by "backprojection" of two intrinsic properties of contours: angle magnitude and curvature.

1 Introduction

A computational theory of vision must explain a very puzzling aspect of human visual experience. How is it that we correctly perceive three-dimensional properties of objects in space from two-dimensional projections (e.g., a single image)? At first it seems that essential information for depth is lost when the retinal image is formed: a ray of light may as well have come from a star light-years distant as from across the room. Nevertheless, we have definite impressions of the distances and orientations of the things we see, even when there is no explicit, unambiguous information about these three-space relations in the image.

There are a few purely physical mechanisms that can account for some modes of spatial perception — in particular, accommodation of the lens for focusing at different distances, binocular stereopsis, and optic flow. But while these mechanisms may account for some spatial perception, their explanation remains insufficient and incomplete. We usually have no trouble interpreting single images with substantial ranges of depth, or even simple line drawings with an infinite number of possible interpretations. Since information is lost in projecting a three-dimensional scene onto a two-dimensional surface, some form of computational "cognitive" model is required to construct percepts from ambiguous, incomplete, and noisy images.

Three important spatial properties that we perceive are **size , shape,** and **depth.** Size and shape are fundamentally different from depth because they are defined relative to an **object,** while depth is defined relative to an **observer.** Size is usually measured with ordinary Euclidean metrics: length, area, and volume. It is difficult to give a precise definition of shape, but the essential principle is that the shape of an object is the spatial



(a) PARALLEL PROJECTION (b) CENTRAL PROJECTION

FIGURE 1 PARALLEL AND CENTRAL PROJECTION

arrangement of the contours and surfaces of which it is composed. While size is independent of the choice of a coordinate system, shape usually is not. Shape is often specified in some "natural" object-centered coordinate system that is selected and aligned to match the symmetry of the object.

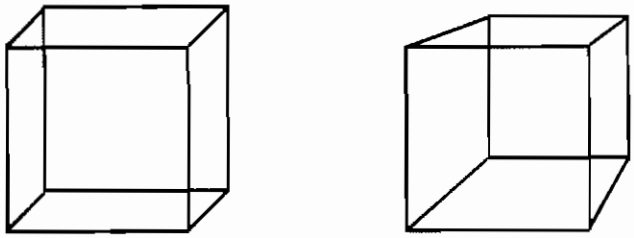
In much of the following discussion we shall assume that shapes can be described adequately by straight lines and planes. These primitive shape descriptors are the simplest geometrical contours and surfaces we can hope to find. They are common in scenes containing man-made objects, less common in natural scenes. If we can develop computational methods for the perception of lines and planes, we can perhaps generalize them to include more complex shapes. In Section 4.3 curved planar contours will be considered.

To recover 3-D shape from 2-D projections, an explicit model of the projective transform is essential. Two models are common: parallel and central projection (Figure 1). In parallel projection an image is formed by parallel rays, usually perpendicular to the image plane. In central projection an image is formed by rays passing through a common point in space called the focal point. The parallel projective transform is called "orthographic," the central projective transform "perspective."

It is important to emphasize that central projection is the correct model both for human vision and for cameras, whereas parallel projection is only an approximation.

The most important parameter that distinguishes perspective from orthographic projection is the included angle of view, which is defined to be the maximum angle between two rays (i.e., the angle between the two rays with the greatest angular separation). The assumption of orthographic projection is essentially equivalent to the assumption of zero included angle of view. Locally, perspective projection is approximately orthographic because the included angle of view is small. When the entire image is considered, however, perspective becomes important.

If the focal length (the perpendicular distance from the focal point to the image plane) is large, compared with the linear dimension of the image, the included angle of view is small and the orthographic approximation is reasonable. Photographs taken with "normal" lenses for a certain film format (e.g., a 50-mm lens on a 35-mm camera) typically cover about 45 degrees of view, and perspective effects are often quite apparent. If a wide-angle lens is used, perspective is dominant and the picture may appear distorted, although the "distortion" is merely the result of viewing the photograph from the wrong distance.



(a) ORTHOGRAPHIC PROJECTION (b) PERSPECTIVE PROJECTION

FIGURE 2 NECKER CUBE ILLUSIONS

Another parameter that causes perspective images to differ from orthographic ones, even when the included angle of view is small, is the ratio of the distances to objects in the scene (or, informally stated, the ratios of "depths" of points). Under perspective, the projected area of an object varies inversely with the object's distance from the focal point (measured along the principal ray, defined to be the ray perpendicular to the image plane). Under orthography, however, the size of an object in an image is independent of depth.

The perspective camera model, therefore, will be required for accurate recovery of size, shape, and depth whenever the image covers a substantial included angle of view, or whenever objects at very different depths are compared.

The difference between orthographic and perspective projection is not only quantitative, but also qualitative. In Figure 2 one of the most familiar of all illusions — the Necker cube — is shown in parallel and central projection. In both cases the images are highly ambiguous because they could have been produced by an infinite number of objects; nevertheless, in each case we perceive only two distinct interpretations. The interpretations of the orthographic image are more or less equally preferable because both have the same symmetry. The interpretations of the perspective image, however, are radically different: one is a symmetrical cube, while the other is a relatively asymmetrical octohedron. There are other qualitative differences between orthography and perspective. For example, vanishing points and vanishing lines are not found in orthographic projections, but are characteristic of perspective projections. (This topic will be covered in detail later.)

The use of explicit models of the projective transform has a long history in computer vision. Mackworth used the concept of **gradient space** [1], based on Huffman's **dual space** [2], to interpret line drawings of polyhedral scenes. For reasons that will be made clear in Section 2, perspective involves more difficult mathematics than does orthography. (See Haralick for a discussion of the mathematics of perspective [3].) Most computer vision approaches, therefore, begin with the assumption of parallel projection. The general approach involves representing physical constraints on the scene as relations in gradient space. Horn, for example, used this approach in his analysis of "shape from shading" [4]. An overview of gradient-space methods can be found in [5].

In this work we shall treat the perspective projection exclusively. We shall use the Gaussian sphere instead of gradient space as a domain for representing geometric constraints. Analytical solutions, although possible in principle, appear to be very difficult, so we will

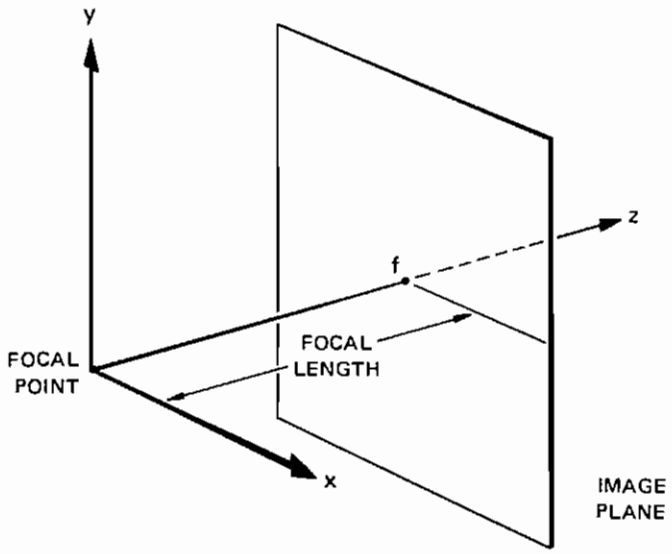


FIGURE 3 STANDARD COORDINATE SYSTEM

present constructive methods that explicitly generate constraints over the entire sphere.

2 Mathematics of Perspective

2.1 Review of Algebraic Models of Projection

The standard coordinate system we shall use is illustrated in Figure 3. The origin is at the focal point, the x-axis is parallel to "right" in the image plane, the y-axis is parallel to "up" in the image plane, and the z-axis is directed through the principal point in the image plane. The focal length is f. Note that this is a left-handed coordinate system. The image plane is one focal length from the origin along the z-axis.

Parallel (orthographic) projection can be represented as a simple linear transform. Given a point $\mathbf{p} = (x, y, z)$, the parallel projection $\mathbf{p}_o$ of $\mathbf{p}$ is given by:

$$\mathbf{p}_o = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{p}^T + (0, 0, f) \ . \tag{2.1}$$

Central (perspective) projection, on the other hand, is an essentially nonlinear transform: image coordinates are determined by dividing scene coordinates by the depth as measured along the principal ray. The central projection $\mathbf{p}_p$ of $\mathbf{p}$ is

$$\mathbf{p}_p = (\frac{xf}{z}, \frac{yf}{z}, f) \ . \tag{2.2}$$

The perspective transform can be expressed elegantly with homogeneous coordinates. Homogeneous coordinates were first developed as an analytical tool for projective geometry [6]; more recently they have been used effectively in computer graphics [7] and industrial automation [8]. The homogeneous coordinates of a point are represented by a four-tuple $\langle x, y, z, w \rangle$, and the ordinary three-dimensional coordinates of the point are obtained by $\langle \frac{x}{w}, \frac{y}{w}, \frac{z}{w} \rangle$. One advantage of using homogeneous coordinates in projective geometry derives from the fact that points "at infinity" are represented as four-tuples with $w = 0$, whereas in ordinary coordinates they have no representation.

The formulation of perspective projection in homogeneous coordinates is as follows. First, we must use a slightly different coordinate system than the standard one illustrated in Figure 3. In the new coordinate system the image plane is the xy plane and the focal point is at $\langle 0, 0, -f \rangle$. A point $\mathbf{P}$ with homogeneous coordinates $\langle x, y, z, w \rangle$ is projected onto the image plane at point $\mathbf{q}$ with ordinary coordinates $\mathbf{q} = \langle \frac{xf}{z+fw}, \frac{yf}{z+fw}, 0 \rangle$. This projection can be expressed in matrix form as

$$\begin{aligned} \mathbf{P}' &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{f} & 1 \end{pmatrix} \mathbf{P}^T \\ &= \langle x, y, z, \frac{z + wf}{f} \rangle \ , \end{aligned} \tag{2.3}$$

followed by conversion to ordinary coordinates

$$\mathbf{p}' = \langle \frac{xf}{z + wf}, \frac{yf}{z + wf}, \frac{zf}{z + wf} \rangle \ , \tag{2.4}$$

and finally a "parallel" projection transform

$$\mathbf{p}_p = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{p}'^T \ . \tag{2.5}$$

Clearly, parallel projection is a special case of central projection. If the focal length is infinite Equation (2.3) becomes the identity transform.

2.2 Gaussian Mapping

In the next two sections we shall use Gaussian mapping [9] to represent the orientation of lines and planes. Gaussian mapping transforms vectors in 3-space into points on a unit sphere centered at the origin (Figure 4). It can be used to represent the orientation of either lines (if the vectors are interpreted as direction cosines) or planes (if the vectors are interpreted as planar normals). Since all parallel vectors in space map to the same point on the sphere, any point on the sphere represents a family of parallel vectors.

Suppose $\mathbf{n}$ is a point on the Gaussian sphere. It can be expressed in either spherical or Cartesian coordinates. The azimuth of $\mathbf{n}$, α, is the angle measured from the z-axis in the xz-plane. The elevation of $\mathbf{n}$, β, is the angle measured from the xz-plane toward the y-axis.

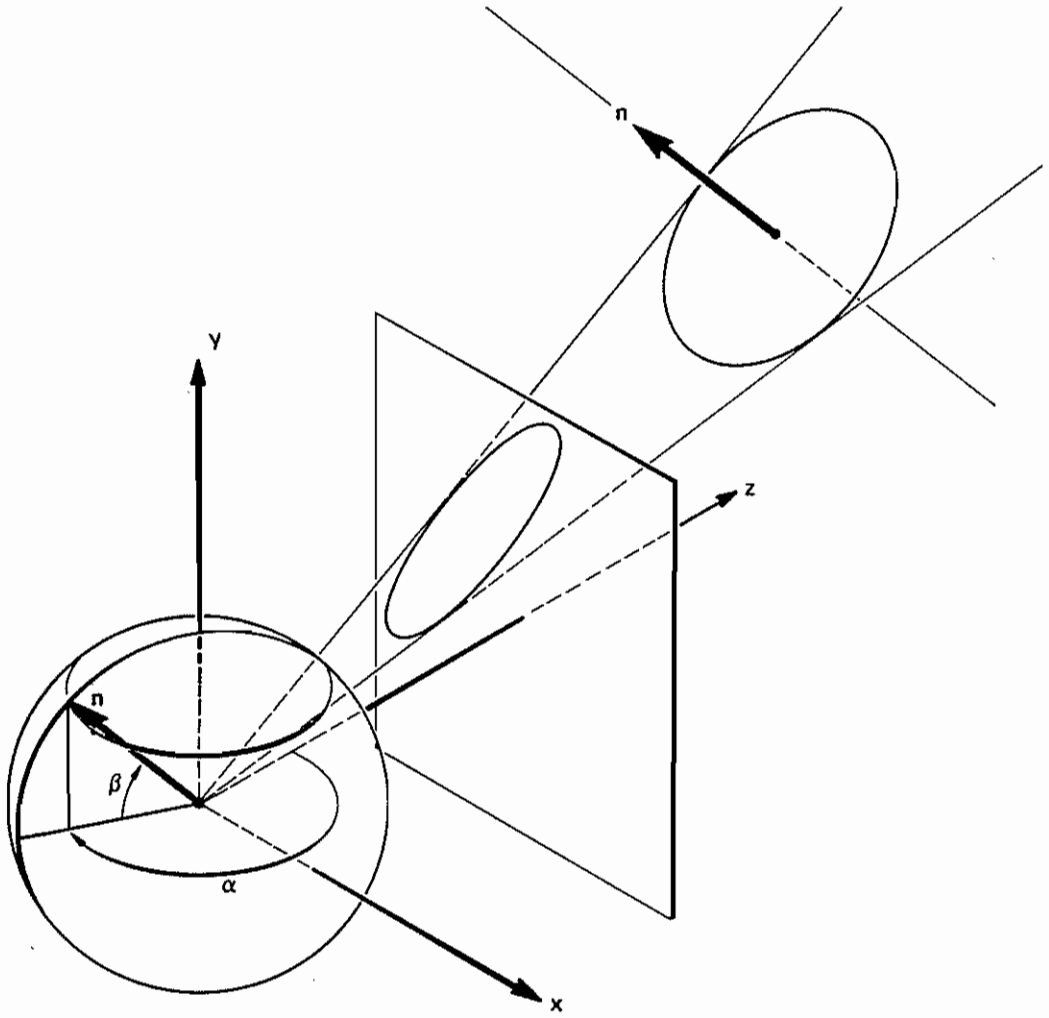


FIGURE 4 GAUSSIAN MAPPING

In Cartesian coordiantes,

$$\mathbf{n} = (\sin\alpha\cos\beta, \sin\beta, \cos\alpha\cos\beta) \ . \tag{2.2.1}$$

The Gaussian sphere is a more convenient domain in which to express geometric constraints under perspective than the more widely used gradient space. All orientations are represented on the sphere, while orientations perpendicular to the line of sight are undefined in gradient space. The sphere is a finite (i.e., closed) space, whereas gradient space is infinite (i.e., open), and therefore more difficult to represent directly in a finite computer memory. More fundamentally, the Gaussian sphere exhibits the same symmetry as the central projection (i.e., it is symmetric with respect to the focal point), while gradient space exhibits the symmetry of orthographic projection (i.e., it is symmetric with respect

to the line of sight). In fact, gradient space is the limiting case of the Gaussian sphere as the focal length approaches infinity and the central projection approaches orthography.

There is an interesting dual relationship between lines and planes in projective space: two lines determine a plane, and two planes determine a line. There is a similar dual relationship on the Gaussian sphere between points and lines (i.e., great circles): a point on the sphere determines a pole through the origin; the dual of the point is the equator associated with the pole. Analogous dual relationships are found in gradient space. Duality can be exploited in the interpretation of perspective images, as we shall see in Section 3.

2.3 A Computational Approach

Geometric properties can be divided into two classes: metric, such as the length and orientation of lines; and descriptive, such as the colinearity of three points or the coincidence of three lines. Metric properties are in general not invariant under projection (either parallel or central), but descriptive ones are. One general approach to using a camera model for image interpretation (and the one used in Section 3) is to first identify instances of descriptive attributes in the image, which, since they are invariant under projection, express strong, unambiguous, and often global information about the scene. These descriptive attributes can then be combined with geometric constraints (the camera model), heuristic rules (such as a preference for symmetrical figures), and specific knowledge of the scene to infer metric properties.

Under orthographic projection the parallelism of lines (a descriptive property) is invariant; under perspective projection, however, it is not — and must therefore be replaced with a more general property. Central projection maps parallel lines in space onto a pencil of lines intersecting at a common point on the image plane. This point of intersection, called a vanishing point, has important implications for image interpretation. In perspective, consequently, parallelism as a descriptive property is replaced by "coincidence." In Section 3 a computational method for finding vanishing points is described.

Under both orthographic and central projection the metric properties are transformed in highly ambiguous ways. For example, an angle on a plane in three-space (defined as the intersection of two lines in the plane) can project to any angle in the image, depending on the orientation of the plane with respect to the image. From a more optimistic standpoint, we can say that angles in the image constrain the orientation of planes in three-space.

In Section 4 two algorithms are described for finding the orientation of planes from metric properties (angles and curvatures). It will be necessary to invoke heuristic assumptions about the symmetry of figures in space to arrive at a meaningful answer. Computational methods for shape perception sometimes make use of known or suspected symmetry in the scene to choose among multiple interpretations. Kanade has used this approach for interpreting orthographic projections [10].

3 Vanishing Points

This section describes a method for finding vanishing points in a perspective image and discusses how to use them to interpret the scene.

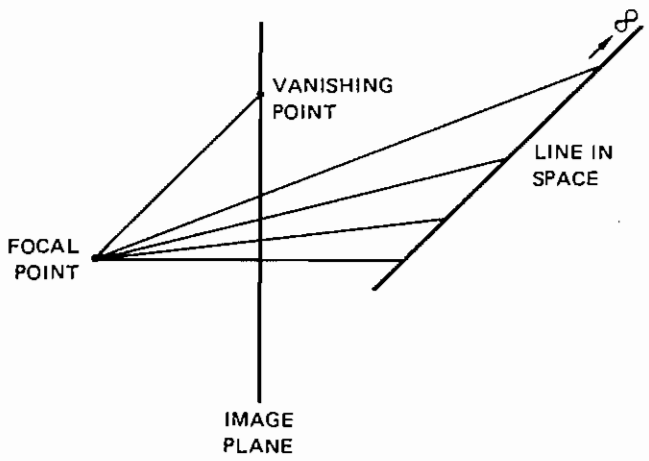


FIGURE 5 CONSTRUCTION OF A VANISHING POINT

The approach is based on the assumption that there exist groups of parallel straight structures in the scene, and that these structures produce line segments in the image. According to the laws of perspective, such a group of image line segments, when extended, will intersect at a common vanishing point. This point has the following interpretation: it is the projection of the intersection of the parallel lines "at infinity." If the focal length is known, the spatial orientation of the group is fully determined by the vanishing point of the lines (i.e., the line from the focal point to the vanishing point is a member of the group). This is illustrated in Figure 5.

The problem of finding vanishing points is divided into (1) finding line segments in the image and (2) finding intersections of the extended line segments that are likely to be vanishing points.

Problem (1) is solved with well-known, conventional techniques. First, "zero-crossing" contours in the image are found [11], then more-or-less straight segments of the zero-crossing contours are found with a recursive contour-splitting technique [12], and finally straight lines are fit to the segments using the method of least-squares.

Problem (2), that of finding intersections, is greatly simplified by using Gaussian mapping. The problem with trying to find intersections directly in the image is that the image plane is an open space, and the vanishing points may occur anywhere, even "at infinity." (The use of gradient space to represent surface orientation raises the same problem: as the surface normal approaches 90 degrees from the z-axis, the gradient space point approaches infinity.)

The **interpretation plane** associated with an image line is defined as follows (Figure 6). Let $\mathbf{p}_1 = (x_1, y_1, f)$ and $\mathbf{p}_2 = (x_2, y_2, f)$ be two distinct image points defining a line l. Then the interpretation plane ϕ associated with l is the plane containing l and the origin (i.e., the focal point), and can be represented by its unit normal row vector:

$$\begin{aligned} \phi &= \frac{\mathbf{p}_1 \times \mathbf{p}_2}{|\mathbf{p}_1||\mathbf{p}_2|} \\ &= (\phi_x, \phi_y, \phi_z) . \end{aligned} \tag{3.1}$$

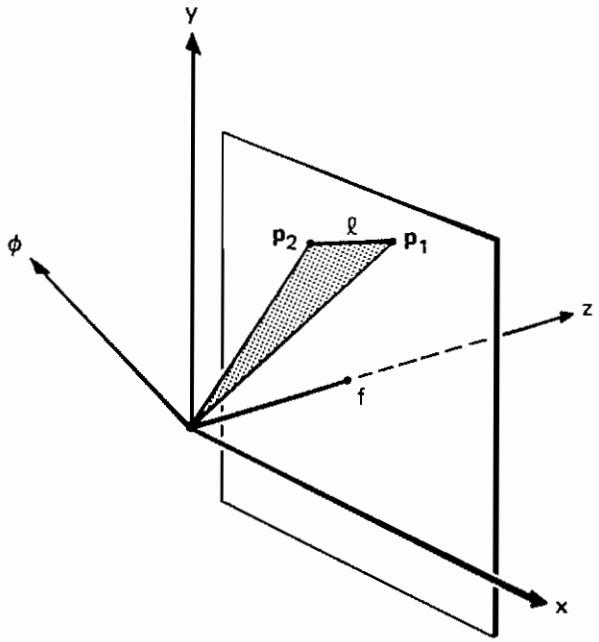


FIGURE 6 THE INTERPRETATION PLANE

The plane ϕ is called the interpretation plane of l because the line in space, the projection of which is l, must lie in ϕ.

The interpretation planes of image lines intersect the Gaussian sphere in great circles, as shown in Figure 7. The intersections of these great circles on the sphere correspond exactly to intersections of their associated lines in the image plane. The procedure for finding vanishing points is then as follows: (1) find lines in the image; (2) determine the interpretation plane of each line; (3) trace the great-circle intersections of the interpretation planes with the Gaussian sphere; (4) find the points on the sphere where several great circles intersect. The vanishing points can then be projected back onto the image plane, if desired.

After a vanishing point has been found, its dual interpretation can also be very useful for interpreting the image. The dual of a vanishing point is a **vanishing line** (a great circle on the sphere). For example, if the vertical vanishing point is found, its dual is the "horizon line" (i.e., the vanishing line of all horizontal planes). As another example, if two horizontal vanishing points are found, their duals intersect at the vertical vanishing point.

The Gaussian sphere can be represented digitally as a two-dimensional array of real numbers, with the row index indicating an azimuth and a column index representing an elevation. Each array element corresponds to a small surface area of the sphere. (In this representation the surface areas are not equal. Other forms, such as tessellated regular polyhedra [13], might be better suited to represent the sphere, but they are rather complicated to implement.)

The procedure for tracing a great circle in the sphere array is as follows. Let α and β be the spherical coordinates (azimuth and elevation) of a point on the sphere. In Cartesian coordinates, the point is

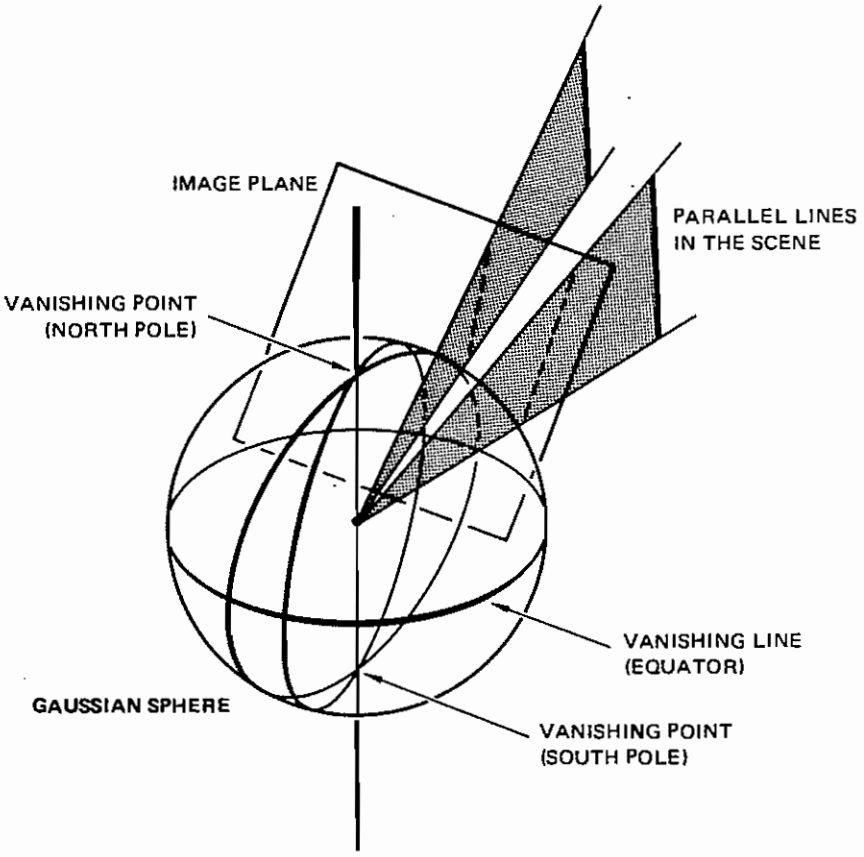


FIGURE 7 VANISHING POINTS ON THE GAUSSIAN SPHERE

$$\mathbf{g} = \langle \sin\alpha\cos\beta, \sin\beta, \cos\alpha\cos\beta \rangle \ . \tag{3.2}$$

The equation of the great circle associated with interpretation plane $\phi = \langle \phi_x, \phi_y, \phi_z \rangle$ is

$$\mathbf{g} \cdot \phi = 0 \ , \tag{3.3}$$

from which we can derive an expression for the elevation β in terms of the azimuth α and the interpretation plane ϕ:

$$\beta(\alpha, \phi) = \tan^{-1} \frac{-\phi_x \sin\alpha - \phi_z \cos\alpha}{\phi_y} \ . \tag{3.4}$$

If ϕ_y is small, this equation can be replaced by a slightly different form that gives azimuth as a function of elevation and has ϕ_x in the denominator.

The array is first initialized to zeros. Since we have the elevation β as a function of the azimuth α and an interpretation plane ϕ (Equation 3.4), we can generate the great circle of ϕ in the array. When a curve is traced into the array, a real value associated with that curve is added to all the array elements containing the curve. (This value is derived heuristically from the length and goodness-of-fit of the image line.) Points at which many curves intersect form clusters of high values; these indicate likely vanishing points.

Figures 8 and 9 show two examples of the method's application to real images. The images were recorded onto 50-mm x 50-mm film areas with a 50-mm-focal-length lens (considered "wide-angle" for this film format) and were digitized at a resolution of 100 microns/pixel.

In Figure 8c, the traces of the interpretation planes of each line segment is shown, and the original image is mapped onto the sphere for reference. This figure shows the "front" hemisphere spread flat. Only the vertical vanishing point is clearly indicated. The dual of the vertical vanishing point is shown to be the horizon line. In Figure 8d, those image line segments whose interpretation planes contain the vertical vanishing point are shown.

In Figure 9 two horizontal vanishing points are found, but the vertical vanishing point is not clearly indicated. The vertical vanishing point is located by intersecting the duals of the two horizontal vanishing points. Again, the horizon line is shown to be the dual of the vertical vanishing point.

Locating vanishing points may be extremely useful for wide-angle stereo image matching. A vanishing point measures the orientation of a group of parallel lines. Two vanishing points are sufficient to determine the rotational transform between the camera's coordinate system and a "natural" world coordinate system defined by the two groups of parallel lines. If two vanishing points in one stereo image are matched with two in the other image, the rotational part of the relative stereo camera model is determined. Matching vanishing points is more robust than matching image points for several reasons: there are fewer of them, they are truly "point-like" features which are not distorted by the projective transform, and they represent a relationship among many contour points covering a large part of the image, and are therefore less dependent of noise that can corrupt a single measurement in the image. Furthermore, matching actual line segments can be constrained to matching between sets of lines passing through corresponding vanishing points.

4 Orientation of Planes

4.1 General Approach

In this section, we shall consider a somewhat different class of algorithms for determining the orientation of planar figures; namely, backprojection of metric image properties. The algorithm for finding vanishing points described in the previous section exploits **descriptive** geometric properties (the intersection of lines in projective space) that are preserved under projection and can be used to infer rotational transforms. The two algorithms treated in this section, however, will exploit **metric** geometric properties (angles and curvatures), which, in general, are not preserved under projection. We shall have to impose heuristic, but very general, constraints (usually expressed as assumptions about the symmetry of the figure in space) to derive a meaningful answer.

Backprojection has been described most thoroughly by Kender [14]. Similar approaches have been used by Witkin [15] and Ikeuchi [16] for 3-D interpretation. The object is to find the most likely planar orientation of an area, given measurements of its geometric features in the image, such as size, slope, angles, and curvature. The essential idea is that

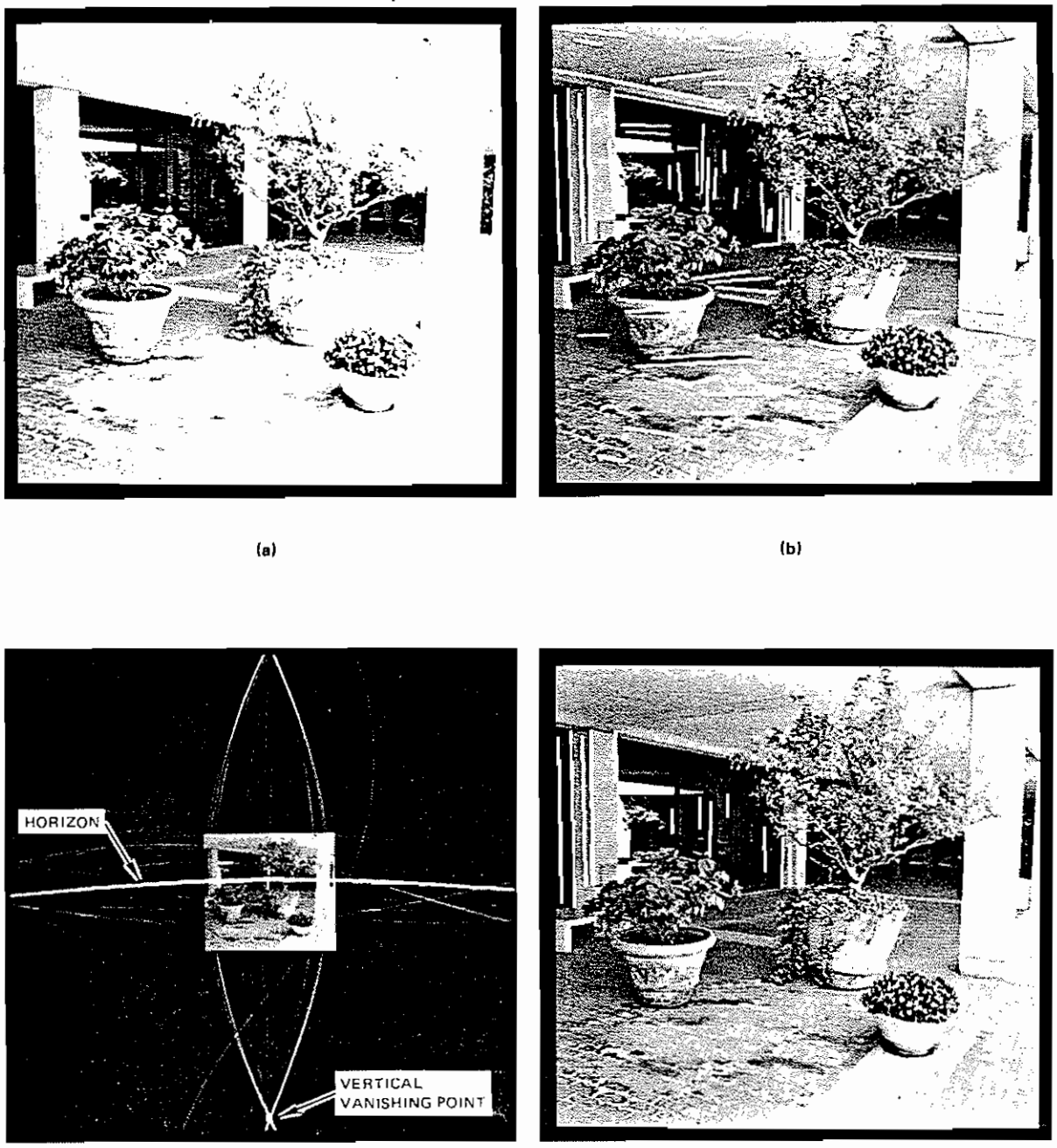


FIGURE 8 A VERTICAL VANISHING POINT: (a) original image; (b) line segments; (c) great circles on one half of the Gaussian sphere accumulator, showing the vertical vanishing point and the implied horizon; (d) vertical line segments.

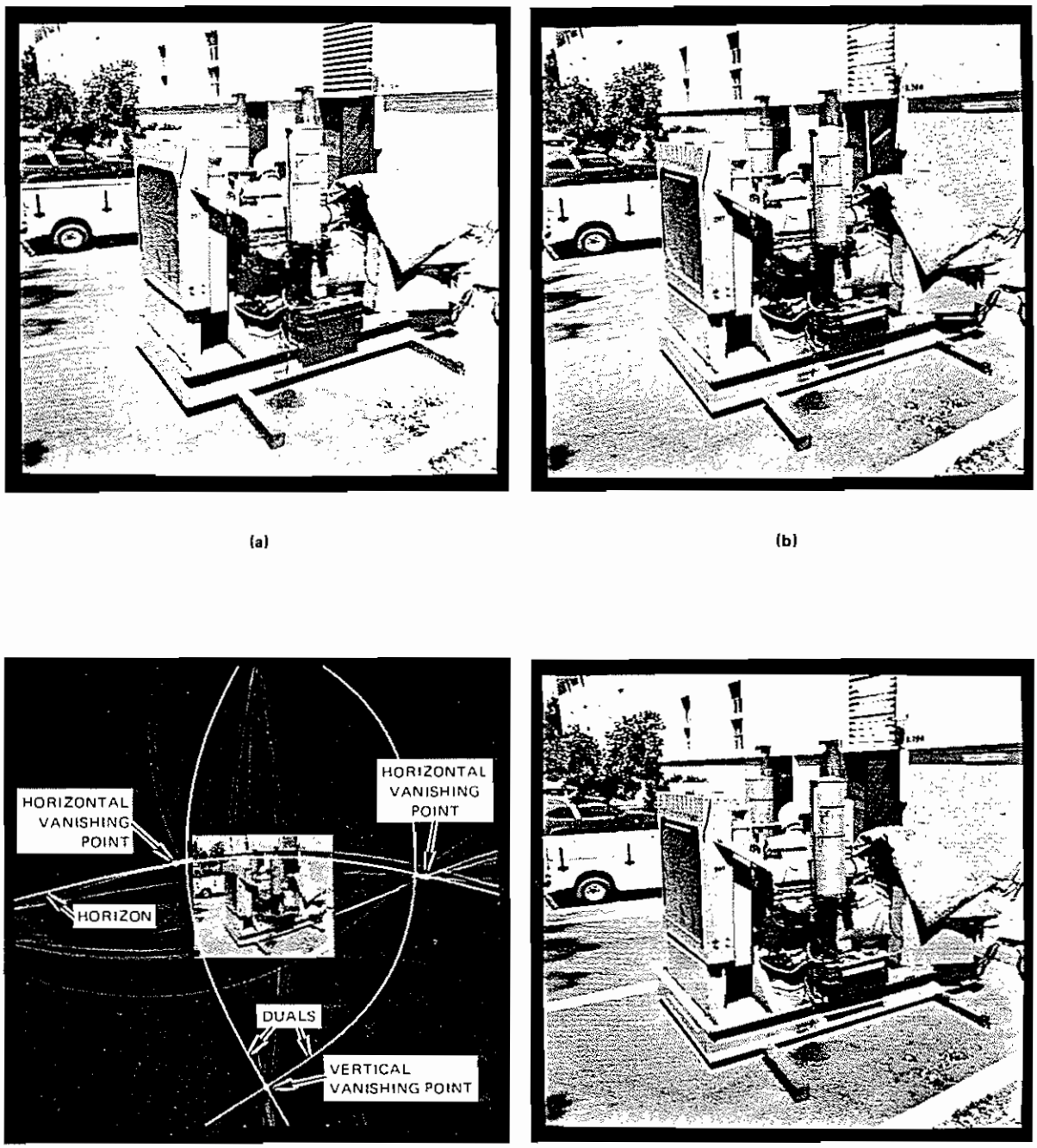


FIGURE 9 HORIZONTAL VANISHING POINTS: (a) original image; (b) line segments; (c) great circles on one half of the Gaussian sphere accumulator, showing two horizontal vanishing points, the implied vertical vanishing point where their duals intersect, and the implied horizon; (d) horizontal and vertical line segments.

the projective transform is reversed by computing, examining, and selecting the most appropriate of all feasible inverse projections. Consider a plane $\gamma(\alpha, \beta)$ tangent to the Gaussian sphere at the point with spherical coordinates (α, β). The plane is represented by its unit normal. In Cartesian coordinates the plane is

$$\gamma(\alpha, \beta) = (\sin\alpha\cos\beta, \sin\beta, \cos\alpha\cos\beta) \ . \tag{4.1.1}$$

The equation of the plane is

$$\gamma(\alpha, \beta) \cdot \langle x, y, z \rangle = 1 \ .$$

Image features can be projected onto $\gamma(\alpha, \beta)$ and measured. We can do this for all values of (α, β) (subject to some quantization of the sphere). The solution is that plane in which the "backprojected" properties satisfy various constraints. Witkin, for example, chooses the backprojection plane in which the distribution of tangents is most uniform [15].

The approach described here uses **intrinsic** properties of line figures and curves for backprojection. An intrinsic property is defined to be a property that is invariant under translation and rotation. Two such properties are the magnitudes of the angles of a planar polygon and the curvatures of a planar contour. Angle magnitude is also invariant under scaling, while curvature is not. As we shall see, this will affect the way we must interpret the curvature results.

4.2 Backprojection of Angle Magnitudes

Assume that the figure in space is a closed planar polygon, such as a triangle. Each pair of adjacent sides defines an angle, and each angle in the planar figure generally projects to a different angle in the image. (The metric property of angle magnitude is not preserved under projection.) Nevertheless, an angle measured in the image constrains the angle measured in the planar figure as a function of the orientation of the containing plane, but the constraint ranges over a family of possible planar orientations. In essence, the angle in the image can "backproject" onto any plane in space.

Consider two interpretation planes, $\phi_1 = (\phi_{1_x}, \phi_{1_y}, \phi_{1_z})$ and $\phi_2 = (\phi_{2_x}, \phi_{2_y}, \phi_{2_z})$, that form angle ω in the backprojection plane $\gamma(\alpha, \beta)$ (see Figure 10). The vectors $\gamma \times \phi_1$ and $\gamma \times \phi_2$ are the intersections of γ and the two interpretation planes. The dot product of these vectors is the product of their magnitudes and the cosine of the angle between them:

$$(\gamma \times \phi_1) \cdot (\gamma \times \phi_2) = |\gamma \times \phi_1||\gamma \times \phi_2| \cos\omega \tag{4.2.1}$$

or

$$\cos\omega = \frac{(\gamma \times \phi_1) \cdot (\gamma \times \phi_2)}{|\gamma \times \phi_1||\gamma \times \phi_2|} \ . \tag{4.2.2}$$

By using (4.2.2), angles measured in the image can be expressed as constraints on the orientation of the plane containing the angle in space.

One approach to using the backprojection constraint would be to solve a system of equations in the form of (4.2.2), but such an explicit solution may be very difficult. Instead, using a highly parallel algorithm, we backproject each image angle onto planes of

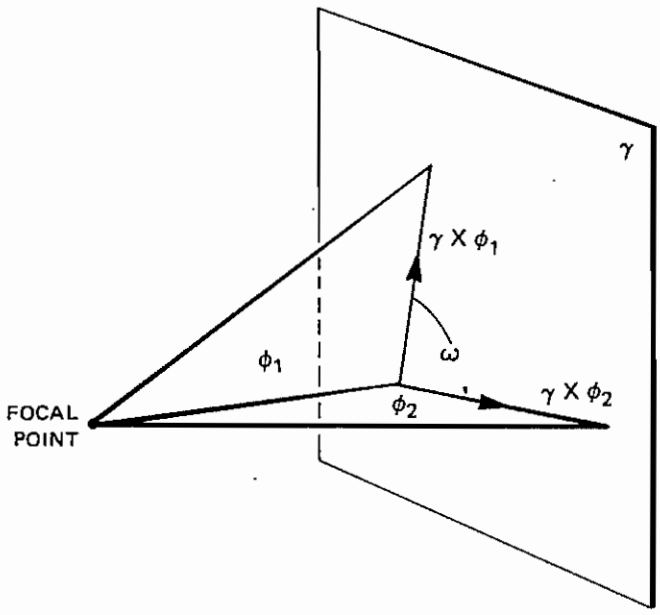


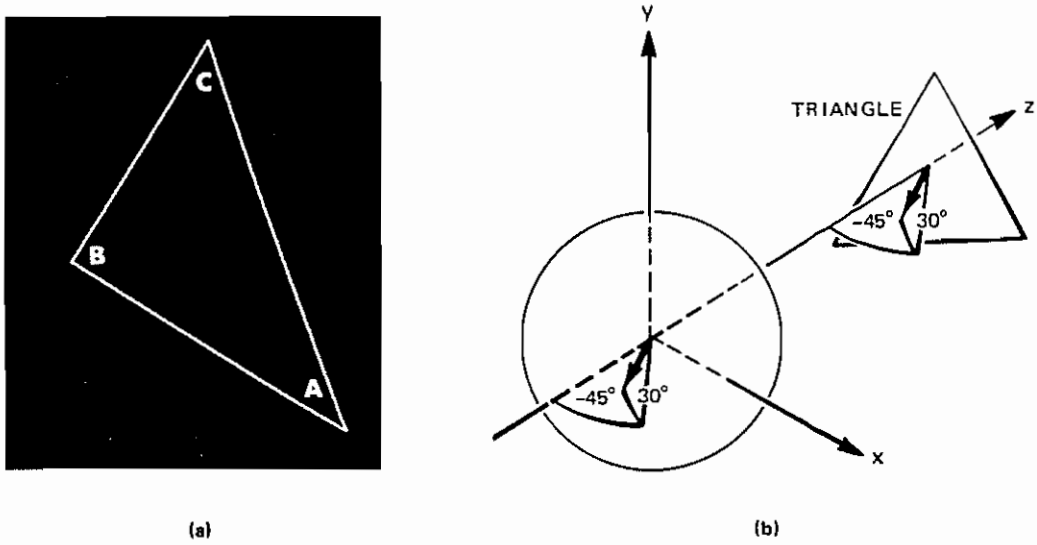
FIGURE 10 THE ANGLE TWO INTERPRETATION PLANES FORM ON ANOTHER PLANE IN SPACE

all possible orientations (subject to some quantization of the Gaussian sphere), obtaining at each orientation a value that expresses the angle the figure **must** make on such a plane.

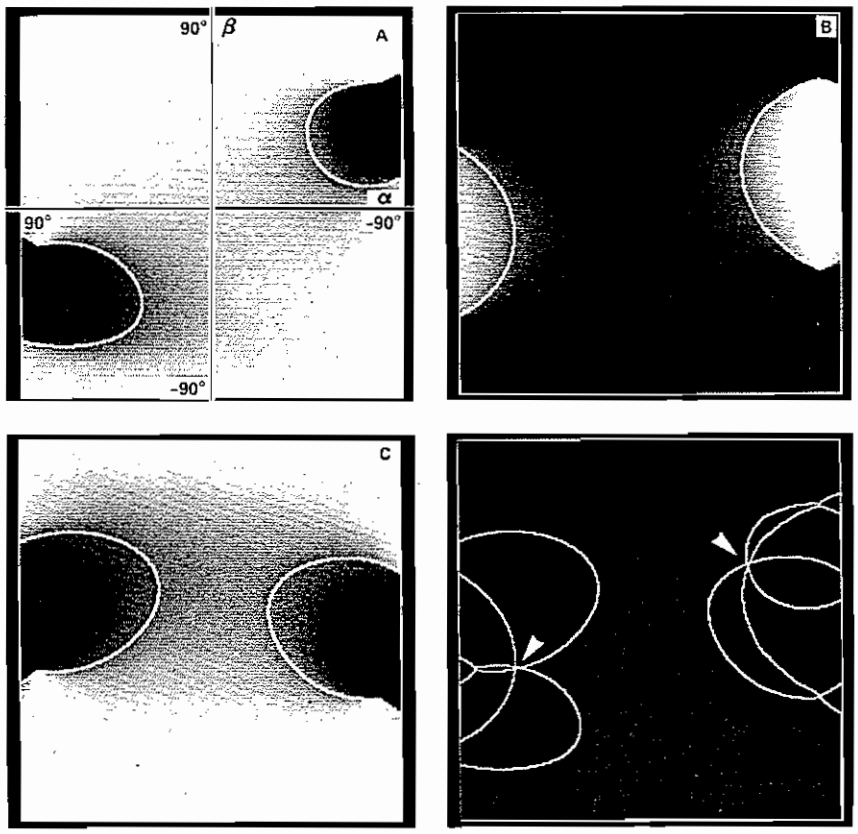
Two examples of this method are illustrated in Figures 11 and 12. The triangle in Figure 11a is interpreted as an image of some other triangle in space (Figure 11b). Each of the three angles is backprojected onto planes of all possible orientations (Figure 11c). Each image in Figure 11c represents a map of the back of the Gaussian sphere. Each point represents a possible planar orientation in terms of the Gaussian spherical coordinates of the plane's normal. The intensity value at each point is directly related to $\cos\omega$ for that orientation (e.g., black indicates $\cos\pi = -1$ and white indicates $\cos 0 = 1$).

Knowledge or heuristic assumptions about the values of angles in space can be used to choose particular interpretations of planar orientation. For example, suppose we interpret the triangle as being as symmetrical as possible — namely, an equilateral triangle ($\omega =$ 60 degrees). Contours for this value of ω are shown in the Figure 11c. Note that the triangle yields two solutions (i.e., the contours all intersect at two points). When asked to interpret Figure 11a as an equilateral triangle, most people have no difficulty in perceiving two possibilities: one in which the plane of the triangle faces "up" (this seems to be the preferred interpretation), and one in which it faces "down." This is actually a simplified version of the Necker cube illusion. The backprojection computation is therefore consistent with human perception.

Similarly, the quadrangle in Figure 12a is interpreted as an image of some other quadrangle in space. In this case four angles are backprojected (Figure 12c). If we assume the quadrangle in space is a rectangle ($\omega = 90$ degrees), and plot the contours for this value (shown in Figure 12c), we find that it yields a unique solution. Again, the result is consistent

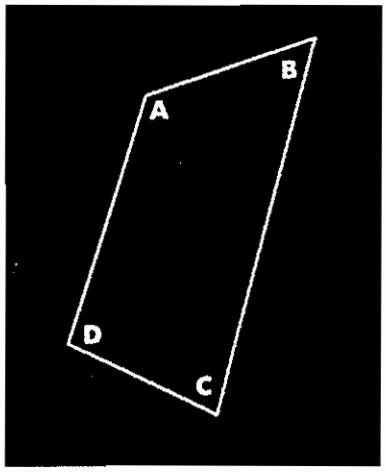


(a) (b)

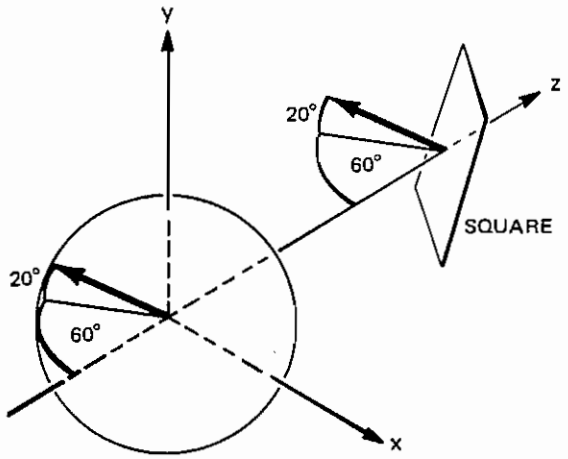


(c)

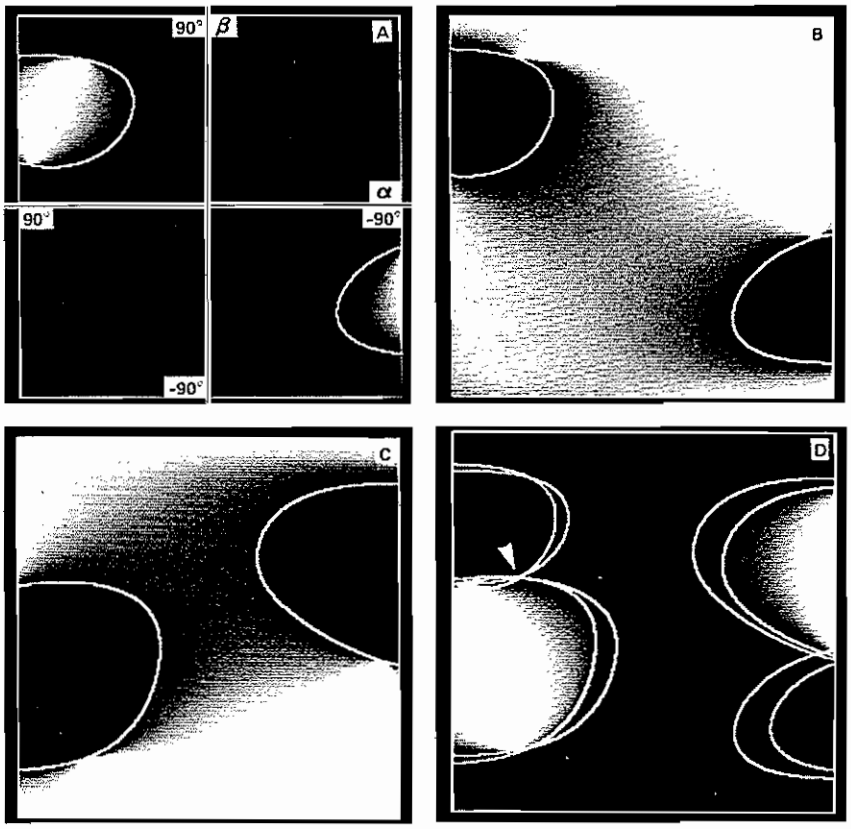
FIGURE 11 BACKPROJECTION OF ANGLES (AN EQUILATERAL TRIANGLE): (a) image of the triangle; (b) spatial configuration; (c) backprojection of the three angles, showing contours for 60° (superimposed contours at the lower right indicate two solutions).



(a)



(b)



(c)

FIGURE 12 BACKPROJECTION OF ANGLES (A RECTANGLE): (a) image of the rectangle; (b) spatial configuration; (c) backprojection of four angles, showing contours for 90° (one solution is indicated in the superimposed contours at the lower right).

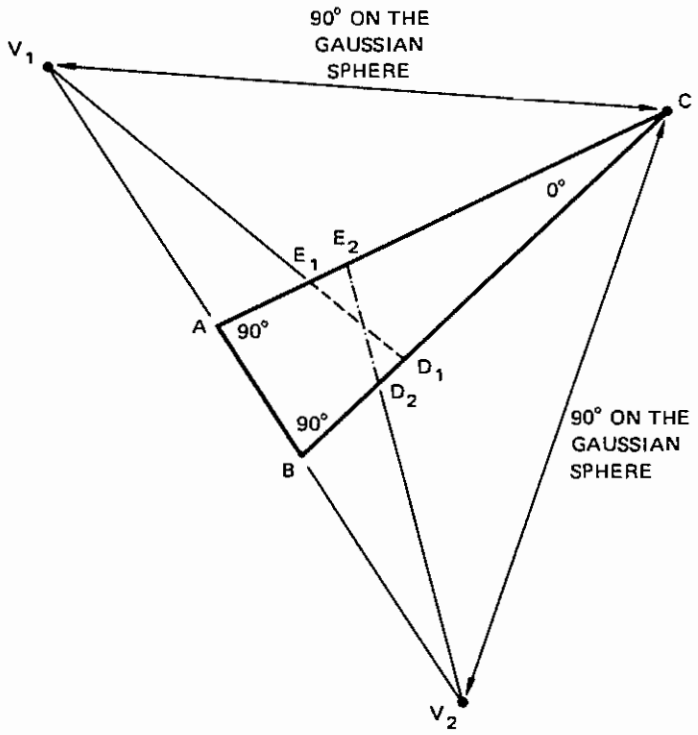


FIGURE 13 A TRIANGLE vs A QUADRANGLE

with human perception. This is not so surprising when consider that the backprojection computation does not merely *estimate* planar orientation, but actually *measures* it under some arbitrary constraint. Consistency of the backprojection computation with human perception in these examples implies that assumptions about or preferences for symmetrical figures is fundamental to shape perception in humans.

Why does a triangle lead to two solutions and a quadralateral to only one? Consider the situation illustrated in Figure 13. The triangle ABC is the perspective projection of an "ideal triangle" in projective space with parallel sides AC and BC. Point C is then a vanishing point. The vanishing point of side AB must be separated from C by 90 degrees on the Gaussian sphere, so it could be either point V_1 or point V_2. These two choices correspond to the two solutions for the orientation of the plane of ABC. (Two vanishing points determine the vanishing line, and hence the orientation, of a plane.) Now suppose we construct a rectangle in projective space from ABC by adding a line DE parallel to AB. This can be done in two ways, depending on which vanishing point is chosen for AB (lines D_1E_1 and D_2E_2, respectively). Therefore, if we interpret the image quadrangle ABD_1E_1 as a rectangle in space we get the solution corresponding to vanishing point V_1, but if we interpret the image quadrangle ABD_2E_2 as a rectangle in space we get the solution corresponding to vanishing point V_2.

4.3 Backprojection of Curvature

Consider the projection of a curved planar contour in space onto an image plane (as in

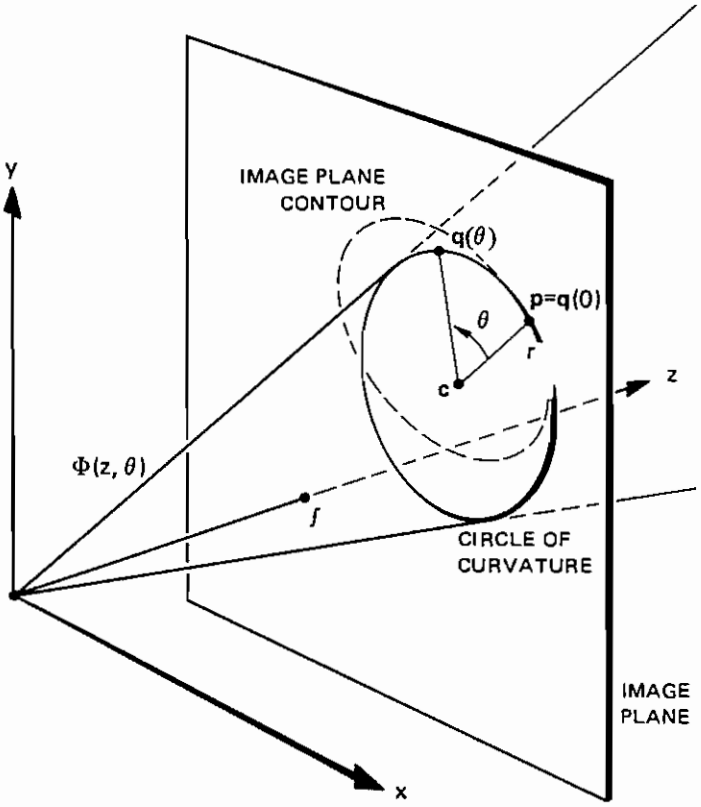


FIGURE 14 THE INTERPRETATION CONE $\Phi(z, \phi)$

Figure 4). The curvature measured at a contour point in space will be different, in general, from the curvature measured at the corresponding point in the image. The difference will depend both on the orientation of the plane in space and on the distance of the plane from the origin.

Suppose a curvature measurement is made in the image. It can be represented by two vectors: $\mathbf{p} = (p_x, p_y, f)$, a point on the image curve, and $\mathbf{c} = (c_x, c_y, f)$, the center of curvature for that point. That is, $\mathbf{p}$ and $\mathbf{c}$ determine a **circle of curvature** $\mathbf{q}(\theta)$ tangent to the image-plane contour (Figure 14). The **curvature** κ of the contour at the point of tangency is the inverse of the radius of the circle of curvature, or $\kappa = \frac{1}{|\mathbf{p}-\mathbf{c}|}$.

The equation of the circle of curvature in the image is

$$\mathbf{q}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} (\mathbf{p} - \mathbf{c})^T + \mathbf{c} \ . \tag{4.3.1}$$

Using the circle of curvature, we can define an **interpretation cone** that is a generalization of the concept of an **interpretation plane.** The interpretation cone Φ is the oblique circular cone with its apex at the origin (the focal point) and its base $\mathbf{q}(\theta)$ in the image plane (Figure 14). It can be represented mathematically as a ruled surface:

$$\Phi(z, \theta) = z\mathbf{q}(\theta) \ . \tag{4.3.2}$$

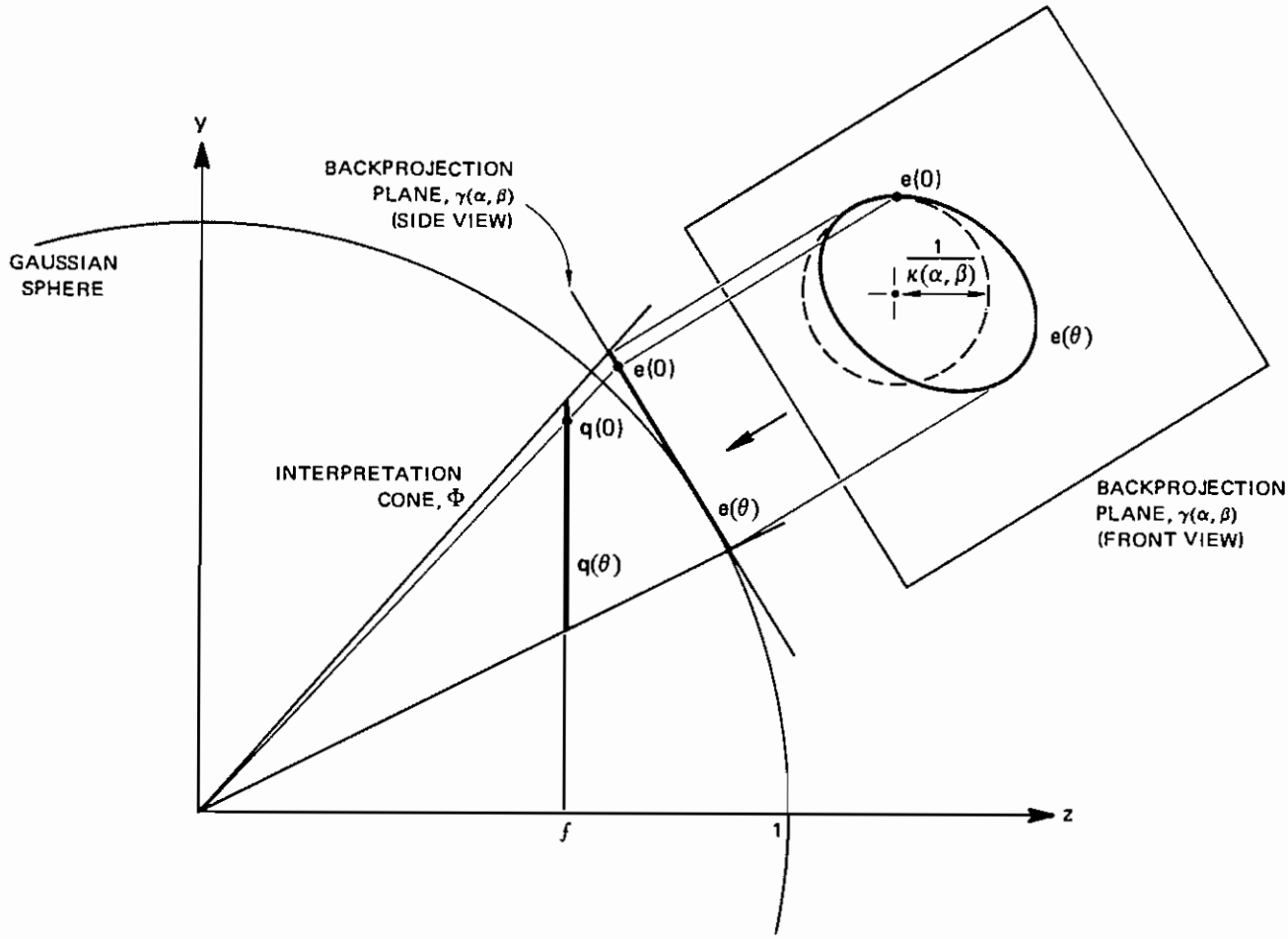


FIGURE 15 BACKPROJECTION OF CURVATURE

In fact, the entire image-plane contour describes a "complex cone." Each interpretation cone shares a line with this complex cone (the ray through the focal point and point **p**). Along this line, the interpretation cone also has the same curvature as the complex cone. (They are called "osculating surfaces.")

To find the backprojected curvature for a pair of image-plane measurements **p** and **c**, we intersect the interpretation cone of **p** and **c** with planes tangent to the Gaussian sphere, and then measure the curvature of the resulting contour. (Figure 15). One of these planes is the Gaussian mapping of the plane of the contour in space, and therefore specifies the orientation of the contour in space. Our object is to find it.

Because the correct backprojection plane is parallel to the plane of the contour in space, the intersection of the correct backprojection plane and the complex cone is a contour that is similar to the contour in space. The backprojected curvatures in the correct plane will differ from the true curvatures on the contour in space only by a common scale factor. This is not true, in general, for other planes. If, therefore, we know (or suspect) something

about the distribution of curvatures on the contour in space, we can use this information to select the correct plane.

The intersection of the backprojection plane $\gamma(\alpha, \beta)$ (Equation 4.1.1) with an interpretation cone Φ is given by

$$\gamma(\alpha, \beta) \cdot \Phi(z, \theta) = z(\gamma(\alpha, \beta) \cdot \mathbf{q}(\theta)) = 1 \ . \tag{4.3.3}$$

We can solve for z:

$$z = \frac{1}{\gamma(\alpha, \beta) \cdot \mathbf{q}(\theta)} \ . \tag{4.3.4}$$

Substituting z into (4.3.2), we get

$$\mathbf{e}(\theta) = \frac{\mathbf{q}(\theta)}{\gamma(\alpha, \beta) \cdot \mathbf{q}(\theta)} \ . \tag{4.3.5}$$

Equation (4.3.5) describes an ellipse that is the intersection of the $\gamma(\alpha, \beta)$ plane and the $\mathbf{q}(\theta)$ interpretation cone. The backprojected curvature $\kappa(\alpha, \beta)$ is the curvature of the ellipse at $\theta = 0$. It may be possible, using the calculus of variations and differential geometry, to solve (4.3.5) analytically. That is, for a collection of image-plane curvature measurements $\{\mathbf{p}_i, \mathbf{c}_i\}$, find the values for α and β that either minimize or maximize some measure over $\{\kappa_i(\alpha, \beta)\}$. In the meantime, we can solve the problem constructively, just as in the preceding section.

Let the first and second derivatives of an image-plane circle of curvature $\mathbf{q}(\theta)$ with respect to θ be denoted as as $\mathbf{q}' = d\mathbf{q}(\theta)/d\theta$ and $\mathbf{q}'' = d^2\mathbf{q}(\theta)/d\theta^2$. These can be found by differentiating (4.3.1).

Similarly, let the first and second derivatives with respect to θ of an ellipse $\mathbf{e}(\theta)$ in some backprojection plane $\gamma(\alpha, \beta)$ be denoted as $\mathbf{e}' = d\mathbf{e}(\theta)/d\theta$ and $\mathbf{e}'' = d^2\mathbf{e}(\theta)/d\theta^2$. These can be found by differentiating (4.3.5), yielding

$$\mathbf{e}' = \frac{\mathbf{q}'}{(\gamma \cdot \mathbf{q})} - \frac{\mathbf{q}(\gamma \cdot \mathbf{q}')}{(\gamma \cdot \mathbf{q})^2} \tag{4.3.6}$$

and

$$\mathbf{e}'' = \frac{\mathbf{q}''}{(\gamma \cdot \mathbf{q})} - \frac{2\mathbf{q}'(\gamma \cdot \mathbf{q}') + \mathbf{q}(\gamma \cdot \mathbf{q}'')}{(\gamma \cdot \mathbf{q})^2} + \frac{2\mathbf{q}(\gamma \cdot \mathbf{q}')^2}{(\gamma \cdot \mathbf{q})^3} \tag{4.3.7}$$

If $\mathbf{e}(\theta)$ were a "natural" representation of the ellipse, backprojected curvature could be computed directly from (4.3.7) as $|\mathbf{e}''(0)|$. (A natural representation of the ellipse $\mathbf{e}(s)$ is one in which $|d\mathbf{e}(s)/ds| = 1$; or, intuitively, it is one in which the natural parameter s is a measure of arc length.) But $\mathbf{e}(\theta)$ is not a natural representation, so the backprojected curvature must be be computed with

$$\kappa(\alpha, \beta) = \frac{|\mathbf{e}'(0) \times \mathbf{e}''(0)|}{|\mathbf{e}'(0)|^3} \ . \tag{4.3.8}$$

See Lipschutz [17] for a derivation of (4.3.8).

Just as Equation (4.2.2) was used to generate constraint surfaces for angle magnitudes, Equation (4.3.8) can be used to generate constraint surfaces for curvature. From a collection

of image-plane curvature measurements, $\{\mathbf{p}_i, \mathbf{c}_i\}$, we generate constraint surfaces $\{\kappa_i(\alpha, \beta)\}$. Examples are shown in Figure 16. The solid ellipse in Figure 16a is the projection of some circle in space, and the four broken circles represent curvature measurements along this image contour. (Lines from the points of each measurement to the centers of curvature are also shown.) Figure 16c shows the backprojected curvatures of each measurement.

The problem of interpreting the curvature results is somewhat more complex than for the case of angle magnitude. Angle magnitudes are invariant under scaling, but curvatures are not. For example, if the size of a rectangle increases the angles remain 90 degrees, but if the size of a circle increases the curvature decreases as the inverse of the radius. We can assign a definite physical meaning to a backprojected angle (e.g., it must be 90 degrees in the backprojection plane). A backprojected curvature, however, is meaningless by itself, and must be interpreted in relation to other backprojected curvatures. (Contours for the correct backprojected curvature are shown in Figure 16c, but this value was computed *a priori* from the known spatial configuration.)

A reasonable heuristic assumption about a contour in space, in the absence of any contradictory information, is that it is the most symmetric contour which is consistent with the projection of the contour in the image. In the following discussion the principle will be illustrated with circular contours. Since the curvature is constant on the circle in space, and the backprojected curvatures in the solution plane (the plane tangent to the Gaussian sphere and parallel to the circle in space) differ from the curvatures on the circle in space by only a common scale factor, we can infer that the backprojected curvatures in the solution plane must be equal. Fortunately, there is an easy way to compute this.

Suppose we have a set of scalar measurements $\{p_i \mid i = 1, n\}$ such that $0 \leq p_i$ and $\sum_{i=1}^{n} p_i = 1$. Then the **entropy** of the set $\{p_i\}$ is defined to be

$$H = -\sum_{i=1}^{n} p_i \log p_i \ . \tag{4.3.9}$$

It is well known that H is a maximum if and only if for all i, $p_i = \frac{1}{n}$.

The curvature backprojection constraint surfaces satisfy the first condition on $\{p_i\}$, but the sum is not in general unity. Let

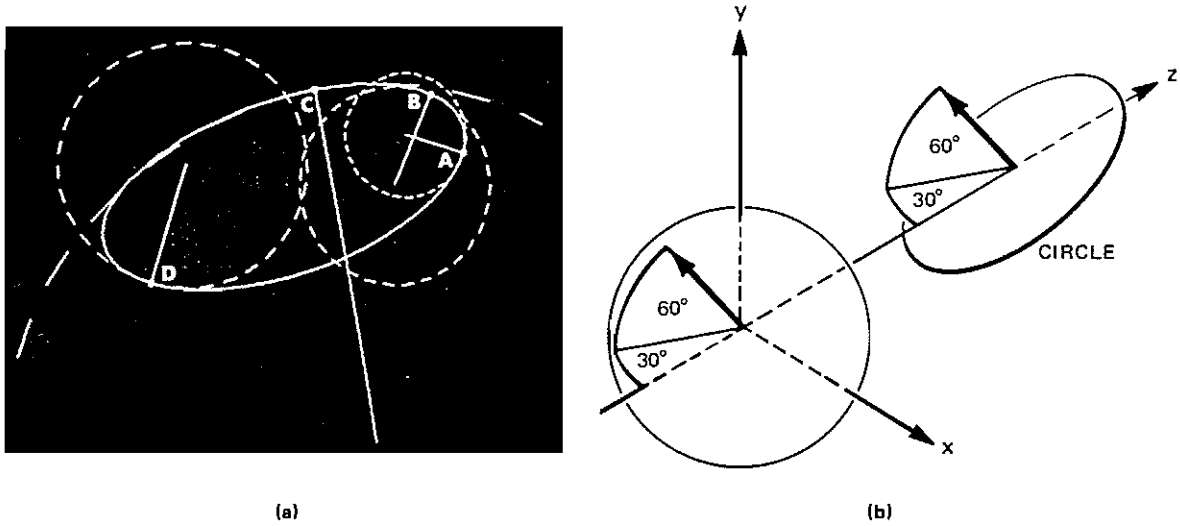
$$S(\alpha, \beta) = \sum_{i=1}^{n} \kappa_i(\alpha, \beta) \tag{4.3.10}$$

be the sum of backprojected curvatures at some α, β. Let

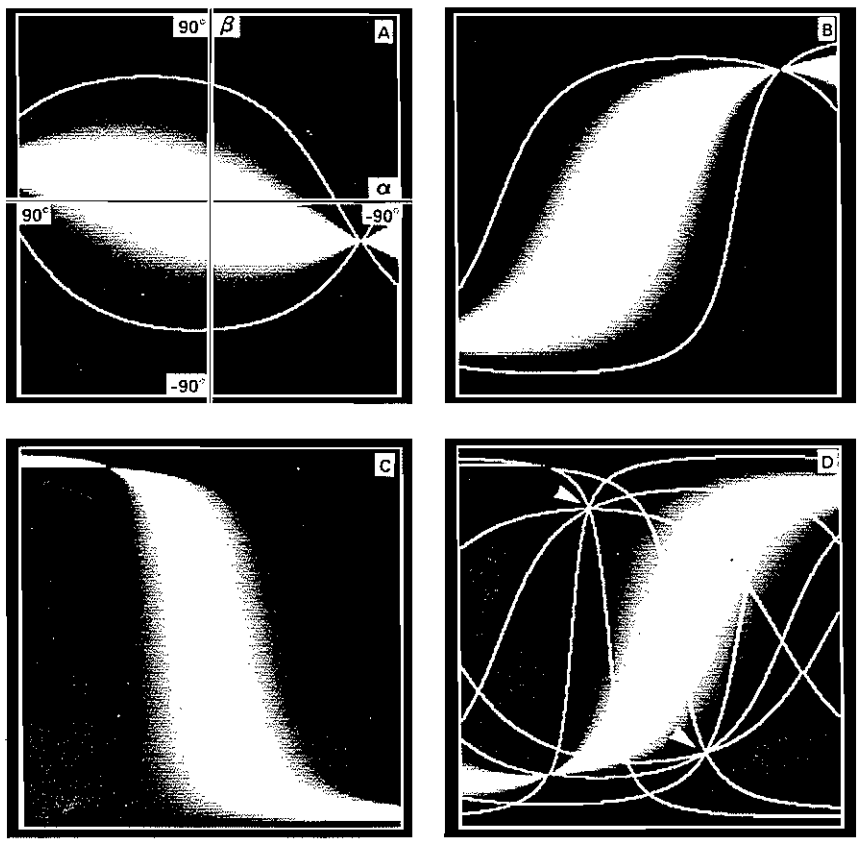
$$H'(\alpha, \beta) = -\sum_{i=1}^{n} \kappa_i(\alpha, \beta) \log \kappa_i(\alpha, \beta) \ . \tag{4.3.11}$$

The curvature measurements can be normalized by dividing each $\kappa_i(\alpha, \beta)$ by $S(\alpha, \beta)$, yielding

$$H(\alpha, \beta) = \frac{H'(\alpha, \beta)}{S(\alpha, \beta)} - \log S(\alpha, \beta) \ . \tag{4.3.12}$$

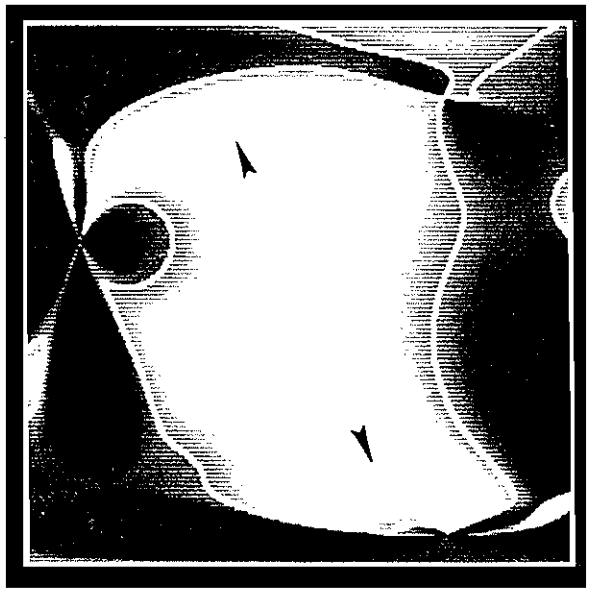


(a) (b)

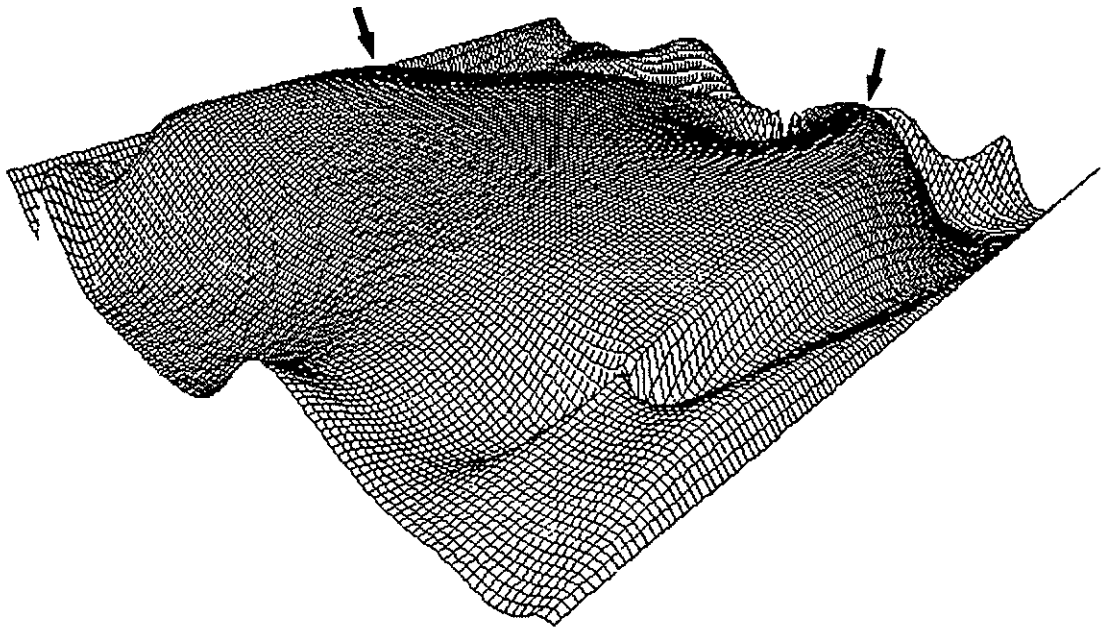


(c)

FIGURE 16 BACKPROJECTION OF CURVATURE (A CIRCLE): (a) image of the circle (the solid ellipse) and four curvature measurements (the broken circles); (b) spatial configuration; (c) backprojection of curvatures, showing contours for the solution value.



(d)



(e)

FIGURE 16 BACKPROJECTION OF CURVATURE (A CIRCLE) (Continued): (d) entropy surface for backprojected curvatures, showing contours for various entropy values and two solutions with maximum entropy; (e) perspective plot of entropy surface.

Equation (4.3.12) gives the entropy of a set of backprojected curvatures in the plane specified by α and β, and it can be used to convert a set of backprojection constraint surfaces into a single entropy constraint surface.

In information theory, entropy is interpreted as the average information per symbol [18]. Here we are interpreting it as a measure of symmetry. The relationship between the two concepts can be understood by thinking of the set of normalized scalar measurements $\{\kappa_i(\alpha,\beta)/S(\alpha,\beta) \mid i = 1,n\}$ as a *message*. The *language* this message belongs to is the set of all possible sets of measurements for a particular curve. If the curve is symmetric, the message will tend to have repeated *symbols*. The maximum-entropy heuristic is then simply an assumption that symmetric figures are more likely than asymmetric ones. It is a very general heuristic, and could be used to interpret the results for backprojected angles in the previous section. It could also be used to interpret backprojection of size, which, like curvature, is another intrinsic metric property invariant under rotation and translation, but not under projection or scaling.

Figure 16 illustrates the use of maximum entropy to select an interpretation for a set of backprojected curvatures. The entropy surface resulting from the backprojected curvature surfaces in Figure 16c is shown in Figure 16d, along with contours for several values, and a perspective plot of the surface is shown in Figure 16e. Two solutions are found, one of which corresponds to the known solution. Two solutions are found because an ellipse in the image plane could be the projection of circles in planes of two different orientations in space. One way to understand this result is to consider the oblique circular cone defined by the origin (the apex) and the circle in space (the base). There are two ways to cut an oblique circular cone into sets of parallel circular sections [19]. These correspond to the two solutions.

5 Conclusions

The perspective camera model is crucial for the interpretation of real images. Although parallel projection provides an adequate approximation when the included angle of view and the range of depth in the scene are small, these conditions are never completely satisfied. Perspective camera modeling entails more difficult mathematics than does orthography, but it also provides more powerful aids to perception (e.g., the Necker cube example in Section 1). In each of the problems we considered effective procedures that use constructive computational techniques were presented.

In Section 3 Gaussian mapping was used to identify descriptive geometric properties (the coincidence of parallel lines), and to infer metric properties (the orientation of groups of parallel lines). The dual interpretation of vanishing points on the Gaussian sphere was used to extend the analysis to finding vanishing lines.

Section 4 described the technique of backprojection of angles and curvatures. Once again the Gaussian sphere was used to represent the space of possible interpretations. Assumptions about the symmetry of figures in space, combined with the constraint surfaces obtained through backprojection, resulted in quantitative measurement of the orientation of the figures. A maximum-entropy heuristic was presented for interpreting the results of

backprojected intrinsic metric properties.

ACKNOWLEDGMENT

The research reported herein was supported by the Defense Advanced Research Projects Agency under Contract MDA903-79-C-0588. This contract is monitored by the U.S. Army Engineer Topographic Laboratory.

I would like to thank Martin Fischler, Graham Smith, Robert Bolles, and Alex Pentland for their invaluable help.

REFERENCES

1. Mackworth, A.K., Interpreting picture of polyhedral scenes, *Artificial Intelligence* **4** (1973) 121-137.

2. Huffman, D.A., Impossible objects as nonsense sentences, in: Meltzer, B. and Mitchie, D., (Eds.), *Machine Intelligence*, **6** (Edinburgh University Press, Edinburgh, 1971).

3. Haralick, R.M, Using perspective transformation in scene analysis, *Computer Graphics and Image Processing*, **13** (1980) 191-221.

4. Horn, B.K.P., Obtaining shape from shading information, in: Winston, P.H., (Ed.), *The Psychology of Computer Vision*, (McGraw-Hill, New York, 1975).

5. Shafer, S.A., Kanade, T., and Kender, J., Gradient space under orthography and perspective, Computer Science Department, Carnegie-Mellon University, CMU-CS-82-123, 1982.

6. Coxeter, H.S.M, *Introduction to Geometry*, (John Wiley & Sons, Inc., New York, 1961).

7. Newmann, W.M. and Sproull, R.F., *Principles of Interactive Computer Graphics*, 2nd. edition, (McGraw-Hill, New York 1979).

8. Paul, R.P., *Robot Manipulators*, (The MIT Press, Cambridge, Massachusetts, 1982).

9. Hilbert, D. and Cohn-Vossen, S., *Geometry and the Imagination* (Chelsea Publishing Co., New York, 1952).

10. Kanade, T., Recovery of the 3-D shape of an object from a single view, *Artificial Intelligence* **17** (1981) 409-460.

11. Marr, D. C. and Hildreth, E., A Theory of Edge Detection, AI Memo 518, Massachusetts Institute of Technology, Cambridge, Massachusetts (1979).

12. Duda, R. O. and Hart, P. E., *Pattern Recognition and Scene Analysis* (John Wiley & Sons, Inc., New York, 1973).

13. Ballard, D. H. and Brown, C. M., *Computer Vision* (Prentice-Hall, Inc., New Jersey, 1982).

14. Kender, J., *Shape From Texture*, Ph.D. Thesis, Carnegie-Mellon University, Computer Science Department, November, 1980.

15. Witkin, A. P., Recovering surface shape and orientation from texture, *Artificial Intelligence* **17** (1981) 17-45.

16. Ikeuchi, K., Shape from regular patterns (an example of constraint progogation in vision), A.I. Memo 567, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, March 1980.

17. Lipschutz, M. M., *Differential Geometry*, Schaum's Outline Series in Mathematics, (McGraw-Hill, New York 1969).

18. Goldman, Stanford, *Information Theory*, (Prentice-Hall, Inc., New York, 1953).

19. Rosenfeld, B. A. and Sergeeva, N. D., *Stereographic Projection*, (MIR Publishers, Moscow, 1977).