Shape Analysis[1]

by

R.J. WOODHAM[2]

Technical Report 86-12

May 1986

Laboratory for Computational Vision
Department of Computer Science
University of British Columbia
Vancouver, B.C., CANADA

1 chapter to appear in *Encyclopedia of Artificial Intelligence* S. Shapiro (ed.), John Wiley & Sons, New York, NY (1986).

2 Fellow of the Canadian Institute for Advanced Research.

INTRODUCTION

Shape analysis is a generic term used to identify methods that determine the form and the spatial arrangement of objects in the world. Vision and touch are the two sensory modalities used for shape analysis in human perception. Of the two, vision has received the most attention in artificial intelligence. Several articles survey work in the field (1,2,3). Textbooks are available for further study (4,5,6). Shape analysis based on touch and active range sensing is an emerging area of importance in robotics (7).

This article is about shape analysis in computational vision. Computational vision is the study of systems that produce descriptions of a world from images of that world. The purpose is to represent those aspects of the world that are required to carry out some task. For most tasks, shape is a necessary component of the description produced. Here, the term *representation* is used to identify a formalism, or language, for encoding a general class of shapes. The term *description* is restricted to mean a specific expression in the formalism that identifies an instance of a particular shape or class of shapes in the representation.

In a general purpose vision system, the mapping from signal input to final shape description is too complex to be treated as a function in a single representation. Shape analysis requires many levels of intermediate representation. Identifying those levels and establishing the constraints that operate both within and between levels is the fundamental challenge of computational vision research. Each level of representation must consider both the processes that derive the representation and the processes that compute with the representation. At the level of the signal, one deals with descriptions that can be derived directly from the image. This leads initially to representations for the two-dimensional shape of image patterns. Interpreting image properties as scene properties leads to representations for the visible surfaces in the scene. Finally, recognition of distinct objects and their spatial arrangement requires representations for three-dimensional shapes that are independent of viewpoint.

Over the past two decades, there has been considerable growth in the theoretical base for computational vision. One major trend has been towards a concentration on topics corresponding to identifiable modules in human perception (8). This has led to the development of three-dimensional vision systems including: shape from stereo (9,10), shape from contour (11,12,13,14,15), shape from motion (16) and optical flow(17) and shape from texture (11,18).

Computational vision distinguishs these three levels of representation: 2-D shape, visible surfaces, and 3-D shape. Unfortunately, there is no general agreement on what are the right representations to use at each of these levels. There is some agreement on necessary criteria these shape representations must satisfy.

CRITERIA FOR SHAPE REPRESENTATION

Several authors suggest necessary criteria that general-purpose shape representations must satisfy (3,19,20,21). No single representation proposed to date satisfies all of the criteria. Nevertheless, the criteria provide a useful framework to discuss representations that have been proposed. The criteria are as follows:

- *The representation of shape must be computable using only local support.* The ability to derive the representation from the input data is the minimal requirement. Local support further stipulates that the representation can be computed locally. This is required to deal with occlusion and to perform detailed inspection. It is also of practical importance since processes that derive the representation can then be implemented efficiently.

- *The representation of shape must be stable.* That is, small changes in the input should cause only small changes in the result. Images are subject to noise. Thus, stability is an important criterion for processes that derive initial descriptions from an image. Stability is also an important criterion for subsequent levels of representation because, without stability, it is difficult to define an effective measure of similarity to compare descriptions.

- *The representation of shape must be rich in the sense of information preserving.* Images are two-dimensional, while objects are three-dimensional. Image projection loses information. An image defines an equivalence class, usually infinite, of worlds that project to the identical image. A representation is rich if it does not arbitrarily restrict or extend this equivalence class. Rich representations are needed to describe a large class of objects, including objects that may never have been seen before.

- *The representation must describe shape at multiple scales.* Representations at multiple scales are useful for several reasons. First, representations at multiple scales suppress detail until it is required. Descriptions at a coarse scale relate to overall shape. Detail emerging at finer scales includes features that are more local. A pinhole in a metal casting is not significant when the task is to identify the part. But, it is critical when the task is to inspect the part for defects. Second, objects must be representable at different levels of detail. This can be accomplished using a hierarchical representation of shape that also takes into account the difference in object appearance owing to scale. For example, a forest is made up of individual trees. A forest can be represented hierarchically as a particular spatial arrangement and species composition of individual trees. At a coarser scale, the forest must still be represented as a forest, even when the individual trees are no longer discernible. Third, in the presence of noise, there is an inherent trade-off between the detectability of an image feature and its precise localization in space. By working at multiple scales, it is possible to optimize this trade-off dynamically, as required. Fourth, a coarse to fine analysis can introduce significant computational speed-up in methods for shape analysis requiring search or convergence. Fifth, to be useful, a representation should be storage efficient. Representations at multiple scales are needed to be both storage efficient and rich.

- *The representation must define an object-based semantics for shape description and segmentation.* In general, comparison of two-dimensional shape descriptions fails because there is no stable similarity measure to use. Large changes in shape description follow from minor changes in either the spatial configuration of the objects in the world, the viewpoint, or the illumination. Shape analysis requires representations in which three-dimensional shape is explicit so that spatial relationships between surfaces can be computed easily. This is necessary to segment complex shapes into simpler components, to predict how objects will appear, and to deal with occlusion.

- *The representation of shape must correspond to human performance on the task.*

Earlier, it was noted that an image defines an equivalence class of worlds that project to the identical image. Similarly, a representation defines equivalence classes of images that produce identical descriptions in the representation. Human perception also defines equivalence classes of worlds/images that produce identical perceptions. A representation of shape corresponds to human performance on some task if two conditions are satisfied. First, images that produce distinct descriptions in the representation are perceived as distinct in the task. Second, images that produce identical descriptions in the representation are perceived as identical in the task. A correspondence to human performance is difficult to achieve, in part because much remains to be understood about human perception. Nevertheless, developing this correspondence is a major motivating factor for current work in computational vision.

THE COMPUTATIONAL TASK

The computational task is to determine the three-dimensional shape of objects from their two-dimensional projection onto images. The principal shape analysis methods studied in computational vision are: shape from contour, shape from stereo, shape from shading, shape from texture, and shape from motion. Although each method differs considerably in precise detail, all share a common characterization as computational tasks. The steps embodied in a shape analysis method are:

- *Identify the visual task.* This involves picking a task domain and a class of locally computable image features for the domain that provide cues to three-dimensional shape.

- *Derive mathematical equations that describe how the world determines the image.* The equations are based on the laws of optics and, in general, consider both geometry and radiometry. The equations determine the mapping from scene to image. Shape analysis, however, requires a solution to the inverse problem. That is, it must determine the mapping from image to scene.

- *Demonstrate that the inverse problem is underconstrained.* It is usually straightforward to demonstrate that the problem is locally underconstrained. In general, the problem is also globally underconstrained although this can be more difficult to demonstrate.

- *Identify additional constraints that lead to a unique solution to the inverse problem.* Image features determine equivalence classes of possible scene features. Conceptually, a unique solution is obtained when a metric is applied to the equivalence classes to select a single preferred solution. Vision has been termed a conservative process (22). The metric is often expressed as a performance index designed to achieve smooth, regular, or minimal energy solutions. Identifying a suitable performance index is not a trivial matter. There are many possible measures to consider for a given visual task. Some degree of mathematical rigor is generally required to demonstrate that a particular choice does, in fact, lead to a unique solution. Finally, even when existence of a unique solution is established, it is still necessary to develop an algorithm to determine the solution.

- *Show that the solution thus obtained agrees with human perception.* Whatever the metric, the correct physical solution cannot be obtained in all cases. Human perception does not always correspond to the correct physical solution either. One level of agreement with human perception is to demonstrate that the computed solution agrees with human perception for the chosen visual task. At a second level, one also compares known algorithms for computing the solution to plausible mechanisms for biological implementation.

SHAPE FROM SHADING

A smooth opaque object produces an image with spatially varying brightness even if the object is illuminated evenly and is made of a material with uniform optical properties. Shading in an image provides essential information about object shape. Methods have been developed to determine shape from shading. These methods are based on an image irradiance equation formulated to determine image brightness as a function of surface orientation. The image irradiance equation cannot be directly inverted because surface orientation has two degrees of freedom and image brightness has only one. Additional information is required to reconstruct the visible surface. Many different constraints have been used in shape from shading methods. Here, two are considered. First, one can impose an overall smoothness metric on the desired solution, using locations in the image where surface orientation is determined locally as initial conditions. Second, one can use multiple images in a technique known as *photometric stereo*. The development given here originated with Horn (23) and includes extensions described in (24,25,26).

The Image Irradiance Equation

Image formation is modeled by an image irradiance equation. To standardize the geometry, consider objects to be defined in a left-handed Cartesian coordinate system with the viewing direction aligned with the negative Z-axis. The equation for a visible surface can be given explicitly as $z = f(x,y)$. In general, optical systems perform a perspective projection defined with respect to the focal point of the lens. If the size of the objects is small compared to the viewing distance, then the orthographic projection is obtained, as illustrated in Figure 1. In an orthographic projection, all rays from object to image are parallel because the focal point is at infinity. Thus, surface point (x,y,z) projects to image point (u,v). Without loss of generality, let $u=x$ and $v=y$. The orthographic projection, defined here, simply discards the z coordinate of each visible surface point (x,y,z).

There are several ways to specify direction in the coordinate system of Figure 1. Consider points on the unit sphere centered at the origin, called the *Gaussian Sphere* after Hilbert and Cohn-Vossen (27). Each point on the Gaussian sphere identifies the unit vector formed by joining the origin to that point. Thus, standard spherical coordinates can be used to specify the full range of directions. Now, the direction of a viewer facing surface normal at any point on the visible surface $z = f(x,y)$ can be found by taking the cross product of any two vectors lying in the tangent plane, provided they are not parallel to each other. Two such vectors are $[1,0,-p]$ and $[0,1,-q]$ where $p = \partial f(x,y)/\partial x$ is the slope in X direction and $q = \partial f(x,y)/\partial y$ is the slope in the Y direction. Their cross product is the vector $[p,q,-1]$. The quantity (p,q) is called the *gradient* of $z = f(x,y)$ and *gradient space* is the two-dimensional space of all such gradients (p,q).

Gradient space corresponds to the projection of points on the viewer facing hemisphere of the Gaussian sphere from the origin to the plane $z = -1$. Because of this, the gradient space has one drawback. Points where a surface smoothly disappears from view form what Marr (28) called an *occluding contour*. Points on an occluding contour have surface normals on the Gaussian sphere that intersect the plane $z = 0$. These points project to infinity in the gradient space. Another projection of the Gaussian sphere can be used instead. Points on the Gaussian sphere can be projected to the plane $z = -1$ from the point (0,0,1) rather than from the origin. This is a *stereographic projection*. Stereographic coordinates will be denoted as (f,g) to avoid confusion with the gradient (p,q). In stereographic coordinates, points on an occluding contour lie on the circle $f^2 + g^2 = 4$. Equations for transforming between spherical coordinates, the gradient, and stereographic coordinates can be found in (6,25).

Two directions are required to specify the local geometry of the incident and the reflected ray. A total of four parameters are required because each direction has two degrees of freedom. Often, however, one considers materials whose reflectance characteristics are invariant with respect to rotation about the surface normal. For surfaces that are isotropic in this way, only three parameters are required. Figure 1 illustrates one way to define the incident and the reflected ray in terms of three angles i, e, and g. This choice has one advantage over other possibilities. For a distant viewer and distant light source, the phase angle g is constant, independent of the surface normal.

The amount of light reflected by a surface element depends on its optical properties, on its microstructure, and on the spectral and spatial distribution of the illumination. The reflectance properties of a surface material are described by its *bidirectional* *reflectance distribution function* (BRDF). The BRDF was introduced as a unified notation for the specification of reflectance in terms of the incident and the reflected beam geometry (29). The BRDF identifies an intrinsic property of a surface material. It determines how bright the surface will appear when viewed from a given direction and illuminated from another.

The surface normal relates surface geometry to brightness because it determines the angles i, e, and g of Figure 1. For a constant scene irradiance and viewer geometry, the image irradiance measured from a given surface material varies only with the surface orientation. A *reflectance map* determines image irradiance as a function of surface orientation (23). When the gradient is used to represent surface orientation, the reflectance map is denoted by $R(p,q)$. When stereographic coordinates are used, it is denoted by $R(f,g)$. A reflectance map is a uniform representation for specifying the reflectance properties of a surface material for a particular light source distribution and viewer geometry. A reflectance map can be derived analytically for a given BRDF and light source distribution (30). More commonly, a reflectance map is measured empirically. A calibration object of known shape is used to determine image brightness as a function of surface orientation. The reflectance map obtained this way can be used to analyze other objects made of the same material and viewed under identical conditions of illumination. The empirical approach has the advantage of automatically correcting for the transfer characteristics of the imaging device.

Image formation can thus be described by a single equation called the *image irradiance equation.* If the gradient is used to represent surface orientation, the image irradiance equation becomes

$$E(x,y) = R(p,q)$$

where $E(x,y)$ is the image irradiance at image point (x,y) and $R(p,q)$ is the reflectance map value at the corresponding gradient (p,q). Shape from shading methods reconstruct a surface $z = f(x,y)$ to satisfy the image irradiance equation.

Shape from Shading and Occluding Contour

An object's boundary provides additional information that can be used to establish initial conditions. Parts of the boundary may correspond to sharp edges on the surface, as with polyhedra. Other parts correspond to places where the surface curves around smoothly. The latter defines an occluding contour, as we have seen. At an occluding contour, surface orientation is determined locally (12,25,28). In an orthographic projection, a normal to the silhouette in the image plane is also a normal to the surface at the corresponding point on the occluding contour.

Ikeuchi and Horn (25) developed an iterative algorithm to determine surface orientation using the image irradiance equation and a smoothness criterion as constraints. Surface orientation at occluding contours is the main source of initial conditions although information from singular points, specular points and self-shadow boundaries can also be included. The stereographic projection is used to represent surface orientation.

Consider the continuous case. The goal is to find functions $f(x,y)$ and $g(x,y)$ that make error in the image irradiance equation small, while keeping the solution surface as smooth as possible. Departure from smoothness is given by

$$\iint \left((f_x^2 + f_y^2) \; + \; (g_x^2 + g_y^2) \right) dx dy$$

where f_x, f_y, g_x and g_y are the first partial derivative of f and g with respect to x and y. Error in the image irradiance equation is given by

$$\int\int (E(x,y) - R(f,g))^2 \, dxdy$$

The problem is to minimize the combined error term e defined by

$$e = \int\int [(f_x^2 + f_y^2) + (g_x^2 + g_y^2) + \lambda(E(x,y) - R(f,g))^2] \, dxdy$$

Error in the image irradiance equation is weighted by λ compared to departure from smoothness. If the reflectance map is known accurately and if the brightness measurements are precise then λ can be made large. On the other hand, if λ is small then a smooth surface is determined despite noise and uncertainties about reflectance and illumination.

Minimization of an integral of the form

$$\int\int F(f,g,f_x,f_y,g_x,g_y) \, dxdy$$

subject to suitable boundary conditions is a problem in the *calculus of variations*. The function F is called the *performance index*. In general, the existence and uniqueness of a solution to problems in the calculus of variations cannot be taken for granted. Careful consideration must also be given to specification of boundary conditions. Correct formulation of a problem requires specifying a performance index that guarantees the existence of a solution and that provides the tightest set of natural boundary conditions that is consistent with the given data. Much of the early work in shape analysis omitted formal analysis of these problems. Recently, formulations of variational principles, called *regularization* in the Soviet literature (31), have been developed that guarantee the existence, uniqueness, and stability of the solution. Variational principles are increasingly being applied to problems in computational vision, both retroactively to analyze earlier work and as the basis for new shape analysis methods (22,32).

In a discrete formulation, the local departure from smoothness is given by

$$s_{i,j} = \frac{(f_{i+1,j} - f_{i,j})^2 + (f_{i,j+1} - f_{i,j})^2 + (g_{i+1,j} - g_{i,j})^2 + (g_{i,j+1} - g_{i,j})^2}{4}$$

The local error in the image irradiance is given by

$$r_{i,j} = (E_{i,j} - R(f_{i,j}, g_{i,j}))^2$$

where $E_{i,j}$ is the measured brightness at image point (i,j) and $(f_{i,j}, g_{i,j})$ is the corresponding surface orientation in stereographic coordinates. The problem is to minimize the error term e given by

$$e = \sum_i \sum_j (s_{i,j} + \lambda\, r_{i,j})$$

where λ is a free parameter as in the continuous case.

The solution method requires that this last equation be differentiated with respect to $f_{i,j}$ and $g_{i,j}$. For a minimum, the partial derivatives are all set to zero resulting in a large, sparse set of linear equations. These equations are solved using an iterative method and the set of values for $f_{i,j}$ and $g_{i,j}$ determine the solution surface.

Empirical results indicate that the method is both stable and robust. It works well when all information is precise. It continues to work reasonably well even when the reflectance map is only a crude approximation. One problem with this straightforward implementation is that convergence can be very slow. On a fine grid, the locations at which initial conditions are specified can be widely separated. The global minimum is achieved by iteratively propagating constraints across the network. A multiresolution algorithm for this formulation of shape from shading has been developed (26), based on *multigrid relaxation methods* of numerical analysis. Empirical results with the multiresolution implementation suggest that order-of-magnitude gains in efficiency are possible.

Photometric Stereo

Another approach to shape from shading uses multiple images to provide additional constraint. These images are taken from the same viewing direction, but under different conditions of illumination. This technique is called *photometric stereo* (33). It allows one to determine surface orientation locally without smoothness assumptions.

Suppose two images $E_a(x,y)$ and $E_b(x,y)$ are obtained by varying the direction of illumination. Since there is no change in the imaging geometry, each picture element (x,y) in the two images corresponds to the same object point and hence to the same gradient (p,q). This means that one does not have the problem of first identifying corresponding points in multiple views, as happens in binocular stereo. The effect of varying the direction of illumination is to change the reflectance map $R(p,q)$ that characterizes the imaging situation.

Let the reflectance maps for the two imaging situations be $R_a(p,q)$ and $R_b(p,q)$ respectively. Suppose a point (x_0,y_0) has measured intensities $E_a(x_0,y_0) = \alpha$ and $E_b(x_0,y_0) = \beta$. One obtains two equations in the two unknowns p and q $R_a(p,q) = \alpha$ and $R_b(p,q) = \beta$. There may be more than one solution because the equations are generally nonlinear. Additional information, such as a third image, can be used to determine the correct solution.

The multiple images required for photometric stereo can be obtained by moving a single light source, by using multiple sources individually calibrated or by rotating the object and imaging device together to simulate the movement of a single light source. An equivalent to photometric stereo can also be achieved in a single view by using multiple sources that can be separated by color.

Photometric stereo is easy to implement. The stereo computation, after an initial calibration step, is purely local and may be implemented by table lookup, allowing real time performance. Photometric stereo is a practical scheme for environments, such as industrial inspection, where the illumination can be controlled. It has been used as the basis for a prototype system to solve the industrial bin-of-parts problem (34).

The gradients computed at neighboring image points may vary considerably due to measurement errors. A smoothness constraint can be used to improve the results of photometric stereo (6). Using a performance index similar to the one discussed above, the problem is to minimize the error term e defined by

$$e = \int\int [(f_x^2 + f_y^2) + (g_x^2 + g_y^2) + \sum_i \lambda_i (E_i(x,y) - R_i(f,g))^2] \, dx dy$$

where $E_i(x,y)$ is the i^{th} image and $R_i(f,g)$ is the corresponding reflectance map. The λ_i weigh the errors in the image irradiance equations relative to the departure from smoothness. They may be different if the measurements from the images are not equally reliable.

Photometric stereo is complementary to binocular stereo. Binocular stereo allows the accurate determination of distance to the surface. Photometric stereo determines surface orientation. Binocular stereo works well on rough surfaces with discontinuities in surface orientation. Photometric stereo is best when surfaces are smooth with few discontinuities. Binocular stereo works well on textured surfaces with discontinuities in surface reflectance. Photometric stereo is best when surfaces have uniform optical properties. Photometric stereo has some distinct advantages. There is no difficulty identifying corresponding points in two images because the images are obtained from the same point of view. Determining correspondence is the major computational task of binocular stereo. In certain circumstances, surface reflectance can also be found because the effect of surface orientation on image brightness can be removed. Binocular stereo does not provide this capability. In some applications, a description of object shape based on surface orientation is preferable to a description based on range.

2-D SHAPE

In two-dimensional shape analysis, all descriptions are given in terms of image properties alone. It is assumed that there is a direct correspondence between image features and requirements of the task. For some tasks, the world is essentially two-dimensional and free of complications that arise from image projection. Examples include optical character recognition (OCR), inspection of printed circuit boards and VLSI layout, and microscopic blood cell image analysis. In some robotics tasks, the three-dimensional objects are confined to a small number of stable configurations that can be characterized by the object silhouette. Shape analysis for these tasks typically takes as input a binary image and is based either on object regions or on the object's bounding contour.

Global Shape Properties of Binary Images

Some global properties of regions in a binary image can be computed using only local support. Two examples are the metric properties perimeter (P) and area (A). The ratio $\frac{4\pi A}{P^2}$ is used as a measure of compactness. The factor of 4π in the numerator normalizes the ratio to one for a circle, the plane figure that encloses the most area for a given perimeter. The compactness ratio is less than one for all other regions. Perimeter, area and the compactness ratio are invariant under translation and rotation. The compactness ratio is invariant also to changes in scale, provided that the computation of perimeter and area is stable as image resolution changes.

In general, topological properties cannot be computed using only local support. The one exception is the topological property known as Euler number. The Euler number E of a binary image is the number of connected objects minus the number of holes. Sometimes it is known *a priori* that the image contains exactly one object. In this case, the Euler number E can be used to compute the number of holes.

The center of area and the direction of the principal axes are features derived when the position and orientation of an object are needed. A complete theory exists for binary images (6). Special purpose hardware is commercially available for binary image analysis. This hardware typically combines elementary binary image processing for smoothing, thinning, and noise suppression with the computation of the global features described above. Unfortunately, when objects overlap, the global features computed for the visible portions bear little relation to those that would be computed for the whole object. It is impossible to recognize occluded objects using only global features.

Other Representations of 2-D Shape

Other representations of two-dimensional shape are discussed in (4;chapter 8). Here, three are briefly described: the Hough transform, the symmetric axis transform, and smoothed local symmetries.

The *generalized Hough transform* (35) is one approach to finding instances of occluded shapes. Object shape is expressed parametrically. Local features add their vote to all locations in an accumulator array of parameter values consistent with the evidence provided by the feature. The final tally of votes determines the values of the parameters and hence the shape. Sufficient evidence can be accumulated even if part of the shape is occluded.

The *symmetric axis transform* (SAT), also called the *medial axis transform*, can be defined as the locus of centers of maximal disks that touch at least two points on the bounding contour of an object (36). The SAT is information preserving but it is not stable because it is very sensitive to small perturbations in the bounding contour. In many cases, the SAT produces unintuitive descriptions.

Recently, the *smoothed local symmetries* (SLS) representation has been developed to overcome deficiencies in the SAT in two ways. First, the representation redefines symmetry to be a local property. The precise details are not described here but can be found in (19,37). Potential axes of the shape are the maximal smooth loci of the local symmetries. Second, any axis whose support region is wholly contained in the support region of another axis is deleted. The surviving axes are the smoothed local symmetries and arguably produce a more intuitive description than the SAT. Smoothed local symmetries are an attempt to produce descriptions that can deal with subobjects at a variety of scales, rather than at a single level.

VISIBLE SURFACES

Descriptions at the level of visible surfaces make object properties explicit in the retinocentric coordinate system defined by the viewing direction. Shape properties made explicit include range and surface orientation. Shape from stereo and shape from motion derive range. Shape from shading, shape from contour, and shape from texture derive surface orientation. Discontinuities in range and surface orientation are included if they are available from earlier segmentation processes. Most often, they are derived subsequently. The representation of visible surfaces is included in the *intrinsic image* idea of Barrow and Tenenbaum (38) and in the *2 1/2-D sketch* of Marr (8). Horn refers to a map of surface orientations as a *needle diagram* (39).

Mackworth (40) popularized the use of gradient space to reason about visible surfaces when interpreting linedrawings of polyhedra. More recently, Draper (41) analyzed the gradient space representation for polyhedral scenes. Geometric properties of the gradient space under orthographic and perspective projection are summarized in (42). Studies of human perception demonstrate that image contours provide cues to three-dimensional shape that apply in a more general setting (43).

Skewed Symmetry

Symmetry in a two-dimensional shape requires a straight line axis for which opposite sides are reflective. That is, symmetrical properties occur along lines perpendicular to the axis of symmetry. Kanade (13) defined a generalization called *skewed symmetry.* Skewed symmetry in a two-dimensional shape requires symmetrical properties along lines not necessarily perpendicular to the axis, but at a fixed angle to it. Figure 2 illustrates.

Under orthographic projection, a three-dimensional planar surface with real symmetry appears as a two-dimensional shape with skewed symmetry. The converse, however, is not always true. Nevertheless, skewed symmetry in an image is often due to real symmetry in the scene. Kanade proposed the following assumption:

> A skewed symmetry depicts a real symmetry viewed from
> some unknown view angle.

This assumption is transformed into constraints in the gradient space. It restricts the gradient (p,q) to lie on a hyperbola in gradient space determined by α and β. In the absence of global information, Kanade suggested that the minimum slant interpretation for the gradient is the surface orientation that is perceived. See (13) for details.

The consequences of Kanade's skew symmetry assumption can be examined. First, when combined with other constraints, it can lead to a unique global solution for the orientation of each visible surface. The assumption provides a performance index to select a preferred solution from an equivalence class of solutions. The solution determined necessarily describes all skew symmetries as oriented real symmetries. Second, no global solution may be obtained because the equivalence class of solutions may not include one in which all skew symmetries are oriented real symmetries. Third, the solution obtained may not be the physically correct one. Consider, for example, a scene containing a circular clock mounted on an otherwise blank wall. The projected shape of the clock will have skewed symmetry, depending on the surface orientation (p,q) of the wall. If one assumes that the clock is circular, then local measurements are sufficient to determine the surface orientation of the wall. (There are two solutions of the form (p_0,q_0) and $(-p_0,-q_0)$.) On the other hand, if the clock was elliptical, its projected shape would still have skewed symmetry. This time, the assumption that the clock was circular would lead to an incorrect determination of the surface orientation of the wall. Fourth, the solution may not agree with human perception. As pointed out in (14), an ellipse is usually perceived as a tilted circle even though an ellipse has real symmetry.

Brady and Yuille (14) have recently proposed an extremum principle to estimate surface orientation from two-dimensional contour. The principle selects the plane orientation that maximizes the compactness ratio $\frac{4\pi A}{P^2}$ introduced earlier. That is, of all true shapes that project to the observed shape, select the one that is the most compact. This is sufficient to determine the gradient (p,q) of the plane containing the true shape. Kanade's skew symmetry assumption is implied by this extremum principle. Skew symmetries are interpreted as oriented real symmetries. Only in special cases, however, does the solution correspond to the minimum slant interpretation.

The Normalized Texture Property Map

Texture is an important visual cue to the properties of a surface material. At the same time, texture gradient is a cue to surface orientation. Kender (18) developed the *normalized texture property map* (NTPM) to represent the local properties of texture as a function of surface orientation. For convenience, suppose the gradient (p,q) is used to represent surface orientation, even though, as above, other choices are possible. The result is analogous to a reflectance map $R(p,q)$. Local geometric features of the primitive texture elements, called *texels*, determine equations that the gradient must satisfy. For some textures, the problem is locally underconstrained and methods analogous to those for shape from shading would be required to determine a solution. Sometimes two or more independent local geometric features can be determined from each texel, each characterized by its own NTPM. Then, the situation is analogous to photometric stereo and surface orientation can be determined locally. There is a strong connection between the local geometric properties of texels and skew symmetry. This connection is explored in (44).

3-D SHAPE

Surface-based descriptions can be derived from an image without specific knowledge of the objects in view. But, surface-based descriptions are different from different viewpoints. It is impossible to derive three-dimensional shape descriptions from an image simply because much of the object is obscured from view. The task in three-dimensional shape description is to interpret descriptions derived from the image as instances of existing knowledge structures. This is termed *model-based vision* (3). Model-based vision systems use models to predict what can be seen. This constrains the equivalence class of solutions to include only those solutions expressible within the knowledge base of the system. Some model-based systems, such as ACRONYM and the EGI representation discussed below, model three-dimensional shapes in a viewpoint independent way. Many others include specific constraints that are viewpoint dependent.

There are many representations for solid objects used, for example, in computer graphics and in computer-aided design and manufacture (CAD/CAM) (45). In shape analysis, the representation is used to predict how the object will appear. This is necessary to generate an inverted index of possible object features corresponding to a given local image feature (46). With this additional stipulation, the number of three-dimensional shape representations used in computational vision is much more restricted.

Generalized Cones

Generalized cylinders were developed by Agin and Binford (47) as a representation of three-dimensional volumes that emphasizes elongation. A shape is described by sweeping a cross-section along an arbitrary three-dimensional curve, called the spine, expanding or contracting the cross-section according to a scaling function, called the sweeping rule. The cross-section need not be circular or at right angles to the spine, although these are simplifications often used in practice. This formulation predates but is similar to the *generalized cones* of Marr (28). Generalized cones are generalized cylinders whose spines are straight lines.

Generalized cones satisfy many of the criteria for shape representation. There is a natural decomposition into a cross-section function and a spine function giving the representation local support. Minor perturbations do make the representation unstable. But, in practice, description is preceded by some degree of smoothing to assure stability. The representation is rich and can be applied at multiple scales (20). The representation is object-based and does support a natural segmentation of complex objects into simpler components. It has also been suggested that this representation corresponds to human perceptions for certain natural structures such as animals (20,46). Unfortunately, not all shapes are represented naturally using generalized cylinders.

Brooks (48) used generalized cones to represent airplane shapes independent of viewpoint in a system called ACRONYM. ACRONYM uses part/whole graphs of generalized cones to predict the appearance of wide-bodied aircraft in aerial photographs. Projections of generalized cones are called *ribbons*. ACRONYM determines the viewpoint simultaneously with the interpretation of ribbons as projected generalized cones.

The Extended Gaussian Image

The *extended Gaussian image* (EGI) of an object records the variation of surface area with surface orientation (49). The EGI is invariant to translation and can be normalized to be invariant also to scale. The EGI is typically represented as a function defined on the Gaussian sphere. Rotations are easy to deal with since an object and its EGI rotate together. The EGI is easy to derive from other representations of three-dimensional objects.

The EGI uniquely represents convex objects and is thus information preserving for this class of objects. An iterative algorithm has been developed to reconstruct a convex polyhedron from its EGI (50). The EGI has also been applied to determine object attitude, that is the three-dimensional rotation required to bring a sample object into correspondence with a prototype. The visible hemisphere of a convex object's EGI can be computed from the visible surface description of surface orientation. Conceptually, determining object identity corresponds to finding a prototype EGI with an identical hemisphere. Determining object attitude corresponds to matching the position and orientation of the visible hemisphere to the prototype EGI. This approach has been applied to the industrial problem of picking parts out of a bin (34). The matching computation can be ill-conditioned. Recently, the mixed-volume function, introduced in (50), has been used as a similarity measure for attitude determination. The result is more stable than direct EGI matching and can support efficient multiresolution attitude determination (51).

CONCLUSIONS

At first glance, shape analysis in computational vision can appear to be a collection of *ad hoc* techniques. In many applications, the environment can be controlled and the computational task can be structured so that a special-purpose vision system will succeed. Special-purpose systems represent an extremely limited repertoire of objects so that simple descriptions often are sufficient to distinguish between them. In contrast, as Binford (3;page 60) says, 0general vision requires strong description and weak classification.0 This article presents a coherent framework for shape analysis in a general-purpose vision system based on: a thorough understanding of image formation, a characterization of the computational task, criteria for representations of shape, and correspondence with human perception. These are the themes. But, much remains before they will be achieved in practice.

Shape from shading was used as the major illustrative example. The methods described satisfy many of the criteria for representation of shape. Surface orientation is computable using local support. The methods based on variational formulations are stable. The representation is rich because the orientation map produced is dense. Description at multiple scales has not explicitly been addressed. But, the choice of λ indirectly influences the level of detail considered. The generation of representations at multiple scales is an obvious application of the multigrid implementation approach. The final description is in terms of an intrinsic object property. The visible surface orientation map can be extended to a full three-dimensional representation through the EGI, although this is currently limited to simple objects. There is no claim that the shape from shading methods correspond to human performance.

The prerequisite problem of how to segment an image to obtain a contour description has not be considered. This problem is generally considered separately as *edge detection* or *region growing*. Some argue that segmentation and shape analysis cannot be separated but are inherently part of the same problem. See, for example, (3,52). Segmentation issues arise again at a higher level. Initial representations for shape are typically dense. It is desirable to segment shape descriptions both to be more storage efficient and to provide symbolic descriptions to facilitate subsequent analysis. There are a myriad of segmentation schemes proposed at each level of shape representation: 2-D shape, visible surfaces, and 3-D shape. Most address storage efficiency foremost but others deal with aspects related to symbolic description. This is an active area of ongoing research.

REFERENCES

1. H.G. Barrow and J.M. Tenenbaum, "Computational vision", *Proc. IEEE* **69**, 572-595 (1981).

2. M. Brady "Computational approaches to image understanding", *ACM Computing Surveys* **14**, 3-72 (1982).

3. T.O. Binford "Survey of model-based image analysis systems", *International Journal of Robotics Research* **1**, (1), 18-64 (1982).

4. D.H. Ballard and C.M. Brown, *Computer Vision*, Prentice-Hall, Englewood Cliffs, NJ, 1982.

5. R. Nevatia *Machine Perception*, Prentice-Hall, Englewood Cliffs, NJ, 1982.

6. B.K.P. Horn *Robot Vision*, MIT Press/McGraw-Hill, Cambridge, MA, 1986.

7. M. Brady "Artificial intelligence and robotics", *Artificial Intelligence* **26**, 79-121 (1985).

8. D. Marr *Vision*, W.H. Freeman, San Francisco, CA, 1982.

9. H.H. Baker and T.O. Binford, "Depth from edge and intensity based stereo", *Proc. of the Seventh International Joint Conference on Artificial Intelligence*, Vancouver, BC, 631-636 (1981).

10. W.E.L. Grimson *From Images to Surfaces*, MIT Press, Cambridge, MA, 1981.

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

12. H.G. Barrow and J.M. Tenenbaum, "Interpreting line drawings as three-dimensional surfaces", *Artificial Intelligence* **17**, 75-116 (1981).

13. T. Kanade "Recovery of the three-dimensional shape of an object from a single view", *Artificial Intelligence* **17**, 409-460 (1981).

14. M. Brady and A. Yuille, "An extremum principle for shape from contour", *IEEE Transactions on Pattern Analysis and Machine intelligence* **6**, 288-301 (1984).

15. T.O. Binford "Inferring surfaces from images", *Artificial Intelligence* **17**, 205-244 (1981).

16. E.C. Hildreth "Computations underlying the measurement of visual motion", *Artificial Intelligence* **23**, 309-354 (1984).

17. B.K.P. Horn and B.G. Schunck, "Determining Optical Flow", *Artificial Intelligence* **17**, 185-203 (1981).

18. J.R. Kender "A computational paradigm for deriving local surface orientation from local texture properties", *Proc. IEEE Workshop on Computer Vision: Representation and Control*, Rindge, NH, 143-152 (1982).

19. M. Brady "Criteria for representations of shape", in J. Beck, B. Hope, and A. Rosenfeld eds, *Human and Machine Vision*, Academic Press, New York, NY, 1983.

20. D. Marr and H.K. Nishihara, "Representation and recognition of the spatial organization of three dimensional structure", *Proc. R. Soc. Lond. B* **200**, 269-294 (1978).

21. F. Mokhtarian and A.K. Mackworth, "Scale-based description and recognition of planar curves and two-dimensional shapes", *IEEE Transactions on Pattern Analysis and Machine intelligence* **8**, 34-43 (1986).

22. M. Brady and B.K.P. Horn, "Rotationally symmetric operators for surface interpolation", *Computer Vision Graphics and Image Processing* **22**, 70-94 (1983).

23. B.K.P. Horn "Understanding image intensities", *Artificial Intelligence* **8**, 201-231 (1977).

24. R.J. Woodham "Analysing images of curved surfaces", *Artificial Intelligence* **17**, 117-140 (1981).

25. K. Ikeuchi and B.K.P. Horn, "Numerical shape from shading and occluding boundaries", *Artificial Intelligence* **17**, 141-184 (1981).

26. D. Terzopoulos "Image analysis using multigrid relaxation methods", *IEEE Transactions on Pattern Analysis and Machine intelligence* **8**, 129-139 (1986).

27. D. Hilbert and S. Cohn-Vossen, *Geometry and the Imagination*, Chelsea Publishers, New York, NY, 1952.

28. D. Marr "Analysis of occluding contour", *Proc. R. Soc. Lond. B* **197**, 441-475 (1977).

29. F.E. Nicodemus J.C. Richmond, J.J. Hsia, I.W. Ginsberg, and T. Limperis, "Geometrical considerations and nomenclature for reflectance", *NBS Monograph 160, National Bureau of Standards*, Washington, DC, 1977.

30. B.K.P. Horn and R.W. Sjoberg, "Calculating the reflectance map", *Applied Optics* **18**, 1770-1779 (1979).

31. A.N. Tikhonov and V.Y. Arsenin, *Solutions of ill-posed problems*, V.H. Winston & Sons, New York, NY, 1977.

32. T. Poggio and V. Torre, "Ill-posed problems and regularization analysis in early vision", *AI-Memo-773, MIT AI Laboratory*, Cambridge, MA, 1984.

33. R.J. Woodham "Photometric method for determining surface orientation from multiple images", *Optical Engineering* **19**, 139-144 (1980).

34. B.K.P. Horn and K. Ikeuchi, "The mechanical manipulation of randomly oriented parts", *Scientific American* **251**, 100-111 (1984).

35. D.H. Ballard "Generalizing the Hough transform to detect arbitrary shapes", *Pattern Recognition* **13**, 111-122 (1981).

36. H. Blum and R.N. Nagel, "Shape description using weighted symmetric axis transform", *Pattern Recognition* **10**, 167-180 (1978).

37. M. Brady and H. Asada, "Smoothed local symmetries and their implementation", in M. Brady and R. Paul eds, *Robotics Research: The First International Symposium*, MIT Press, Cambridge, MA, pp 331-354, 1984.

38. H.G. Barrow and J.M. Tenenbaum, "Recovering intrinsic scene characteristics from images", in A.R. Hanson and E.M. Riseman eds, *Computer Vision Systems*, Academic Press, New York, NY, pp 3-26, 1978.

39. B.K.P. Horn "Sequins and quills - representations for surface topography", *AI-Memo-536, MIT AI Laboratory*, Cambridge, MA, 1979.

40. A.K. Mackworth "Interpreting pictures of polyhedral scenes", *Artificial Intelligence* **4**, 121-137 (1973).

41. S.W. Draper "The use of gradient and dual space in line-drawing interpretation", *Artificial Intelligence* **17**, 461-508 (1981).

42. S.A. Shafer T. Kanade, and J.R. Kender, "Gradient space under orthography and perspective", *Computer Vision Graphics and Image Processing* **24**, 182-199 (1983).

43. K.A. Stevens "The visual interpretation of surface contours", *Artificial Intelligence* **17**, 47-75 (1981).

44. T. Kanade and J.R. Kender, "Mapping image properties into shape constraints: skewed symmetry, affine-transformable patterns, and the shape-from-texture paradigm", in J. Beck, B. Hope, and A. Rosenfeld eds, *Human and Machine Vision*, Academic Press, New York, NY, pp 237-257, 1983.

45. A.A.G. Requicha "Representations for rigid solids: theory, methods, and systems", *ACM Computing Surveys* **12**, 437-464 (1980).

46. R. Nevatia and T.O. Binford, "Description and recognition of curved objects", *Artificial Intelligence* **8**, 77-88 (1977).

47. G.J. Agin and T.O. Binford "Computer description of curved objects", *Proc. of the Third International Joint Conference on Artificial Intelligence*, Palo Alto, CA, 629-640 (1973).

48. R.A. Brooks "Symbolic reasoning among 3-D models and 2-D images", *Artificial Intelligence* **17**, 285-348 (1981).

49. B.K.P. Horn "Extended Gaussian images", *Proc. IEEE* **72**, 1671-1686 (1984).

50. J.J. Little "An iterative method for reconstructing convex polyhedra from extended Gaussian images", *Proc. of the Third National Conference on Artificial Intelligence*, Washington, DC, 247-250 (1983).

51. J.J. Little "Determining object attitude from extended Gaussian images", *Proc. of the Ninth International Joint Conference on Artificial Intelligence*, Los Angeles, CA, 960-963 (1985).

52. J.M. Tenenbaum and H.G. Barrow, "Experiments in interpretation-guided segmentation", *Artificial Intelligence* **8**, 241-274 (1977).

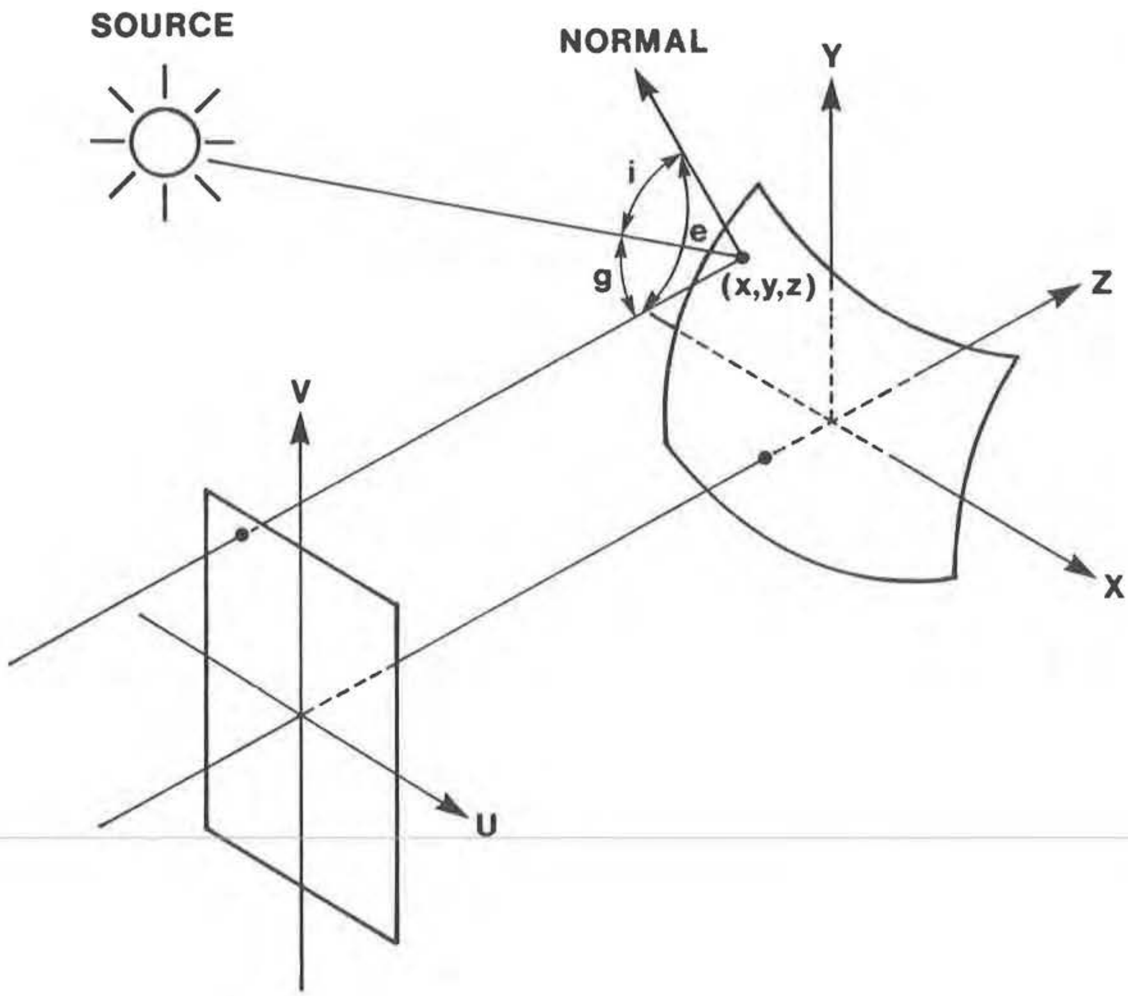

Figure 1. In an orthographic projection all rays from object surface to image are parallel. With appropriate scaling of the image plane, image coordinates (x,y) and surface coordinates (x,y) can be used interchangeably. The incident angle i is the angle between the incident ray and the surface normal. The emergent angle e is the angle between the emergent ray and the surface normal. The phase angle g is the angle between the incident and emergent rays.

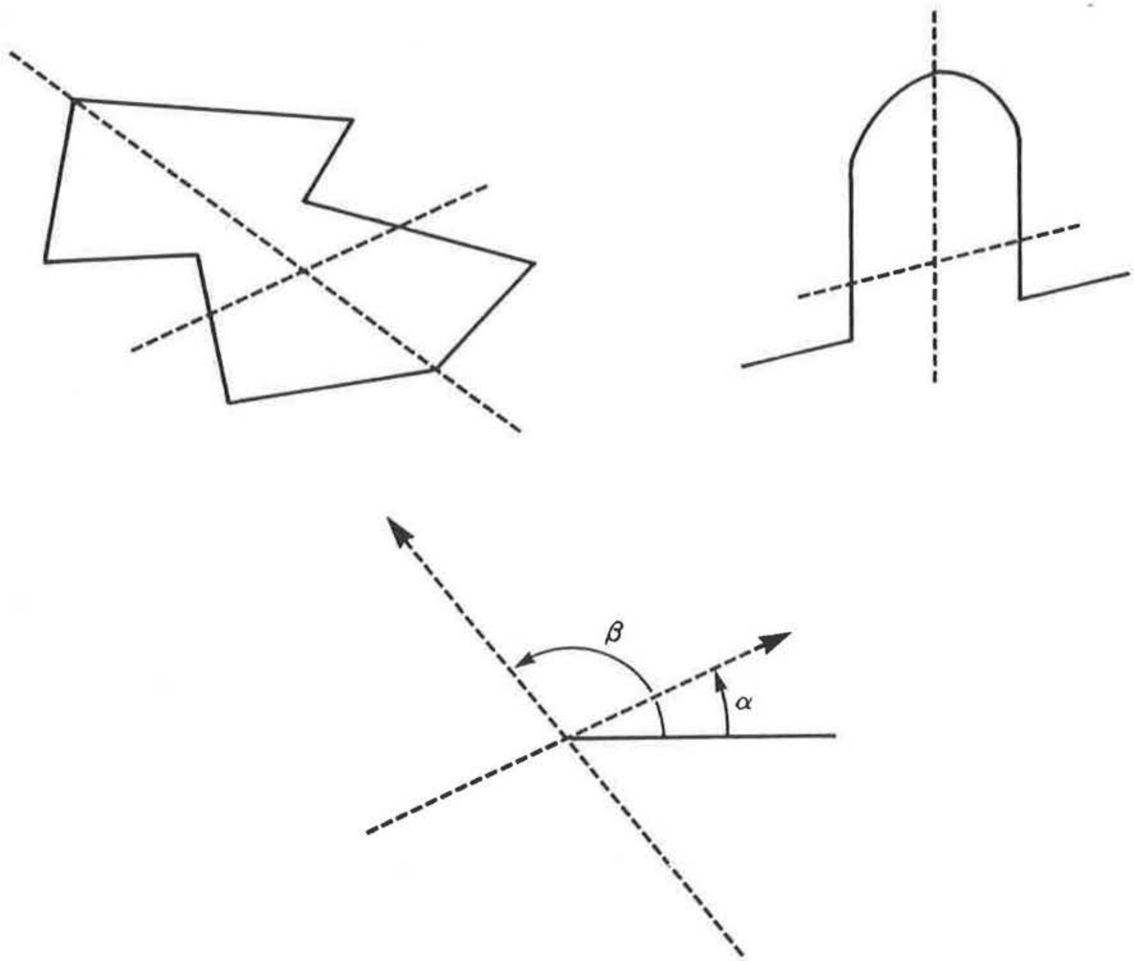

Figure 2. Examples of skew symmetry. A skew symmetry defines two directions: the skewed transverse axis and the skewed symmetry axis. These directions are denoted α and β in the figure.