Stable Representation of Shape

by

R.J. Woodham[1]

Technical Report 87-5

February 1987

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

[1] Fellow of the Canadian Institute for Advanced Research.

1. INTRODUCTION

Computational vision is the study of intelligent systems that produce descriptions of a world from images of that world. The purpose is to determine those aspects of the world that are required to carry out some task. For most tasks, shape is a necessary component of any description produced. Thus, shape representation is a central concern to designers of computer vision systems.[1]

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 2D 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 3D shape that are independent of viewpoint. Computational vision thus distinguishs three levels of representation: 2D image, visible surfaces, and 3D objects [1-5]. The principal shape representations considered at each of these levels are discussed in [6].

---

[1] To avoid confusion, 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.

This chapter identifies stability as one of several design criteria that a shape representation should satisfy. The definition of stability adopted here is the standard one from numerical analysis. That is, a computation is stable if small changes in the input produce correspondingly small changes in the output. This definition is well-routed in mathematics but it is difficult to make precise when dealing with symbolic descriptions. One small step in this direction is the use of the mixed volume as a measure of the similarity between two convex polyhedra. Technical details are found in the Ph.D. thesis of Little[7] and related publications [8-10]. Here, we interpret Little's results, compared to alternatives, in terms of stability.

Section 2 outlines a general approach to vision research and identifies one particular research strategy to follow. Section 3 describes design criteria for shape representation. Section 4 introduces orientation-based representations for 3D shape and describes a prototype vision system suitable for automatic bin-picking. Section 5 provides the necessary background detail for convex polyhedra. Section 6 discusses similarity measures for convex polyhedra. Concluding remarks are in Section 7.

2. AN APPROACH TO VISION RESEARCH

A complete theory of human vision must account for the relationship between the natural world and human visual perception. It is here taken as a given the the natural world consists of 3D objects and that perceptions ultimately can be represented as symbolic descriptions. In this view, the 2D image acts as an intermediate representation that mediates between the 3D world and visual perception. Figure 1 illustrates. To

understand the relationship between the 3D world and perception, there are four components to consider, as suggested by the four labelled arrows in Figure 1.

Arrow 1 characterizes the mapping from the 3D world to the 2D image. Given a spatial arrangement of objects made of a particular set of materials and illuminated in a particular way, the laws of physical optics determine the image. Geometric equations determine where each point on a visible surface will appear in the image and corresponding radiometric equations determine its brightness and colour. This is properly the domain of computer graphics, although it is relevant to point out that computer graphics is primarily concerned with producing "realistic" images, to convey information, as opposed to images that necessarily depict physical reality.

Arrow 2 characterizes the inverse mapping from 2D image to 3D world. All so-called "shape from" methods, including shape from binocular stereo [11-12], shape from shading [13-17] and photometric stereo [18], shape from contour [19-23], shape from motion [24] and optical flow [25], and shape from texture [19,26] are formulated as problems at this level. Since the mapping from the 3D world to the 2D image (arrow 1) is many-to-one, the inverse mapping (arrow 2) is underconstrained. That is, there are many 3D worlds that produce the identical 2D image. In most situations, the inverse problem is ill-posed in that there is no unique, stable solution [27].

Arrow 3 characterizes the mapping from the 2D image to perception. This is properly the domain of perceptual psychology. The mapping from the 2D image to perception also is many-to-one in that many different 2D images produce identical perceptions. Familiar examples include colour and texture metamers and edge contrast effects, such as the Cornsweet illusion. Many constraints on the visual perception of 2D images

derive from the assumption that a 2D image is an image of a 3D world. Thus, one cannot deal only with the perception of images (arrow 3). Necessarily, one must consider the relationship between 3D worlds and their 2D images (arrows 1 and 2).

Arrow 4 characterizes the inverse mapping from a perception to the 2D image. Again, this inverse mapping is underconstrained. But, the question can be posed as, "What is the equivalence class of images that produces a given perception?" Answering this question corresponds to identifying perceptual metamers.

2.1 The Research Strategy

"Shape from" methods define the computational task in terms of arrow 2 of Figure 1. That is, the task is to determine the 3D shape of objects from their 2D projection onto images. Although each method differs considerably in precise detail, all share a common characterization as computational tasks. Each embodies the following steps:

- *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 3D 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 (arrow 1 of Figure 1). The equations determine the mapping from scene to image. Shape analysis, however, requires a solution to the inverse problem. That is, one must determine the mapping from image to scene (arrow 2 of Figure 1).

- *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 stable solution to the inverse problem.* Image features determine equivalence classes of possible scene features. Conceptually, a unique stable solution is obtained when a suitable metric is applied to the equivalence classes to select a single preferred solution. 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 rigour is generally required to demonstrate that a particular choice does, in fact, lead to a unique stable solution. Finally, even when the existence of the desired solution is established, it is still necessary to develop an algorithm to compute 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.

3. SHAPE REPRESENTATION

Several authors propose design criteria that general-purpose shape representations should satisfy [4,6,28-30]. 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 designed criteria are summarized below:

- *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 2D 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 3D 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 and 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.

4. ORIENTATION-BASED REPRESENTATION OF SHAPE

A *depth map* is one way to represent the shape of a visible surface. A depth map determines distance to the surface along parallel rays on a dense, regularly spaced grid. A range finder is a sensor that produces surfaces descriptions of this form. Several "shape from" methods, including shape from binocular stereo and shape from motion, also produce surface descriptions in the form of a depth map. The depth map representation has certain deficiencies when the task is to determine 3D object identity, position and attitude. For one thing, depth maps do not transform in a simple way when the object rotates or, equivalently, when the viewpoint changes. Thus, it is difficult to compare a sensed depth map directly to stored object models.

Alternatively, one can represent the shape of a visible surface by specifying surface orientation on a dense, regularly spaced grid. This representation has been called a *needle diagram* by Horn [5]. Photometric stereo [18] produces surface descriptions of this form. Other "shape from" methods, including shape from shading and shape from texture gradient, also produce surface descriptions in the form of a needle diagram. A depth map description can be transformed into a needle diagram by numerical differentiation. The needle diagram itself still depends on both the position and attitude of the object in view. It is possible, however, to transform a needle diagram in a simple way to compute an *orientation histogram*. The orientation histogram corresponds to a hemisphere of the object's Extended Gaussian Image (EGI), as we shall see. Object identity and attitude is readily determined by comparing a sensed orientation histogram to object models stored using the EGI representation.

Consider the continuous case illustrated in Figure 2. Let U be the unit sphere, termed the *Gaussian sphere*. Each point on the Gaussian sphere identifies the unit vector formed by joining the origin to that point. Let E be a portion of a surface S bounded by a closed curve. The image of E under the *Gauss map*

$$G(p) = \omega, \quad p \in S, \omega \text{ unit normal at } p$$

is $G(E)$, the *Gaussian image* of E.

The Gauss map allows us to define three related concepts. First, the *Gaussian curvature* at p, denoted by $K(p)$, is

$$K(p) = \lim_{|E| \to 0} \frac{|G(E)|}{|E|}$$

where E is a compact region on S enclosing p. Second, the *area function* at ω, denoted by $A(\omega)$, is

$$A(\omega) = \lim_{|R| \to 0} \frac{|G^{-1}(R)|}{|R|}$$

where R is a compact region on U enclosing ω. Third, a 3D object's *Extended Gaussian image* (EGI) is simply its area function defined on U. The EGI of an object records the variation of surface area with surface orientation, using U as the reference system. If S has strictly positive curvature, $p \in G^{-1}(\omega)$ is unique. That is to say, the EGI uniquely represents convex objects (up to translation) and is thus information preserving for this class of objects.

In the case of polyhedra, the set of surface orientations is finite. Let $\{\omega_i\}$ be the set of unit (outward) surface normals of the planar faces of a given convex polyhedron. Then its EGI can be represented as $\{A(\omega_i)\}$ where $A(\omega_i)$ is the area of the face with orientation ω_i. One can imagine translating each ω_i to a common point of application

and scaling it so that its length is $A(\omega_i)$. This produces another representation, equivalent to the EGI, that has been called a *spike model* [5].

The EGI is invariant to translation and can be normalized to be invariant also to scale. 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. (See [31] for a primer on extended Gaussian images.)

Figure 3 illustrates a prototype vision system for automatically picking parts out of bin. Versions of this system have been built and reported in the literature[32-34]. A 3D object, made of a given material and illuminated in a given way, determines the 2D image. Using photometric stereo, or other techniques, a description of the visible surface is computed from the image in the form of a needle diagram. The needle diagram is converted into a discrete orientation histogram, using a standard tesselation of the Gaussian sphere U. The EGI's of known objects are stored internally in the form of discrete orientation histograms. The measured orientation histogram determines a hemisphere of the object's EGI. Object recognition is achieved by matching the visible hemisphere to the correct stored EGI. The three degrees of freedom in object attitude also are determined by the position and orientation of the visible hemisphere at the correct match. This allows a robot to pick mixed parts out of a bin.

Figure 3 provides the overview. Unfortunately, the matching computation can be ill-conditioned. Recently, the mixed-volume has been used as the similarity measure. The result is more robust than direct EGI matching and can support efficient multiresolution attitude determination.

5. CONVEX POLYHEDRA

The discussion of convex polyhedra given here follows Little[7]. (See [35-36] for a more comprehensive treatment.)

5.1 Basic Definitions

A *plane* can be represented as the set

$$\{ x \mid <\omega,x> = c \}$$

where $< \cdot , \cdot >$ denotes vector inner product, ω is a unit normal vector to the plane and c is a scalar constant. A plane also defines the *half space* given by

$$\{ x \mid <\omega,x> \leq c \}$$

The intersection of a finite number of half spaces forms a *convex polyhedron* denoted by

$$\bigcap_{i=0}^{n} \{ x \mid <\omega_i,x> \leq c_i \} \tag{1}$$

A bounded polyhedron is termed a *polytope*. A 2D polytope is called a *polygon*. A 3D polytope is called a *polyhedron*.

The *support function*, $H(\omega)$, of a convex polytope is defined as the perpendicular distance to the closest tangent plane, with normal vector ω, that touches but does not pass into the interior of the polytope. Distance is measured from an arbitrary origin chosen inside the polytope. Figure 4 illustrates the geometric construction of the support function, $H(\omega)$, for a sample polygon. The polygon is shown in Figure 4(a). The orientation, ω, is defined with respect to an origin, interior to the polygon, and a reference direction, as shown in Figure 4(b). For any ω, consider the tangent line having normal vector in the direction ω. $H(\omega)$ is the distance of closest approach of the tangent line to the origin. The points of closest approach all lie on a set of circles, as shown in

Figure 4(c). Each circle passes through the origin and one vertex of the polygon. The resulting support figure is drawn in Figure 4(d). (For polyhedra, the support function, $H(\omega)$, determines a piece-wise spherical support figure.) Even though the set of orientations, $\{\omega_i\}$, corresponding to faces is discrete, the support function, $H(\omega)$, is a defined for all ω.

The volume of a convex polytope can be expressed in terms of its area function, $A(\omega_i)$, and support function, $H(\omega_i)$. For the 2D case, the area of the polygon is $1/2 <H(\omega_i),A(\omega_i)>$. For the 3D case, the volume of the polyhedron is $1/3 <H(\omega_i),A(\omega_i)>$. In all cases, the measure is independent of choice of origin in the definition of $H(\omega)$.

Let $\{\omega_i\}$ be a fixed set of orientation vectors. Then, a convex polytope can be uniquely represented in terms of $\{\omega_i\}$ either by specifying the corresponding set $\{c_i\}$ or the corresponding set of areas $\{A(\omega_i)\}$. If $\{c_i\}$ is given, the polytope is determined by equation (1). At first, it might seem that some information is lost if only $\{A(\omega_i)\}$ is given since the positions of the surface normals is not preserved. Said another way, the adjacency information of the faces is not explicit. Minkowski first showed in 1897 that the EGI (i.e., $\{A(\omega_i)\}$) uniquely determines (up to translation) a convex polyhedron. Reconstructing the convex polyhedron from its EGI was first demonstrated using an iterative algorithm developed by Little [8].

Now, two polyhedra P and Q are *homothetic* if

$$P = \{\, x \mid x = \lambda \cdot y + t,\ y \in Q,\ \lambda \in \mathbb{R}^1,\ t \in \mathbb{R}^3 \,\}$$

That is, two polyhedra are homothetic if they differ only by a translation and by a scal-

ing. (Two polyhedra are not homothetic if they differ by a rotation.) If two polytopes are homothetic, their EGI's can be made equal by a scaling.

The *convex sum (or mixture)* of two polyhedra P and Q is

$$\lambda \cdot P + (1-\lambda) \cdot Q = \{ \lambda \cdot x + (1-\lambda) \cdot y \mid x \in P, y \in Q, 0 \leq \lambda \leq 1 \}$$

Let $R = \lambda \cdot P + (1-\lambda) \cdot Q$ where $0 \leq \lambda \leq 1$. Then, the EGI, $A_R(\omega)$, of the convex mixture, R, is the convex sum of the EGI's $A_P(\omega)$ and $A_Q(\omega)$. That is,

$$A_R(\omega) = \lambda \cdot A_P(\omega) + (1-\lambda) \cdot A_Q(\omega)$$

Similarly,

$$H_R(\omega) = \lambda \cdot H_P(\omega) + (1-\lambda) \cdot H_Q(\omega)$$

The support function, $H_R(\omega)$, of the convex mixture, R, is the convex sum of the support functions $H_P(\omega)$ and $H_Q(\omega)$.

5.2 The Mixed-Volume

The volume, $V(R)$, of the convex mixture, R, is more difficult to express. For the 3D case, one obtains

$$V(R) = \lambda^3 V(P) + 3\lambda^2(1-\lambda) V_3(P,Q) + 3\lambda(1-\lambda)^2 V_3(Q,P) + (1-\lambda)^3 V(Q) \qquad (2)$$

where

$$V_3(S,T) = \frac{1}{3} <H_S(\omega_i), A_T(\omega_i)>$$

is called the *mixed volume* of S and T. $V_3(S,T)$ is the inner product of the support function of S with the area function of T. For the 3D case, the mixed volume is not, in general, symmetric (i.e., $V_3(S,T) \neq V_3(T,S)$). For the 2D case, the mixed volume is symmetric.

As above, let $R = \lambda \cdot P + (1 - \lambda) \cdot Q$ be the convex sum of two convex polyhedra P and Q. Then, by the *Brunn-Minkowski Theorem*,

$$V(R)^{1/3} \geq \lambda\, V(P)^{1/3} + (1 - \lambda)\, V(Q)^{1/3} \tag{3}$$

with equality if and only if P and Q are homothetic. Combining equations (2) and (3), we obtain

$$V_3(P,Q)^3 \geq V(Q)^2\ V(P)$$

with equality holding if and only if P and Q are homothetic. The mixed volume captures the relationship between the shapes of two polytopes. When the mixed volume is minimal, the polytopes are homothetic. Mixing the two does not cause a shape change, only a scaling. Little's iterative algorithm [8] minimizes the mixed volume to reconstruct object shape, determined by $H(\omega_i)$, from the given EGI, $A(\omega_i)$.

6. SIMILARITY MEASURES FOR CONVEX POLYHEDRA

When the task is object recognition, the similarity measure should be invariant to translation and scaling, since these typically vary with viewpoint rather than with object identity. Rotation also is a viewpoint dependent measure. When the task requires the determination of object attitude, the similarity measure must be sensitive to rotation. The mixed volume captures exactly these properties.

Initially we were mislead into thinking the adjacency structure of a polytope was important. (Adjacency structure determines which faces share an edge, which edges meet at a vertex and which vertices lie on a face.) Interestingly, Little's reconstruction method, based on the mixed volume, does not deal explicitly with adjacency structure.

Adjacency structure is determined as a consequence of the minimization and can change many times as the algorithm iterates. It is also important to note that adjacency structure is not a stable property of a polytope. Small changes in the relative positions of faces can produce large changes in the adjacency structure.

6.1 Determining Attitude by Comparing Area Functions

Determining the attitude of a known object is equivalent to finding the rotation, ϕ, that brings the known area function into correspondence with the sensed area function. Let $\{A(\omega_i)\}$ be the sensed area function of the visible surface of an object. Let $\{A_\phi(\omega_i)\}$ be the area function of prototype object with rotated attitude ϕ. At each sampled attitude ϕ, the measure of similarity is given by

$$\sum_i (A_\phi(\omega_i) - A(\omega_i))^2$$

Determining the best match corresponds to finding the attitude, ϕ, that minimizes this measure. (Equivalently, one can find the attitude, ϕ that maximizes $<A_\phi(\omega_i), A(\omega_i)>$.)

A direct comparison of area functions is the method used by Horn and Ikeuchi [32]. But, there are difficulties. As the resolution is increased, effectiveness of area matching decreases. The tesselation of the Gaussian sphere, U, becomes finer resulting in more empty cells in the orientation histogram of the known object. Thus, even when the attitude difference between the sensed and the known area functions is small, the match may be poor. In fact, the match may be poor even at the correct attitude. Said another way, the similarity measure used is not stable.

6.2 Determining Attitude with Mixed Volumes

Let $\{A(\omega_i)\}$ be the sensed area function of the visible surface of an object. Let $\{H_\phi(\omega_i)\}$ be the support function of the known object with rotated attitude ϕ. Rotating an object preserves volume. Thus, the mixed volume

$$<H_\phi(\omega_i), A(\omega_i)>$$

is minimized at the attitude, ϕ, that brings the known object into correspondence with the sensed object.

For polytopes, the area function, $A(\omega)$, is non-zero for only finitely many values of ω. When the attitude of the sensed object is slightly different from that of the known object, the correlation of area functions can be zero. (It is possible to consider smoothing the area functions directly, to improve correlation, as Brou has observed [37].) The support function, $H(\omega)$, is a continuous function of ω and the mixed volume achieves this smoothing in a more rigourous way.

Little's experiments [7] support this approach. The effects of the magnitude of the difference in attitude between object and prototype are significantly smaller for the mixed volume method. This suggests that it is possible to trade-off resolution on the Gaussian sphere, U, with the number of test attitudes, ϕ, to achieve an efficient and accurate coarse-to-fine determination of object attitude. The justification for the mixed-volume method depends on the area function and support function being defined over the entire Gaussian sphere. But, it appears to work well in practice when only the visible hemisphere of the sensed object is available.

7. CONCLUDING REMARKS

The prototype bin-picking system demonstrates that robust, practical machine vision systems can be designed and built. The approach is based on a careful analysis of the physics of imaging and on the view of machine vision as an inverse problem. This resulted in the concepts of photometric stereo, the Extended Gaussian image and mixed volumes.

The research strategy exemplified by this work is applicable to a broader class of vision tasks. The discussion of the research strategy and of design criteria for shape representations is intended to suggest that this is so. We have only just begun, however, and much remains to be learned.

Stability, in particular, is relevant to a number of the design criteria discussed above. It is not yet clear how the computations of early vision, which are primarily numeric in form, can interface with knowledge representations, which are primarily symbolic in form.

Currently, we do not know how to define stability for symbolic representations. The basic mechanism to compare two symbolic expressions is to test for equality. More general matching of expressions is possible when syntactic transformations are allowed to reduce the comparison to an equality test. For example, the substitution of terms in one expression for variables in the other to make the expressions identical is called *unification*. Algorithms exist to find the most general (i.e., simplest) unifier of any finite set of unifiable expressions, or to report failure if the set cannot be unified. Finally, when transformation includes the ability to do deductive inference, two symbolic expres-

sions can be considered equivalent if each implies the other. Stability, as it's currently defined, requires the ability to quantify similarities and differences. If a significant portion of a computational vision system is to be symbolic, rather than numeric, it will be necessary extend the notion of stability to cover the symbolic domains too.

ACKNOWLEDGMENTS

This report describes research done at the Laboratory for Computational Vision of the University of British Columbia. Support for the Laboratory's research is provided, in part, by the UBC Interdisciplinary Graduate Program in Remote Sensing, by the Natural Sciences and Engineering Research Council of Canada (NSERC) under grants SMI-51, A0383 and E0008, and by the Canadian Institute for Advanced Research. The work on shape representation also was supported, in part, by NSERC grant A3390.

Technical details on convex polyhedra and the mixed volume derive from the Ph.D. thesis of J.J. Little. The author also thanks J.J. Little and A.K. Mackworth for many discussions on aspects of shape representation.

REFERENCES

[1] Marr, D. (1982), *Vision*, W.H. Freeman, San Francisco, CA.

[2] Barrow, H.G. & J.M. Tenenbaum (1981), "Computational vision", *Proc. IEEE* (69)572-595.

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

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

[5] Horn, B.K.P. (1986), *Robot Vision*, MIT Press/McGraw-Hill, Cambridge, MA.

[6] Woodham, R.J. (1987), "Shape analysis", in *Encyclopedia of Artificial Intelligence*, S. Shapiro (ed.), (in press), John Wiley & Sons, New York, NY.

[7] Little, J.J. (1985), "Recovering shape and determining attitude from extended Gaussian images", TR-85-2, UBC Dept. of Computer Science, Vancouver, BC.

[8] Little, J.J. (1983), "An iterative method for reconstructing convex polyhedra from extended Gaussian images", *Proc. 3rd National Conference on Artificial Intelligence*, pp 247-250, Washington, DC.

[9] Little, J.J. (1985), "Extended Gaussian images, mixed volumes, and shape reconstruction", *Proc. ACM Symposium on Computational Geometry*, pp 15-23, Baltimore, MD.

[10] Little, J.J. (1985), "Determining object attitude from extended Gaussian images", *Proc. 9th International Joint Conference on Artificial Intelligence*, pp 960-963, Los Angeles, CA.

[11] Baker, H.H. & T.O. Binford (1981), "Depth from edge and intensity based stereo", *Proc. 7th International Joint Conference on Artificial Intelligence*, pp 631-636, Vancouver, BC.

[12] Grimson, W.E.L. (1981), *From Images to Surfaces*, MIT Press, Cambridge, MA.

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

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

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

[16] Woodham, R.J. (1984), "Photometric method for determining shape from shading", in *Image Understanding 1984*, S. Ullman & W. Richards (eds.), pp 97-125, Ablex Publishing Corp., Norwood, NJ.

[17] Horn, B.K.P. & M.J. Brooks (1986), "The variational approach to shape from shading", *Computer Vision Graphics and Image Processing* (33)174-208.

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

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

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

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

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

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

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

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

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

[27] Poggio, T. & V. Torre (1984), "Ill-posed problems and regularization analysis in early vision", AI-Memo-773, MIT AI Laboratory, Cambridge, MA.

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

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

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

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

[32] Horn, B.K.P. & K. Ikeuchi (1984), "The mechanical manipulation of randomly oriented parts", *Scientific American* (251)100-111 (August, 1984).

[33] Ikeuchi, K., B.K.P. Horn, S. Nagata, T. Callahan & O. Feingold (1984), "Picking up an object from a pile of objects", in *Robotics Research: The First International Symposium*, M. Brady & R. Paul (eds.), pp 139-162, MIT Press, Cambridge, MA.

[34] Ikeuchi, K., H.K. Nishihara, B.K.P. Horn, P. Sobalvarro & S. Nagata (1986), "Determining grasp configurations using photometric stereo and the PRISM binocular stereo system", *International Journal of Robotics Research*, 5(1)46-65.

[35] Lyusternik, L.A. (1963), *Convex Figures and Polyhedra*, Dover Publications, New York, NY.

[36] Grunbaum, B. (1967), *Convex Polytopes*, John Wiley & Sons, New York, NY.

[37] Brou, P. (1984), "Using the Gaussian image to find the orientation of objects", *International Journal of Robotics Research* 3(4)89-125.

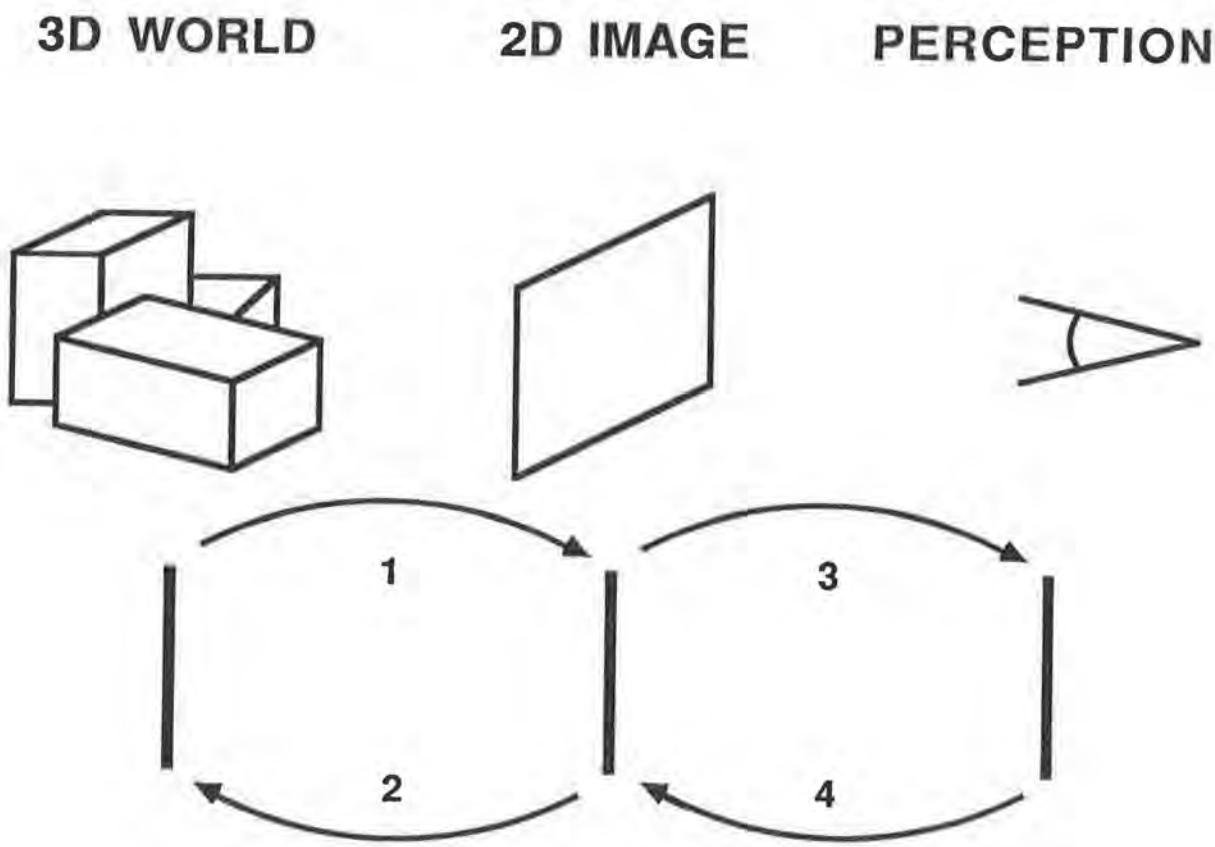

Figure 1. Visual perception of the natural world. The natural world consists of 3D objects made out of different materials illuminated in different ways. An image is a spatially varying brightness pattern. Visual perception consists of symbolic descriptions of the natural world computed from the image. There are four mappings to describe. Arrow 1 is the mapping from the 3D world to the 2D image. Arrow 2 is the inverse mapping from the 2D image to the 3D world. Arrow 3 is the mapping from the 2D image to perception. Arrow 4 is the inverse mapping from a perception to the 2D image. Since the mappings 1 and 3 are, in general, many-to-one, the inverse mappings 2 and 4 determine equivalence classes, respectively, of worlds that produce identical images and of images that produce identical perceptions.

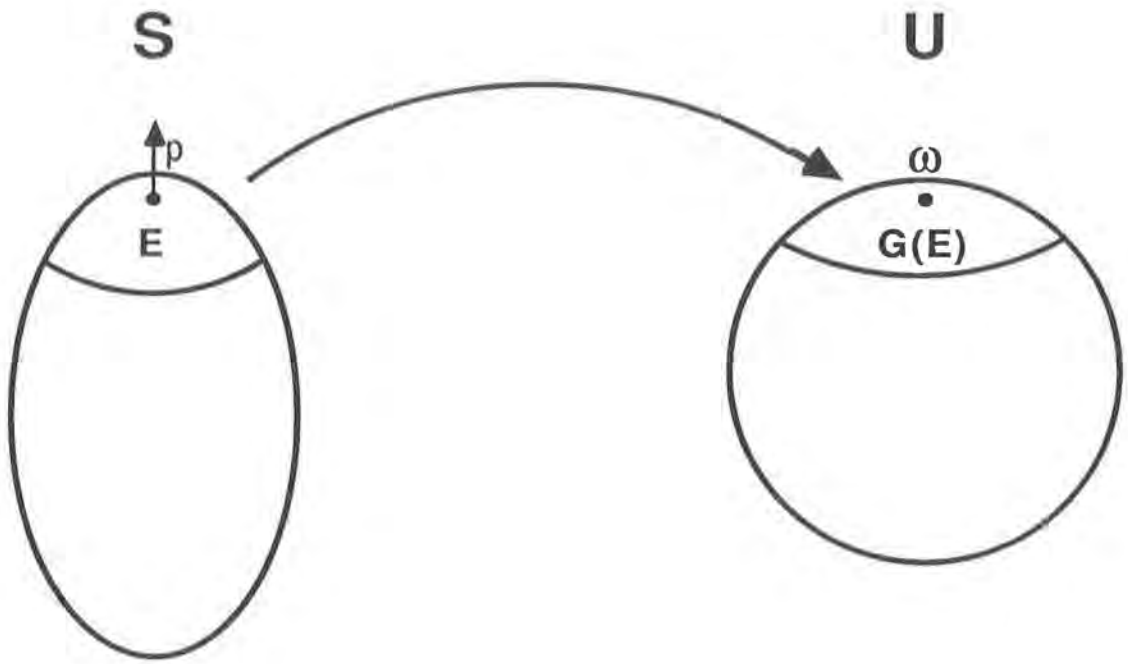

Figure 2. The Gauss map. Let p be a point on a surface S. The unit surface normal vector at p determines a point ω on the Gaussian sphere U. $G(E)$ is the Gaussian image of E under the Gauss map $G(p) = \omega$.

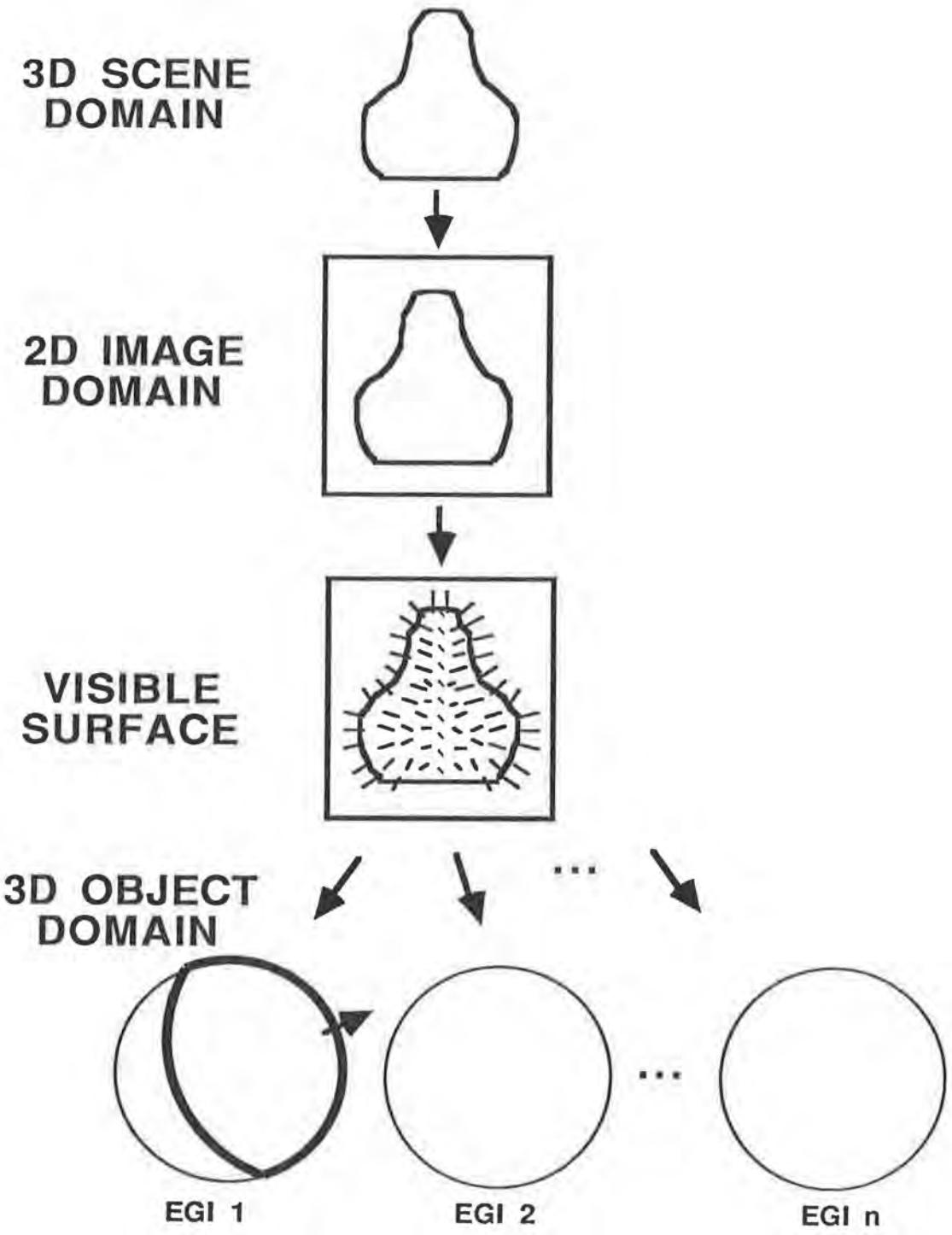

Figure 3. A 3D object, made of a given material and illuminated in a given way, determines the 2D image. Using photometric stereo, or other techniques, a description of the visible surface is computed from the image. Here, the description is of surface orientation at each visible point, called a needle diagram. Each of several 3D object's is represented by its Extended Gaussian Image (EGI). The needle diagram determines a hemisphere of the object's EGI. Matching the visible hemisphere to known EGI's determines both object identity and object attitude. This, for example, allows a robot to pick mixed parts out of a bin.

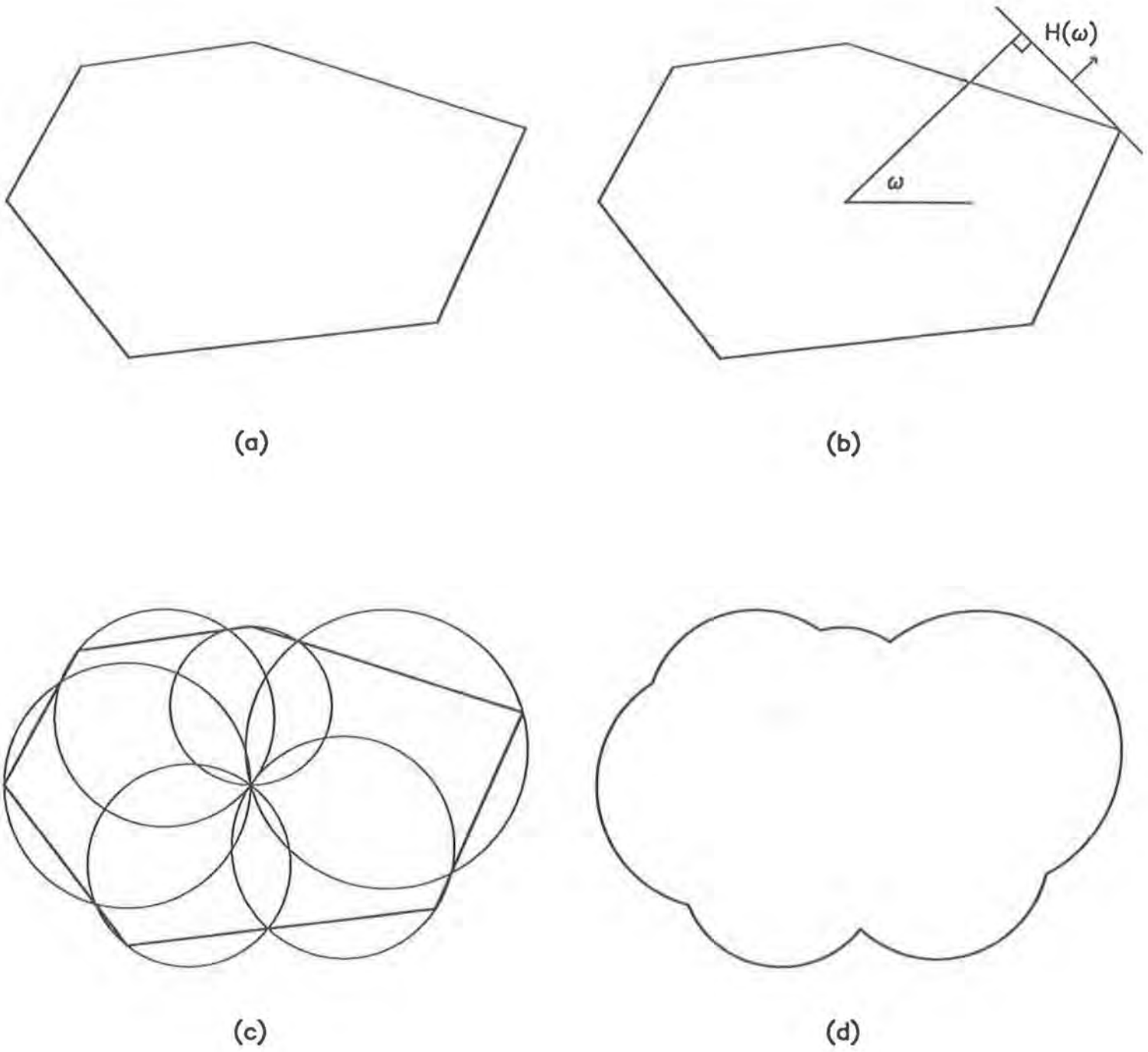

Figure 4. Geometric construction illustrating the definition of a support function, $H(\omega)$, shown here for the 2D case. A sample polygon is shown in (a). Orientation, ω, is defined with respect to an origin, interior to the polygon, and a reference direction, as shown in (b). For any ω, consider the tangent line having normal vector in the direction ω. $H(\omega)$ is the distance of closest approach of the tangent line to the origin. The points determined by $H(\omega)$ all lie on a set of circles, as shown in (c). Each circle passes through the origin and one vertex of the polygon. The resulting support figure is shown in (d).