PROCEEDINGS

Of SPIE-The International Society for Optical Engineering

Volume 726

Intelligent Robots and Computer Vision

Fifth in a Series

David P. Casasent
Chair/Editor

Sponsored by
SPIE—The International Society for Optical Engineering

Cooperating Organizations
Carnegie Mellon University
Sira Ltd.
Tufts University/Electro-Optics Technology Center

28-31 October 1986
Cambridge, Massachusetts

Proceedings of SPIE—The International Society for Optical Engineering

Volume 726

Intelligent Robots and Computer Vision

Fifth in a Series

David P. Casasent
Chair/Editor

Sponsored by
SPIE—The International Society for Optical Engineering

Cooperating Organizations
Carnegie Mellon University
Sira Ltd.
Tufts University/Electro-Optics Technology Center

28–31 October 1986
Cambridge, Massachusetts

Published by
SPIE—The International Society for Optical Engineering
P.O. Box 10, Bellingham, Washington 98227-0010 USA
Telephone 206/676-3290 (Pacific Time) • Telex 46-7053

SPIE (The Society of Photo-Optical Instrumentation Engineers) is a nonprofit society dedicated to advancing engineering and scientific applications of optical, electro-optical, and optoelectronic instrumentation, systems, and technology.

The papers appearing in this book comprise the proceedings of the meeting mentioned on the cover and title page. They reflect the authors' opinions and are published as presented and without change, in the interests of timely dissemination. Their inclusion in this publication does not necessarily constitute endorsement by the editors or by SPIE.

Please use the following format to cite material from this book:
Author(s), "Title of Paper," *Intelligent Robots and Computer Vision: Fifth in a Series,* David P. Casasent, Editor, Proc. SPIE 726, page numbers (1987).

Library of Congress Catalog Card No. 86-63728
ISBN 0-89252-761-7

Printed in the United States of America.

Shape and Function

Franc Solina and Ruzena Bajcsy

General Robotics and Active Sensory Perception Laboratory
UNIVERSITY of PENNSYLVANIA
Department of Computer and Information Sciences
Philadelphia, Pennsylvania 19104-6389

Abstract

We propose a modeling system for generic objects in order to recognize different objects from the same category with only one generic model. The representation consists of a prototype, represented by parts and their configuration. Parts are modeled by superquadric volumetric primitives which are combined via Boolean operations to form objects. Variations between objects within a category are described by allowable changes in structure and shape deformations of prototypical parts. Each prototypical part and relation has a set of associated features that can be recognized in the images. These features are used for selecting models from the model data base. The selected hypothetical models are then verified on the geometric level by deforming the prototype in allowable ways to match the data. We base our design of the modeling system upon the current psychological theories of categorization and of human visual perception.

Introduction

Computer vision has not yet adequately addressed the problem of recognizing generic objects. Most current vision systems must have a precise model of any particular object that they are supposed to recognize. This is because the present shape representations are inflexible and usually tailored to a particular domain or to a particular type of objects. These representations may well be sufficient wherever the number of objects and their shape can be controlled, but not for a vision system of a robot that has to function in an unrestricted environment. Since people handle the complexities of the world by using categories for recognition, understanding, handling, and naming of objects, we propose to use category-like models in computer vision.

Categories which people form to achieve larger cognitive economy are not arbitrary collections of objects but reflect the structure of the world. Categories that are linked most to the structure of the perceived world are basic categories.[1] Basic category is the highest level of abstraction for which a generalized outline form can be recognized and the highest level for which an image can be generated. They are the preferred level of reference, are recognized faster then superordinate or subordinate categories and children learn them first (i.e. dog vs. animal vs. German shepherd). Superordinate categories seem to share primarily functional features - vehicles are for transportation, and tools are for fixing. They do not share perceptual features in sharp contrast to basic categories which share both *functional and perceptual* features. Subordinate categories subdivide basic categories according to one or very few perceptual or functional features. Basic categories seem also to be mutually exclusive which is not the case for subordinate or superordinate categories.*

---

* There might be sometimes a marginal overlap of basic categories, like cups and mugs but both appearance and function are similar for them. Wittgenstein (1953)[2] noted that categorical judgements become a problem only if one is concerned with boundaries.

Category prototypes. Rosch[1] has shown that the categories we use contain one member that seems most representative. These best examples are called *prototypes.* We use prototypes to judge all other members of the category which are perceived *in terms of* or as *deviating from* them.[3] Experiments have shown that prototypes are the most rapidly recognized in comparison to other objects in the same category.[1] This can be because these prototypical members represent the common properties in the best way. Goldmeier[3] referred to such objects as singular because they become the standard of reference for other designs. By singular or " Prägnant " we understand the tendency to make the structure as clear cut as possible. Leyton[4] has shown that prototypification is decomposed into a sequence of well-defined stages. Prototypification occurs in parallel, at different regions (parts) of the figure and removes deformation that differentially varies over a region.

The role of parts. For objects and biological categories basic category cuts seem to follow natural breaks in the structure of the world which is determined by part configuration. Tversky and Hemenway[5] pointed out the particular salience of parts on the basic category level. In comparison to the basic level the proportion of parts of all the common features decreases for both super and subordinate categories. Parts and part configuration form a natural bridge connecting perception (appearance) of objects and behavior (activity) toward them, and in turn communication about them. Perceived part configuration underlies both perceived structure and perceived function, and forms the basis of intuitive causal reasoning and naive induction. The basis of naive induction is that separate parts have separate functions, similar parts have similar functions, and more salient parts have more important functions.

Function and *functionality* are meant here as a proper action or a design which fulfills its purpose. Functions of man-made objects are defined in terms of the user (agent) model and goal model. But parts and function are related also independent of human users when an organism or objects are studied as a self-contained system. The functional basis of shape in nature was first investigated in depth by Thompson.[6] He pointed out that the repetition of shapes is not accidental; same shapes are encountered across species for the same function since they were molded by the same physical laws. There are only a limited number of types of leaves, crystals etc. The "form follows function"† hypothesis carries over also to the man-made objects, where despite the large diversity found in the design certain *functional* dimensions must be met for adequate usage. This seems to define objects on the basic category level. The subordinate category level of man-made objects is the result of available technologies, skill in manufacturing, habits, and esthetic preference. Since this whole set of often disparate requirements influences the design process, the task of synthesis of form is not easily formalized.[7] Good designers, however, always start with the basic function in mind (Figure 1). It is interesting to note that archeologists work in a similar frame set but try instead to solve the inverse problem (reverse the arrows in Figure 1). Given excavated artifacts or only some of their parts they want to determine their function among other facts. In a way, the whole process of visual perception is to the instant visual input the same, as archeology is to the excavated artifacts.

Perceptual organization showed the importance of recovering structure and clustering of features for image understanding. A new wave of research in computer vision was triggered by this attention to the structure in the

Figure 1 Synthesis procedure for design of objects. Function or purpose that define the basic category dictate the part configuration and the shape of individual parts.

† Architect Louis Sullivan is credited for coining the phrase.

images that in turn reflects the structure in the world.[8,9,10,11,12] The visual process was reformulated as the task of discovering structure based on some primitive spatiotemporal regularities like least-distortion and non-accidentalness. The task of recovering structure, also known as perceptual organization, was first investigated by the Gestalt school in psychology. A fact that demonstrates the role of perceptual organization in vision is that people are able to perceive structure in images apart from recognizing familiar objects. Humans can see and describe the contents of images from various sources like black and white photographs, x-rays, line drawings, and electron micrographs as long as the basic structure of the image is preserved. The addition of semantic context rarely affects this spontaneous, pre-attentive organization of images into parts. The basis of this remarkable capability may be the fact that regular relationships like parallel lines, curvilinearity of arcs, symmetry, two or more terminations of vertices at a common point and so on, are very *unlikely to arise by chance*.

Partitioning of objects into parts. Building object representations by putting together smaller objects as building blocks is a common practice in computer vision and computer graphics or design. The apparent complexity of our environment is also produced from a limited vocabulary of parts by applying a small set of generic processes over and over again. Hence rather than build point-by-point descriptions of surfaces and volumes to eventually arrive at high-level models, the immediate recognition of part models in images is preferred.[10,11] While, in general, for computer modeling almost any set of primitive building blocks will do, as long as they enable easy manipulation, we want to concentrate here on parts or primitives that have a perceptual salience and hence reflect the structure in the world. When looking for a partition rule, two possibilities exist. One is to formalize the decomposition of objects into parts by defining part boundaries in terms of differential geometry.[13,9] The other is to define the shape of possible parts. Pentland[10] proposed superquadrics as a set of part primitives which very closely correspond to Hoffman's notion of parts.[9] Superquadrics can be compared to lumps of clay that can be further deformed and glued together into very realistic looking models. Since superquadrics are a set of parametric volumes where the parameters can change continuously, there are an infinite number of them.

Deformation of shape is a process that affects both natural and man-made objects. It is a highly intuitive way of describing and thinking about objects. Deformation is not just any alteration of shape. It conveys the impression that the shape as a whole has undergone a change that can be modeled by some physical process. Thompson[6] pointed out that several natural forms are easily explained when they are regarded as deformations of a simple structure. A deformation is highly intuitive and easily visualized process which helps not only to explain natural forms but simulates some manufacturing processes for fabrication of objects. Attributes that are responsible for the subdivision of a basic category into subordinate categories often correspond to some deformation. It is not surprising then that for verbal descriptions of objects we often use adjectives that reflect some underlying deformation of a simpler, prototypical shape.[14] The deformation terms are often perceived as manifestations of physical processes acting on an imagined physical object with a prototypical shape. These ideas about prototypes and deformation were taken up for shape representation in computer vision. Based on work by Barr[15,16] Pentland[10] developed a complete real-time graphics system called "SuperSketch," which uses deformations to mold superquadrics into more natural looking forms.

Motivation. Ideas from categorization, perception, and functionality meet at the basic category level which is a preferred level of abstraction. The goal of a computer vision system should be to model and recognize basic categories, their parts and configuration and hence their function, too! Superordinate and subordinate categories could be derived from basic categories. Recognizing categories opens up new ways of using machine vision. Instead of just assigning labels to objects, like "this object is a cup," questions about the usage (function) of the object could be answered.

The next two sections define the proposed category model for computer vision and describe the recognition procedure. In the discussion the proposed model is compared to other attempts of modeling generic objects and future work is outlined.

Modeling Categories for Computer Vision

Shape representation of generic objects and categories is difficult because the representation must allow for *variations* of shape within a category and yet *differentiate* between categories. The proposed category model consists of combining prototypical part models and specifying possible variations in structure together with deformations of parts to account for the variation inside the category (Figure 2).

Individual superquadrics that model parts are combined to represent articulated objects. This structure will be described by trees, where the internal nodes correspond to Boolean union and the leaves to superquadrics with positional information, or by hierarchical structures, where the position of each part is defined in the coordinate system of a part on the preceding, higher level. Each part model and structural relation has also the information about the allowable shape deformations of parts and allowable changes in structure that account for deviations among the members of the category. The changes of part shapes are interrelated, i.e. a little cup cannot have a big handle although little cups and large handles are themselves legal. The proposed model for categories is hence part based as opposed to feature based because shape deformations operate on parts and changes are made in the way parts fit together. The amount of deformation of the prototype must be limited since not all kinds of deformations can be allowed. Deformations must be specified by the direction and amount of change. The variability of the structural description will be described by specifying a range of possible parameters to allow for relative changes of position among the parts. The variability of superquadrics, on the other hand, will be described by

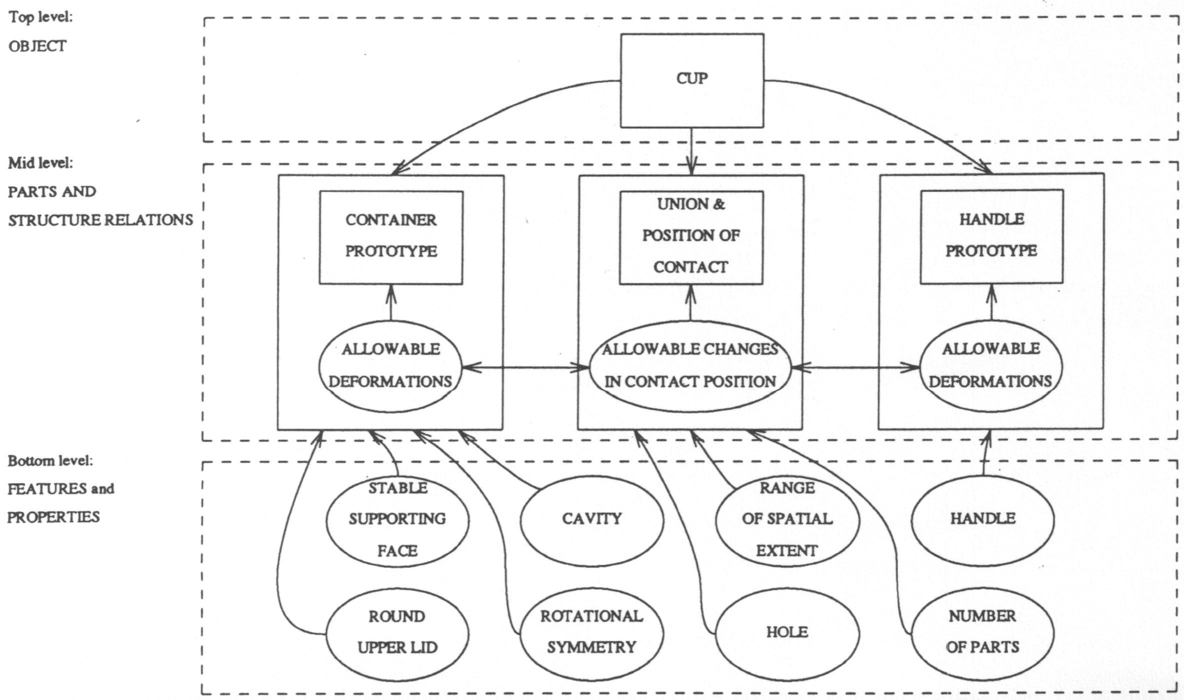

Figure 2 A category model for cups. Category models consist of three levels. The upper level is the category label that applies to all objects in the category. The middle level are the prototypical parts and their structural relation together with allowable changes in structure and allowable shape deformations of part primitives to accommodate variations inside the category. The bottom level specifies the associated features and properties that point to parts or their relation/structure on the second level.

specifying a range of values for superquadric parameters. Deformations must not, however, change any of the features on the bottom level of the category model. Each part and relation has a set of associated features and properties that serve as an interface to the vision modules that recognize features. Each of these features points to one or more parts on the second level or to a structural relationship among parts. All features must be universal enough to be representative for all objects that belong to a particular category.

Superquadrics as part-models. Superquadrics are a family of parametric shapes that extend the basic quadric surfaces and solids.[15] They can model a large variety of "standard" building blocks like spheres, cylinders and prisms (Figure 3). Superquadrics are defined in implicit form or by the following surface column vector:

$$\vec{x}(\eta,\omega) = \begin{bmatrix} a_1 \cos^{\varepsilon_1}\eta \cos^{\varepsilon_2}\omega \\ a_2 \cos^{\varepsilon_1}\eta \sin^{\varepsilon_2}\omega \\ a_3 \sin^{\varepsilon_1}\eta \end{bmatrix} \qquad \begin{array}{c} -\frac{\pi}{2} \le \eta \le \frac{\pi}{2} \\ -\pi \le \omega < \pi \end{array}$$

The parameters η and ω correspond to latitude and longitude angles of the 3-D vector $\vec{x}$ in spherical coordinates. The scale parameters a_1 , a_2 , a_3 define the size of superquadrics in directions x, y and z respectively. ε_1 is the squareness parameter along the z axis and ε_2 is the squareness parameter in the x-y plane. The surface that they describe is everywhere derivable and deformations are easily defined on them.[16]

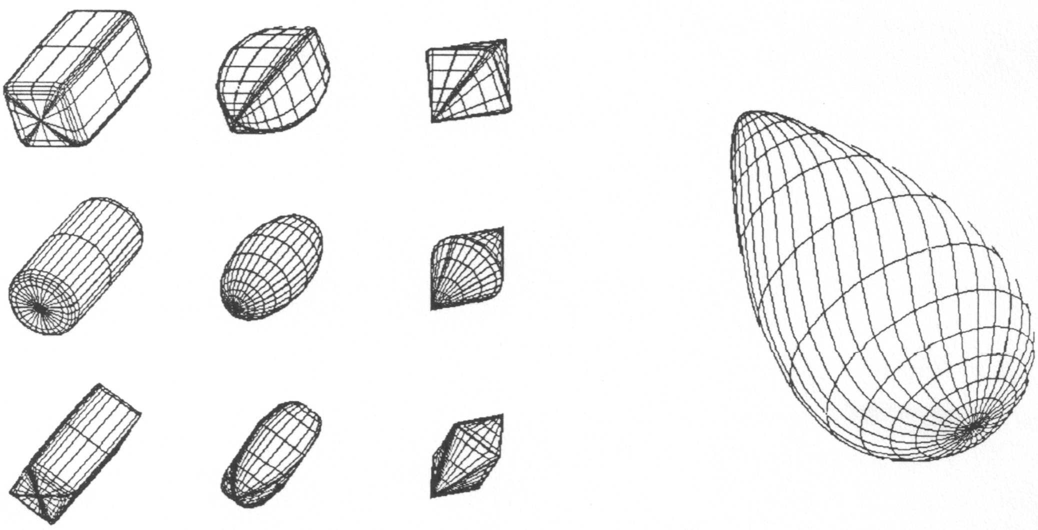

Figure 3 On the left is a set of superquadrics. The control parameters a_1 and a_2 are = ½ a_3. The shape parameters ε_1 and ε_2 change from 0.2 to 1 and 2 from left to the right and from top to the bottom of the figure respectively. On the right is an ellipsoid-like superquadric deformed by tapering and twisting along the *z* axis. Tangent and normal vectors on the undeformed surface can be transformed into tangent or normal vectors on the deformed surface by multiplying with the Jacobian matrix of the deformation function. Deformations are easily combined; each level in the deformation hierarchy requires an additional matrix multiplication.

Recognition Procedure

The input to the recognition procedure is a pair of reflectance images of an object in canonical position, i.e. a cup in its upright position. Low level vision processing includes: thresholding the background, finding edges, connecting edges and growing regions to form closed contours, and using stereo to get 3-D points. To select a model, features typical of all members of a category and yet specialized enough to allow for distinction among categories must be recognized (i.e. holes, cavities, handles, parallel surfaces, parallel edges, rotational symmetries, number of parts, and relative measures like ratio of width/height, or ranges with lower and upper bounds of object dimensions). What can be detected depends on the sophistication of the low level vision and the geometric reasoner that combines the results of low level vision into features.

Hypothesis generation and verification. When a feature is found in an image, all objects that have that feature can be selected and their position and orientation hypothesized.[17] Two techniques are used for comparison, one looks for similarity or common features, the other for differences or distinctive features to set two objects or concepts apart. Similarity seems a good mechanism for initial hypothesis generation while differences seem suited for selecting the best hypothesis among the like. Models that are not consistent with all extracted features are eliminated from the set of possible models. Hypothesis verification has to measure how well the model fits the data. If more then one hypothesis is generated then all of them must be tested. Features alone are not sufficient for verification since different arrangements of same features result in different objects. Verification with shape comparison on the geometric level seems more general. This is against the prevailing trend in object recognition where geometric information is transformed to symbolic representation as soon as possible.[18] We suggest keeping the geometric information around and working with it whenever it is more convenient. Verification in the proposed procedure consists of checking how the parts are put together and deforming each part of the prototype to match the appropriate part of the object. Since a brute force approach for solving the correspondence is not possible, the following parameters must be solved to match the object with the model; the location and orientation of the whole object, subdivision into parts, and the deformation of individual parts. When the model hypothesis is made, based on the position of extracted features, the hypothesis can predict also the location of the local coordinate system. By knowing the relations between the local coordinate systems of individual parts it is possible to check if the parts fit together. Next, the shape of individual parts is verified by matching the prototypical superquadric model to the 2-D contour and 3-D points of the corresponding part. The presence of deformations greatly complicates the problem. By comparing the prototype part with the object part, direction and amount of deformation can be estimated. In this process a predefined sequence of deformations is advantageous. Regarding prototypes as shapes with global symmetry structure Leyton[4] demonstrated strong differential-geometric constraints on the decomposition of prototypification for 2-D shapes. This comparison can give only estimates of the position and orientation of the local coordinate system, superquadric parameters a_1, a_2, a_3, ε_1, and ε_2, and deformation parameters. These estimates are used as the input to a nonlinear modeling procedure (i.e. Levenberg-Marquardt[19]) to compute the actual parameter values. To measure the fit between the deformed superquadrics and the image data the sum of squares of distances between corresponding points is used.

If more than one model matches, either the data is not sufficient to discriminate among the possible categories or the object is a borderline case (since some objects actually match two different categories). The more a prototype must be deformed, the less typical is the object for the category. By comparing two competing models one can conclude whether the information hidden due to self-occlusion can resolve the ambiguity. Changing the viewing direction or manipulating the object (i.e. rotating it) to get more information may help.

Equipment and objects for experiments. Pictures of objects, placed on a rotational platform, will be taken with the remotely controlled Pennsylvania Active Camera System.[20] We use kitchen objects for investigating the relationship between shape and function since their gross shape corresponds to a variety of classes (functions); blob-like (for containment), flat (for support), and elongated (for manipulation). They have holes, handles, cavities, and a clearly defined function or purpose.

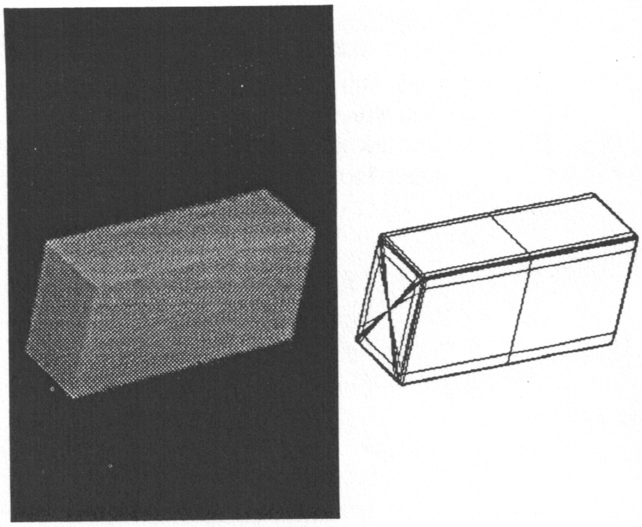

Figure 4 On the assumption that the correct model was selected (superquadric with fixed ε_1 and ε_2), its dimensions (a_1, a_2, and a_3) and the transformation from the global coordinate system to the local coordinate system which is embedded in the superquadric were computed from the 2-D projection of three edges around a common vertex and the assumption that they are perpendicular to each other in 3-D.

Conclusions

We made two basic assumptions for representing the shape of basic object categories; first, that categories can be modeled with a prototype, which has the most representative shape in the category, with a set of shape deformations of that prototype to account for the variations inside the category; and second, that on the basic category level parts (shape) correspond to function. Our model assumes that each object in a category has the same set of parts. Cups, for example, have one handle. Separate categories would have to be made for cups with two or no handles which would join on the next higher level in the category hierarchy, followed by containers, dishes, household objects etc.

Other work concerned with the relation between shape and function is the naive physics program,[14] work by Davis,[21] by Winston et. al[18] on learning physical descriptions of objects from functional descriptions, examples, and precedents and by Stansfield[22] who is concerned with object recognition within an active, multisensor system.

Acknowledgements

This work is supported in part by NSF-CER/DCR82-19196 A02, NSF/DCR-8410771, Air Force/F49620-85-K-0018, and ARMY DAAG-29-84-K-0061.

References

1. Rosch, E., ''Principles of Categorization,'' in *Cognition and categorization*, ed. E. Rosch and B. Lloyd,Erlbaum, Hillsdale, NJ (1978).
2. Wittgenstein, L., *Philosophical Investigations,* Macmillan, New York (1953).
3. Goldmeier, E., *Similarity in Visually Perceived Forms,* International Universities Press, New York (1972).
4. Leyton, M., ''Constraint-Theorems on the Prototypification of Shape,'' *Proceedings AAAI*, pp. 702-706 (August 1986).
5. Tversky, B. and K. Hemenway, ''Objects, Parts, and Categories,'' *Journal of Experimental Psychology: General* **113**(2) pp. 169-193 (June 1984).
6. Thompson, D'Arcy, *On growth and form,* Cambridge University Press, Cambridge, England (1961). Abridged edition, Ed. J. T. Bonner, (First edition 1917)
7. Alexander, C., *Notes on the Synthesis of Form,* Harvard University Press, Cambridge, MA (1964).
8. Witkin, A. P. and J. M. Tenenbaum , ''On the role of structure in vision,'' in *Human and Machine Vision*, ed. J. Beck, B. Hope, A. Rosenfeld ,Academic Press, New York (1983).
9. Hoffman, D. D. and W. A. Richards , ''Parts of recognition,'' *Cognition* **18** pp. 65-96 (1985).
10. Pentland, A. P., ''Perceptual Organization and the Representation of Natural Form,'' *Artificial Intelligence* **28**(3) pp. 293-331 (1986).
11. Biederman, I., ''Human image understanding: Recent research and theory,'' *Computer Vision, Graphics, and Image Processing* **32** pp. 29-73 (1985).
12. Lowe, David G., *Perceptual Organization and Visual Recognition,* Kluwer, Boston (1985).
13. Koenderink, J. and A. van Doorn , ''The shape of objects and the way contours end,'' *Perception* **11** pp. 129-137 (1982).
14. Hobbs, J. R., T. Blenko, B. Croft, G. Hager, H. A. Kautz, P. Kube, and Y. Shoham , ''Common sense Summer: Final Report,'' Report No. CSLI-85-35, Center for the Study of Language and Information, Stanford University (1985).
15. Barr, A. H., ''Superquadrics and angle-preserving transformations,'' *IEEE Computer Graphics and Applications* **1** pp. 11-23 (1981).
16. Barr, A. H., ''Global and local deformations of solid primitives,'' *Computer Graphics* **18**(3) pp. 21-30. (1984).
17. Knoll, T. F. and R. C. Jain , ''Recognizing partially visible objects using feature indexed hypotheses,'' *IEEE J. of Robotics and Automation* **RA-2** pp. 3-13 (1986).
18. Winston, P. H., T. O. Binford, B. Katz, and M. Lowry , ''Learning physical description from functional descriptions, examples, and precedents,'' *Proc. National Conf. on AI*, pp. 433-439 AAAI, (1983).
19. Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vettering, *Numerical Recipies,* Cambridge University Press, Cambridge, England (1986).
20. Krotkov, E. P., J. F. Summers, and F. Fuma , ''The Pennsylvania Active Camera System,'' Technical Report TR-86-15, Department of Computer and Information Science, University of Pennsylvania,, Philadelphia (1986).
21. Davis, E., ''Shape and function of solid objects: some examples,'' Technical report No. 137, Courant Institute, New York (1984).
22. Stansfield, S. A., ''Representation and control within an intelligent, active, multisensor system for object recognition,'' *Proc. 1985 Int. Conf. IEEE Systems, Man, and Cybernetics*, pp. 215-219 (1985).