ADVANCE PRINTING OF PAPER SUMMARIES

SPSE'S 41ST ANNUAL CONFERENCE

May 22-26, 1988
Hyatt Regency Hotel Crystal City

ARLINGTON, VIRGINIA

Sponsored by SPSE
The Society for Imaging Science and Technology

Carnegie Mellon

Department of Computer Science
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213-3890
(412) 268-3016

3 November 1987

Professor Franc Solina
CIS Department/SEAS
University of Pennsylvania
Philadelphia, PA 19104-6389

Dear Franc:

I was asked to organize the session for Robotics and Machine Vision in the SPSE Conference in Washington May 22-27, 1988. Enclosed is a copy of the conference brochure.

As I mentioned, I would like to invite you to give a talk on your work in that session. If you can, please send a short abstract (100-200 words) to the Conference Chairman listed in the enclosed brochure.

Sincerely,

T Kanade

Takeo Kanade
Professor of Computer Science
and Robotics
Acting Director of
the Robotics Institute

Enclosure

ADVANCE PRINTING OF PAPER SUMMARIES

SPSE's 41st ANNUAL CONFERENCE
May 22-26, 1988
Hyatt Regency Hotel
Crystal City, Arlington, Virginia

Published by:
SPSE - THE SOCIETY FOR IMAGING SCIENCE AND TECHNOLOGY
7003 Kilworth Lane
Springfield, Virginia 22151 USA
(703) 642-9090

SPSE - The Society for Imaging Science and Technology
7003 Kilworth Lane
Springfield, Virginia 22151 USA
(703)642-9090

The cover: The version of the United States Capitol Building printed in black is an unmodified digital image. Printed over it in blue ink is the same image modified to enhance the high spatial frequency areas of the scene. Courtesy of LogEtronics Corporation.

ISBN: 0-89209-138-4

This book was reproduced from copy submitted by the authors. No editorial changes have been made to the content.

Color Forming Chemistry and Silver Halide Photography (continued)

Robotics and Machine Vision

Session Chairman: Takeo Kanade, Carnegie Mellon University, Pittsburgh, Pennsylvania

• Invited

RECOVERY OF DEFORMABLE PART MODELS - OR HOW TO MODEL BANANAS AND OTHER ASSORTED FRUITS

Franc Solina and Ruzena Bajcsy
Department of Computer and Information Science
University of Pennsylvania, Philadelphia PA

MODELING OF SHAPE IN COMPUTER VISION

Why are shape models necessary?

Visual perception enables intelligent interaction with the environment. It provides us with information that makes it possible to locate and recognize objects without direct physical contact. Although we are far from a complete understanding of human visual perception, we are trying to endow machines with the sense of vision for the very same reason - to enable machines to interact with a changing environment. The main problem in vision research is that images - which are two-dimensional - underdetermine the three-dimensional world. Different 3-D scenes can produce the same 2-D image. Additional information about image formation and the structure of the world is required to invert the projection and infer meaningful descriptions. In computer vision systems this information is incorporated in models, most of them models for representing shape. The way information in a vision system can be organized and processed is closely linked to the selected representation of shape.

What kind of shape models?

The input for machine vision systems consists of digitized images made out of pixels. These pixels must be organized into larger entities or models. The prevailing method in computer vision is a stepwise introduction of higher and higher level models.[1] From low level shape models such as edges, corners, and surface patches, which are all computed locally, higher level models such as generalized cylinders are constructed. Since at each step of data reduction only the next highest model is considered, errors are common when this approach is used on natural scenes where image formation parameters can change abruptly from image point to image point. For example, shadow edges can be constructed which are not consistent with real physical edges.

There is psychological evidence showing the particular salience of parts.[2] People seem to describe, remember, recognize and categorize objects on the level of part descriptions. Parts and part configuration seem to form a natural bridge connecting perception of objects and behavior toward them, and in turn communication about them. Descriptions are usually in terms of modifiers or deformations of prototypical shapes. Prototypes are those members of categories that seem the most representative.[3] Using deformations is an efficient way of representing a large variety of shapes and has a casual structure because it is related to the process of how the shape evolved over time.[4] These ideas about the structure of the world (perceptual organization) triggered a new wave of research in computer vision.[5] Pentland[6] called for direct recovery of mid-grain level models that correspond to human notion of parts. As models he suggested superquadrics combined with global deformations.

Superquadrics are a family of parametric shapes which are an extension of basic quadric surfaces. Two additional shape parameters ε_1 and ε_2 control the squarness in the x–y plane and along z axis so that solid shapes falling between ellipsoids and parallelepipeds can be defined.[7] Modeling capabilities of superquadrics can be enhanced by deforming them in different ways, including tapering and bending[8] (Figure 1). Superquadrics are parametric models which use a small number of parameters to describe a large number of 3-D points. To find parameters of the model that best fit the data is an overdetermined optimization problem. A result of this

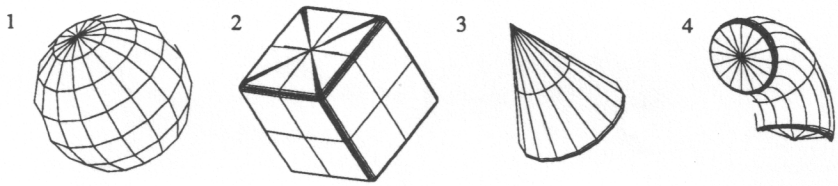

Figure 1: Four examples of shapes produced with superquadric models and global deformations (*1*: $\varepsilon_1 = \varepsilon_2 = 1$, *2*: $\varepsilon_1 = \varepsilon_2 = 0.1$, *3*: tapering, *4*: bending).

optimization is normally also a measure of how well the model fits the data - an important aspect for computer vision where results of interpretation are normally not directly verifiable. The shape recovery of parametric models can be explained also in terms of intrinsic and extrinsic forces. Intrinsic forces are the internal properties of the model, governing its possible arrangements and its possible shape. Extrinsic forces are the influences which direct the shape options allowed by the internal constraints. Shape is a result of the interaction of intrinsic and extrinsic forces.

MODEL RECOVERY

Recovery of single-part models

We introduced a new method for recovery of superquadrics with parametric deformations (bending, tapering, and cavity deformation).[9] The input are range images were obtained with a passive laser imager.[10] Let us first assume one single-part object in the scene at a time, with the supporting surface removed. Model recovery is formulated as a least-squares minimization of a cost function causing the set of input range points to lie on or close to the surface of the model. The cost function is based on the inside-outside function

$$F(x, y, z) = \left[\left[\left[\frac{x}{a_1} \right]^{\frac{2}{\varepsilon_2}} + \left[\frac{y}{a_2} \right]^{\frac{2}{\varepsilon_2}} \right]^{\frac{\varepsilon_2}{\varepsilon_1}} + \left[\frac{z}{a_3} \right]^{\frac{2}{\varepsilon_1}} \right]^{\varepsilon_1}$$

which for a given 3-D point (x, y, z) tells where the point lies relative to the model's surface. Parameters a_1, a_2, a_3 set the size of the superquadric in x, y and z direction of the model's internal coordinate system. The actual cost function has six additional parameters to express any position and orientation of models in space and includes also a size factor in order to recover the smallest possible model that fits the input points. Each shape deformation requires additional parameters in the fitting function. During the iterative gradient descent minimization process, all model parameters are adjusted simultaneously recovering the position, orientation, size and shape of the model (Figure 2). Initial estimates required for minimization are rough position, orientation and size of the object, acquired by computing the center of gravity of input range points, central moments of inertia, and the extend of range points along the axis of least inertia. The solution space can be searched efficiently with a gradient descent method when combined with a stochastic technique that prevents the procedure to get trapped in a shallow local minimum.

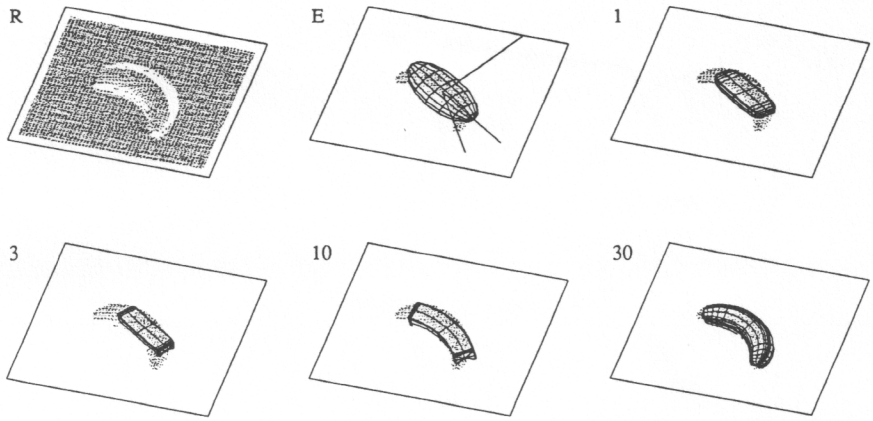

Figure 2: Shape recovery of a bent object - a banana. Top left (*R*) is the original range image. *E* is the initial estimate superimposed on the object range points (the supporting surface was removed). *1, 3, 10* and *30* are the iterations during the model recovery sequence when a total of 13 model parameters were adjusted simultaneously.

Segmentation

When an object consists of several parts, a variable number of range points in a model allows the model to actively search for a better fit (by compressing itself and expanding) resulting in a subdivision of the object into a model representing the largest part of the object and points belonging to the rest of the scene.[9] Using the same method, the remaining points can be recursively subdivided into parts each represented with a single compact volumetric model (Figure 3).

CONCLUSIONS

So far we used range data as input for model recovery. Other shape cues could be used, even in combination, as long as they can be interpreted as position (passive stereo, focusing) or orientation (shading, texture, occluding contours). The proposed method for segmentation and shape representation recovers gross shape features before shape details. This asymmetric order is in accord with how people perform planning, execute drawing or identify objects.[4] One has to find out where an object is, how it is located, how big it is, before its general outline and shape details can be recovered. Different vision tasks differ by how far they have to go up this sequence from general to specific. For path planning, just the location of objects suffice. For object avoidance or grasping, orientation and size become important too. Only for recognition, recovery of gross or fine shape features becomes necessary. For handling of mail pieces, for example, a simple classification into four groups, based on the recovered superquadric model parameters is possible.[11]

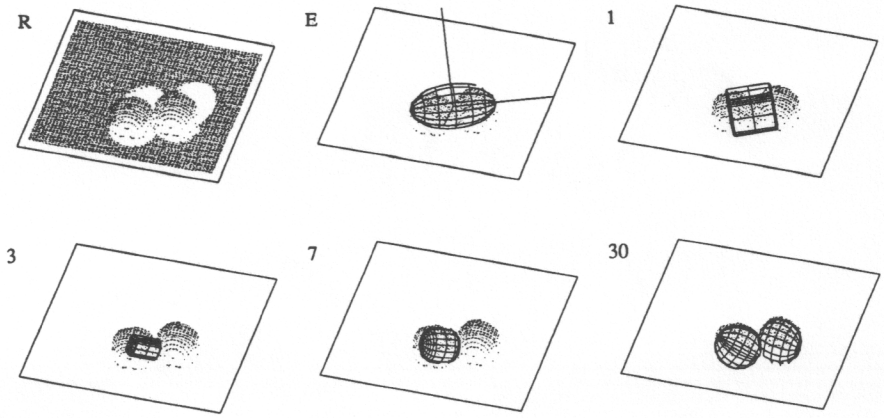

Figure 3: Segmentation and shape recovery of two oranges. Top left (*R*) is the original range image. *E* is the initial model superimposed on input object range points after the supporting surface was removed. *1*, *3*, *7* and *30* are the iterations during the model recovery sequence. The sequence shows the initial model (*E*) shrink (*1-3*) and then expand (*7-30*) to recover the shape of that part. *30* shows also the model of the second object, recovered from the remaining points.

REFERENCES

1. D. Marr, 1982. *Vision.* San Francisco: Freeman.
2. B. Tversky and K. Hemenway, 1984. "Objects, Parts, and Categories." *Journal of Experimental Psychology: General* **113** (2): 169-193.
3. E. Rosch, 1978. "Principles of Categorization." In *Cognition and categorization*, ed. E. Rosch and B. Lloyd. Hillsdale, NJ: Erlbaum.
4. M. Leyton, 1987. "Nested structures of control: an intuitive view." *Computer Vision, Graphics, and Image Processing* **37**: 20-53.
5. A. P. Witkin and J. M. Tenenbaum, 1983. "On the role of structure in vision." In *Human and Machine Vision*, ed. J. Beck, B. Hope, A. Rosenfeld. New York: Academic Press.
6. A. P. Pentland, 1986. "Perceptual Organization and the Representation of Natural Form." *Artificial Intelligence* **28** (3): 293-331.
7. A. H. Barr, 1981. "Superquadrics and angle-preserving transformations." *IEEE Computer Graphics and Applications* **1**: 11-23.
8. A. H. Barr, 1984. "Global and local deformations of solid primitives." *Computer Graphics* **18** (3): 21-30.
9. F. Solina, 1987. "Shape recovery and segmentation with deformable part models." Ph.D. dissertation, University of Pennsylvania, Philadelphia.
10. G. Tsikos, 1987. "Segmentation of 3-D scenes using multi-modal interaction between machine vision and programmable mechanical scene manipulation." Ph.D. dissertation, University of Pennsylvania, Philadelphia.
11. F. Solina and R. Bajcsy, 1987. "Range Image Interpretation of Mail Pieces with Superquadrics." In *Proceedings AAAI-87*, 733-737. Seattle.