Range Image Interpretation of Mail Pieces with Superquadrics

Franc Solina and **Ruzena Bajcsy**

GRASP Laboratory
Department of Computer and Information Sciences
University of Pennsylvania
Philadelphia, Pennsylvania 19104-6389, USA

Abstract

Although mail pieces can be classified by shape into parallelopipeds and cylinders, they do not conform exactly to these perfect geometrical shapes due to rounded edges, distorted corners, and bulging sides. Segmentation and classification of mail pieces hence cannot rely on a limited set of specific models. Variations and deformations of shape can be conveniently expressed when using superquadrics. We show how to recover superquadric models for mail pieces and segment the range image at the same time.

I. Introduction

Postal services are currently facing the problem of automating mail piece handling. At present only letter handling is fully automated. The rest of the mail pieces is handled at least partially if not completely by hand due to their large variability in size and shape [Owen, 1986]. Any automatic system for handling mail pieces has to determine location, orientation, size, and shape of mail pieces in order to manipulate them accordingly. Computer vision is a promising way to satisfy these requirements.

The problem of characterizing mail pieces is somewhere between scene description and object recognition. For scene description, a unique description of objects is not necessary. It is generally sufficient to generate, using a bottom up strategy, a succession of representations that depend on the viewing direction and orientation of objects which results in a geometric representation such as surface patches or polyhedral approximations. On the other hand, to recognize an object in the scene as one from a set of predefined models, a computer vision system must have models of these objects which it compares to the input data. For recognition of 3-D objects, view point independent, 3-D models are required. Most working recognition systems rely on fixed, definitive models intended only for environments where a limited, preselected number of objects are encountered. Mail pieces, however, do not come just in a few uniform shapes and sizes. Thus, having individual models for each mail piece is not feasible. This is why segmenting and representing mail pieces is not object recognition in the strict sense which is normally understood as selecting the right ready-made model from a predefined set of models.

Classifying mail pieces is related to categorization. People form categories by picking out the essential and separating it from the accidental [Rosch, 1978]. This sorting of instances into categories reflects the structure of the world [Pentland, 1986a], [Bajcsy and Solina, 1987]. Like any other objects, mail pieces can be grouped into classes or categories. Shape classification which is used for manual handling of mail pieces and which identifies parcels, flats, tubes, rolls, and irregular packages, reflects such structure. An automated mail handling system must also divide mail pieces into appropriate classes, give their shape description by identifying the necessary parameters of the class model, and provide the position and orientation in a world coordinate system. The difficulty in modeling mail pieces is their nonuniform shape and size. They do not conform to perfect geometrical shapes because of rounded edges, distorted corners, bulging sides, and wrinkled wrapping. With standard 3-D shape representations, like generalized cylinders or polyhedral approximations, such degradations from ideal prototypes are difficult to express. Superquadrics, on the other hand, have the advantages of generalized cylinders and direct control over the roundness/squareness of edges. In general, only a single superquadric model is required for a single mail piece.

The rest of the paper is organized as follows: we first describe the recovery of superquadric models from range data, outline the recognition procedure, including some new ideas and preliminary results about segmentation and, at the end, compare our recovery method with other related work and discuss future research.

II. Model recovery

Superquadrics are a family of parametric shapes that were invented by the Danish designer Peit Hein [Gardiner, 1965] as an extension of basic quadric surfaces and solids (see also [Barr, 1981]). Pentland [Pentland, 1986a] suggested them first for analysis of scenes in computer vision. A superquadric surface is defined by the following implicit equation:

$$\left[\left(\frac{x_S}{a_1}\right)^{\frac{2}{\varepsilon_2}}+\left(\frac{y_S}{a_2}\right)^{\frac{2}{\varepsilon_2}}\right]^{\frac{\varepsilon_2}{\varepsilon_1}}+\left(\frac{z_S}{a_3}\right)^{\frac{2}{\varepsilon_1}}=1 \qquad (1)$$

x_S, y_S, and z_S are coordinates of a point on the superquadric surface. Subscript S indicates a superquadric centered

This research was made possible by the following grants and contracts: USPS 104230-87-H-0001/M-0195, ONR Subcontract SB35923-0, NSF/DCR-8410771, ARMY/DAAG-29-84-K-0061, NSF-CER/DCR82-19196 A02, Airforce/F49620-85-K-0018, and DARPA/ONR. We wish to thank Sandy Pentland for his continuous encouragement to use superquadric models and Max Mintz for helping with the minimization procedure.

coordinate system. Parameters a_1, a_2, a_3 define the superquadric size in x, y and z directions, respectively. ε_1 is the squareness parameter along the z axis and ε_2 is the squareness parameter in the x-y plane. By changing the two shape parameters, superquadrics can model a large set of standard building blocks, like ellipsoids, cylinders, parallelopipeds, and all shapes in between. Global deformations like tapering, twisting, and bending further enhance superquadric modeling capabilities [Barr, 1984].

We define the "inside-outside" function for superquadrics as:

$$F(x_S, y_S, z_S) = \left[\left[\left(\frac{x_S}{a_1} \right)^{\frac{2}{\varepsilon_2}} + \left(\frac{y_S}{a_2} \right)^{\frac{2}{\varepsilon_2}} \right]^{\frac{\varepsilon_2}{\varepsilon_1}} + \left(\frac{z_S}{a_3} \right)^{\frac{2}{\varepsilon_1}} \right]^{\varepsilon_1} \tag{2}$$

When $F(x_S, y_S, z_S) = 1$, the point (x_S, y_S, z_S) is on the surface of a superquadric. If $F(x_S, y_S, z_S) > 1$, the corresponding point lies outside and if $F(x_S, y_S, z_S) < 1$, inside the superquadric. With the outermost exponent ε_1 we force F to grow quadratically instead of exponentially. This ensures faster convergence during model recovery.

Superquadrics are suitable models for computer vision because we can form *overconstrained* estimates of their parameters. This overconstraint comes from using models defined by a few parameters to describe a large number of 3-D points. This enables us to verify our estimated models and measure the "goodness of fit." For a superquadric in an arbitrary position we must recover 11 parameters: location in space (3 par.), orientation in space (3 par.), size (3 par.), and two shape parameters, ε_1 and ε_2. On the other hand, many more 3-D points are typically available on the surface of the modeled object from either range imaging or passive stereo. To find the parameters so that the model best fits the data is called an overdetermined optimization problem.

We introduce here a relatively fast *iterative* fitting procedure based on the "inside-outside" function. Eq. (2) defines the surface in a superquadric centered coordinate system (x_S, y_S, z_S). 3-D points from passive stereo or range imaging, however, are given in a world coordinate system (x_W, y_W, z_W). We express these 3-D points in the superquadric centered coordinate system by a translation and a sequence of rotations. A convenient way of expressing such transformation in homogeneous coordinates is with a 4×4 matrix T:

$$\begin{bmatrix} x_S \\ y_S \\ z_S \\ 1 \end{bmatrix} = T \begin{bmatrix} x_W \\ y_W \\ z_W \\ 1 \end{bmatrix} \tag{3}$$

where $T = Trans(p_1, p_2, p_3) \cdot Rot(\phi, \theta, \psi)$. We use Euler angles to express the orientation in terms of rotation ϕ about the z axis, followed by a rotation θ about the new y axis, and finally, a rotation ψ about the new z axis. Substituting eq. (3) into eq. (2) we get the "inside-outside" function for a superquadric in general position and orientation:

$$F(x_W, y_W, z_W) = F(x_W, y_W, z_W, a_1, a_2, a_3, \varepsilon_1, \varepsilon_2, \phi, \theta, \psi, p_1, p_2, p_3) \tag{4}$$

The independent parameters expressed in vector notation are: $\vec{a} = [a_1, a_2, \cdots, a_{11}]^T$. Suppose we have N 3-D surface points (x_W, y_W, z_W) which we want to model with a superquadric. Eq. (4) predicts the position of a point (x_W, y_W, z_W) relative to the surface of the model. We want to vary the 11 adjustable parameters a_j, j = 1, ... , 11 in eq. (4) to get such values for a_j's that most of the 3-D points will lay on or close to the model's surface. Since for points on the surface of a superquadric: $F(x_W, y_W, z_W; a_1, \cdots, a_{11}) = 1$, we achieve this by minimizing:

$$\sum_{i=0}^{N} [1 - F(x_{W_i}, y_{W_i}, z_{W_i}; a_1, \cdots, a_{11})]^2 \tag{5}$$

However, due to self-occlusion the solution to eq.(5) is unbounded in the sense that an infinite number of superquadric models of different size fit objects like cylinders or parallelopipeds. Obviously only the model with the *smallest* possible volume that still fits the given points is the desired solution. We want a modified fitting function which has a minimum corresponding to the smallest superquadric that fits a set of 3-D points *and* such that the function value for surface points is known before the minimization. Using function:

$$R = a_1 a_2 a_3 (F - 1) \tag{6}$$

we fulfill the first requirement with the factor $a_1 a_2 a_3$, which corresponds to the superquadric size. The second requirement is met by the factor $(F - 1)$, since function R has value 0 for all points on the surface and does not depend on knowing the correct size. Now we have to minimize:

$$\sum_{i=0}^{N} [R(x_{W_1}, y_{W_1}, z_{W_1}; a_1, \cdots, a_{11})]^2 \tag{7}$$

Since R is a nonlinear function of 11 parameters a_j, j = 1, ... , 11, the minimization must proceed iteratively. Given trial values for $\vec{a}$, we evaluate eq. (6) and employ a procedure to improve the trial solution. The procedure is then repeated with new trial values until the sum of least squares (eq. 7) stops decreasing, or the changes are statistically meaningless. Since first derivatives $\partial R / \partial a_i$ for $i = 1, \cdots, 11$ can be computed, we use the Levenberg-Marquardt method for nonlinear least squares [Press *et al.*, 1986]. The first trial set of parameters, $\vec{a}$, must be set experimentally to some initial estimates $\vec{a}_i$. We found out that very rough estimates for position, size and orientation are sufficient. Initial estimates for both shape parameters, ε_1 and ε_2 can always be 1, while position, orientation, and size can be estimated by computing the center of gravity and moments of inertia for the given 3-D points. During the fitting procedure we introduce "jitter" by adding Poisson distributed noise to the evaluation of function R. Small local minima caused by the complicated topology of the fitting function and the noise in the input data are thus avoided and a global convergence assured [Pentland, 1986b].

Deformed superquadrics can be recovered using the same technique of minimizing the "inside-outside" function (Fig. 1). Global deformations like tapering, bending, and twisting require

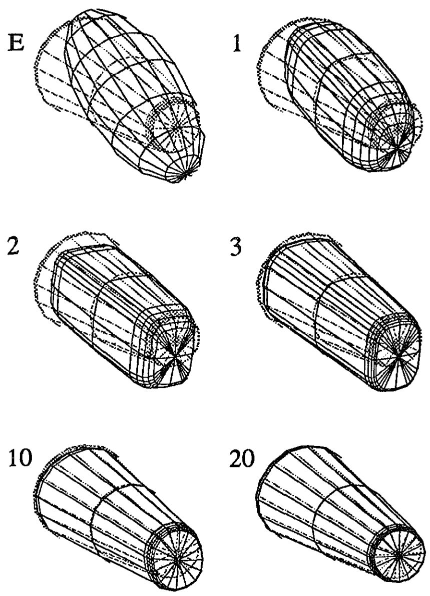

Figure 1: Recovery of a tapered cylinder with an iterative process through which the estimated shape converges to the actual range data. The initial estimate and some of the following iterations (solid lines) are shown superimposed on the superquadric (broken lines) that represents the input data (300 3-D points). A total of 13 model parameters (11 + 2 for tapering) were adjusted *simultaneously* to achieve a least squares fit. The whole fitting procedure took about three minutes on a VAX 785.

just a few additional parameters. *Any* shape deformation can be recovered in this way as long as the inverse transformation is available [Bajcsy and Solina, 1987].

We tested the fitting procedure on synthetic (Fig. 1), and real range data (Fig. 2). The described recovery procedure is *fast* and *stable* in the sense that it always converges to a good approximation of the actual object. We are able to fit *simultaneously* all 11 parameters and achieve a good fit in just a few iterations (Fig. 3). Speed depends on the number of 3-D points for which the fitting function and their derivatives must be evaluated, the number of necessary parameters and the accuracy of initial parameter estimates. We investigated the robustness of the minimization procedure by studying the relation between independent parameters of the fitting function and the sum of least squares (Fig. 4).

III. Recognition procedure

The goal of a vision system for mail piece handling is to classify each mail piece into a class of like objects and report its position, orientation and size so that appropriate manipulation can be performed. The whole process can be divided into image acquisition, segmentation, model recovery and classification. Model recovery was already described. The rest of this section is devoted to segmentation and classification.

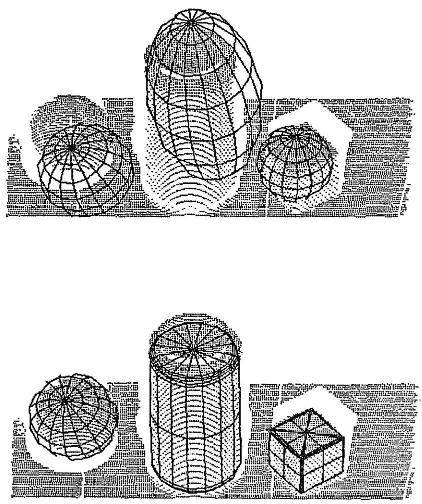

Figure 2: Interpretation of a real range image [Hansen and Henderson, 1986] with superquadric models. On top are the initial model estimates, on the bottom the recovered models after 12th iteration. Segmentation into individual objects was done by hand.

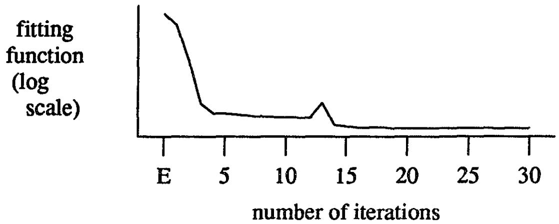

Figure 3: Rate of convergence for the cylinder in Fig. 3. The notch around 13th iteration is due to the addition of Poisson distributed noise which pushed the fitting process out of a local minimum and towards a better solution. One iteration using about 500 range points took about 15 seconds on a VAX 785.

A. Scene segmentation

Under the assumption that only single mail pieces are present in the scene, segmentation consists of removing the supporting surface. The remaining range points are then used for model recovery. If several, possibly overlapping, mail pieces are present, segmentation must divide the scene into regions corresponding to single objects.

Segmentation is a data driven process and normally applies image formation models like edges, corners, regions, normals, and surfaces to the image. A review of low level range image processing research [Besl and Jain, 1985] reveals that there are two principal approaches. One extracts edges, the other segments surfaces into planar or cylindrical surfaces. The "edges first" approach is successful when the objects have nice, clear edges. Mail pieces, however, have crumpled edges and beaten corners and this shape noise degrades the performance of edge finders. Crumpled paper on mail pieces can also mislead a region growing algorithm, causing it to subdivide a single face into a number of small surface patches. Using extracted features, regions corresponding to a single object or part can be hypothesised and verified by model fitting.

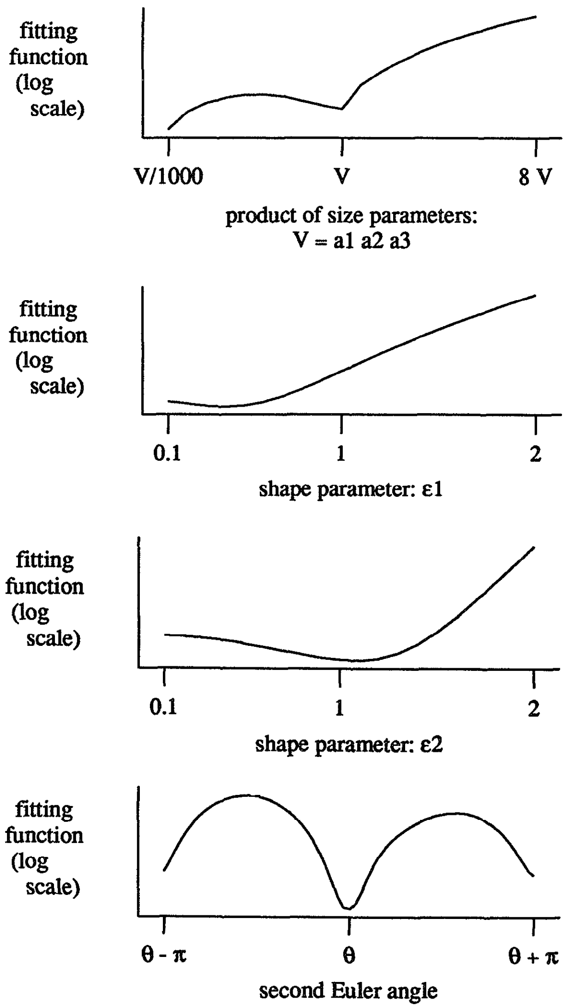

Figure 4: Influence of inside-outside function parameters on the fitting function for the cylinder in Fig. 2. Although all parameters are interdependent, these 2-D plots give some insight into the behavior of the inside-outside function. Note that factor $a_1 a_2 a_3$ in modified function R (eq. 6) introduced a new minimum when any of a's is 0. If the initial values for size are not much to underestimated, this does not cause a problem.

We are currently investigating the use of superquadric models for segmentation also. The recovery procedure described in the previous section uses a fixed number of range points which are assumed to belong to the same mail piece. Now consider the case where only very gross segmentation is available, a result of simple histogramming for example, or even no segmentation at all. We want to take the whole scene as a large block and like a sculptor carve out the objects or parts that make up the scene. The shape of possible recovered parts depends on the capabilities of our models. Superquadric primitives combined with some global deformations can describe a large class of man-made and natural objects [Pentland, 1986a]. The problem can be interpreted as a global minimization problem over the space of model parameters *and* number of models. First, we want to recover the model that accounts for the largest number of data points and repeat the process for remaining chunks until an appropriate level of representation for the task at hand is reached. The number of points during model recovery is not fixed. Points that are to far outside from the model's surface in the current iteration do not contribute to the estimation of model parameters, while other points, not used in a previous iteration but close enough in the present iteration, are used again. The changing number of points from iteration to iteration must be taken into account when comparing the goodness of fit.[1]

B. Mail piece classification

Classification of mail pieces is necessary because differently shaped mail pieces require different handling. A classification scheme must reflect the shape of mail pieces but can also depend on the nature of the automated manipulation (robot arms equipped with grippers or suction pumps, fixed automation). Using recovered superquadric parameters, different geometric classification schemes can be easily designed. For example, the classification currently used for manual handling is:

- letters and flats ($a_1 \ll a_2$, a_3 and ε_1, $\varepsilon_2 \ll 1$),
- box-like packages (ε_1, $\varepsilon_2 \ll 1$),
- tubes and rolls ($\varepsilon_1 \ll 1$ and $\varepsilon_2 = 1$),
- irregular objects ($1 < \varepsilon_{1,2} > 2$, global deformations).

IV. Discussion

Pentland has shown that superquadric primitives can describe a large class of man-made and natural objects [Pentland, 1986a]. We believe that they are appropriate as part-based models, especially for the class of basic categories, since the *prototype and deformation* paradigm common in human perception can easily be applied [Bajcsy and Solina, 1987]. At that level very detailed shape descriptions are not necessary. With a small set of parameters a large set of primitives can be *uniformly* handled. Superquadrics model the whole object, including parts hidden by self-occlusion and parts occluded by other objects, by assuming symmetry. Verification which normally comes as an afterthought is here an integral part of model recovery.

[1] Instead of comparing the sum of least squares, we divide the sum first by the number of participating points. The treshold for rejecting points that are too far *outside* from the model's surface is a function of goodness of fit. The better the fit, the stricter the rejection criteria.

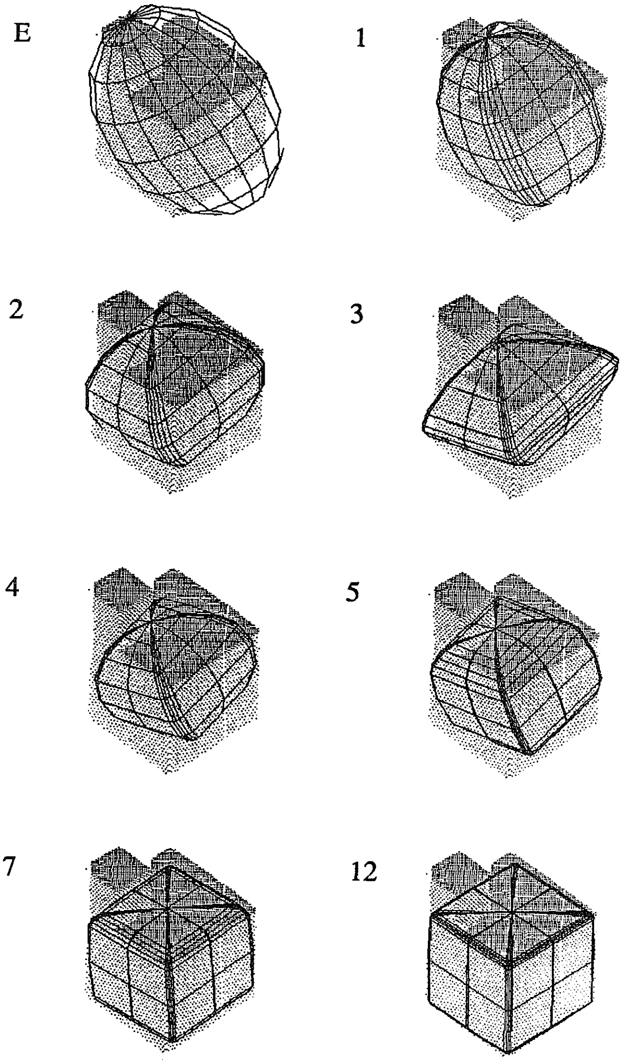

Figure 5: Segmentation by model recovery. The above image sequence shows the iterative process through which the estimated shape based on the non-segmented range image converges to a model that accounts for the largest part in the scene. The small cube on top of the large one can be recovered simply by applying the fitting to the remaining range points.

To recover superquadric models, Pentland [Pentland, 1986a] first suggested an analytic solution of parametric superquadric equations. Using linear regression, one could compute parameter values that provide the best fit. Pentland [Pentland, 1986b] currently recovers superquadrics from range data by computing a heuristic "goodness-of-fit" functional in a coarse grain search over the *entire* parameter space. We believe that, due to complexity, an analytic solution for superquadric parameters is not practical. A heuristic approach, on the other hand, lacks precision, and global search is computationally expensive. Recovery using the "inside-outside" function and a steepest descent method combined with addition of Poisson noise has proved to be more efficient. The speed of the fitting procedure depends on the number of range points, the number of function parameters and the accuracy of first parameter estimates. Since the "inside-outside" function and its partial derivatives can be evaluated for all range points in parallel, the fitting procedure may be speeded up on a parallel architecture.

Segmentation by model recovery looks promising but more research is in order. Global search is a possible but costly proposal [Pentland, 1986b]. We will investigate if the method would benefit by using simulated annealing.

References

[Bajcsy and Solina, 1987]
R. Bajcsy and F. Solina, "Three Dimensional Shape Representation Revisited," Proceedings ICCV, London, England (June 1987).

[Barr, 1981]
A. H. Barr, "Superquadrics and angle-preserving transformations," *IEEE Computer Graphics and Applications* 1 pp. 11-23 (1981).

[Barr, 1984]
A. H. Barr, "Global and local deformations of solid primitives," *Computer Graphics* 18(3) pp. 21-30. (1984).

[Besl and Jain, 1985]
P. Besl and R. Jain, "Range Image Understanding," *IEEE Proceedings on Computer Vision and Pattern Recognition*, (June 1985).

[Gardiner, 1965]
M. Gardiner, "The superellipse: a curve that lies between the ellipse and the rectangle," *Scientific American*, (September 1965).

[Hansen and Henderson, 1986]
C. Hansen and T. Henderson, "UTAH Range Database," Technical Report UUCS-86-113, Computer Science Department, University of Utah, Salt Lake City (April 1986).

[Owen, 1986]
J. Owen, "Characterization of Live Mail," Proceedings USPS Advanced Technology Conference, Washington, DC pp. 2-22 (1986).

[Pentland, 1986a]
A. P. Pentland, "Perceptual Organization and the Representation of Natural Form," *Artificial Intelligence* 28(3) pp. 293-331 (1986).

[Pentland, 1986b]
A. Pentland, "Recognition by parts," SRI Technical note No. 406, Menlo Park, CA (1986).

[Press *et al.*, 1986]
W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vettering, *Numerical Recipies,* Cambridge University Press, Cambridge, England (1986).

[Rosch, 1978]
E. Rosch, "Principles of Categorization," in *Cognition and categorization*, ed. E. Rosch and B. Lloyd, Erlbaum, Hillsdale, NJ (1978).