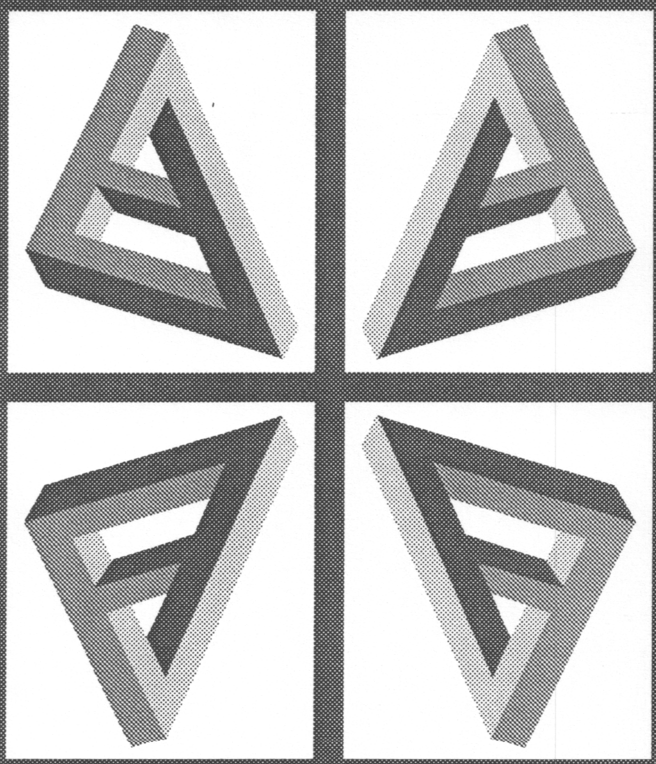
ASPECTS OF VISUAL FORM PROCESSING



Edited by
Carlo Arcelli
Luigi P. Cordella
Gabriella Sanniti di Baja

World Scientific

ASPECTS OF VISUAL FORM PROCESSING

Edited by

Carlo Arcelli
Istituto di Cibernetica
Consiglio Nazionale delle Ricerche
Naples, Italy

Luigi P. Cordella
Dipartimento di Informatica e Sistemistica
Università di Napoli Federico II
Naples, Italy

Gabriella Sanniti di Baja
Istituto di Cibernetica
Consiglio Nazionale delle Ricerche
Naples, Italy

World Scientific
Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 9128

USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661

UK office: 73 Lynton Mead, Totteridge, London N20 8DH

ASPECTS OF VISUAL FORM PROCESSING
2nd International Workshop on Visual Form

ISBN 981-02-2011-1

Printed in Singapore by Uto-Print

SHAPE DECOMPOSITION USING PART-MODELS OF DIFFERENT GRANULARITY

FRANC SOLINA and ALEŠ LEONARDIS

Computer Vision Laboratory, University of Ljubljana
Faculty of Electrical Engineering and Computer Science,
Tržaška 25, 61001 Ljubljana, Slovenia

ABSTRACT

In this paper we study the possibilities of recovering object structure from range images using part-level models of variable granularity. The approach is based on a new method of reliable and efficient recovery of part-descriptions in terms of superquadric models. Instead of pre-segmenting range images and then recovering volumetric part models from isolated regions, a direct and *simultaneous* recovery of part models from the whole unsegmented range image is achieved. This is possible by integrating a *superquadric* model fitting technique into the *recover-and-select* paradigm. By changing the compatibility constraints in this segmentation/part recovery scheme, models of different granularity can be recovered. In this way, the system can be adapted to different tasks which require different levels of description.

1. Introduction

The significance of detecting geometric structures in images has long been realized in the vision community. One of the primary intentions has been to build primitives that would bridge the gap between low-level features and high-level symbolic structures useful for further processing. To represent the "natural" structuring of the world and support recognition and learning of such "natural" structures from images people employ a part structure. Perceptually, the world can be broken down into parts, and the goal of computer vision is to recover from images this part structure (segmentation) and the metric properties of individual parts (shape recovery). This structure normally exhibits a hierarchy of parts. Larger parts are made up of smaller ones which are again made up of even smaller ones. There can be several such levels of details that are visible even to the unaided human eye. This part hierarchy is also reflected in the language since parts at different hierarchical levels have distinct names. Marr already envisioned such a part hierarchy for description (i.e., see the human body constructed of generalized cylinders in Fig. 1 after [17]).

For efficient communication people normally use words that refer to the highest level of this part hierarchy which makes sense in a given situation. Computer vision systems must therefore also provide descriptions at the appropriate level of details. The purpose of machine vision is not to reconstruct the scene in its entirety, but rather to search for specific features that enter, via data aggregation (i.e., model recovery), into a symbolic description of the scene at a level necessary to achieve a specific task.

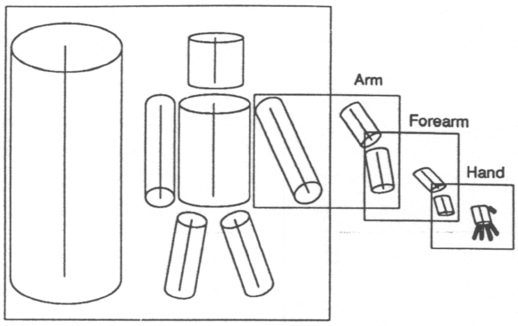


Figure 1: A hierarchy of different levels of part descriptions. Each model is a self-contained unit of shape information and has a limited complexity. (After [17].)

Many theories have emerged which emphasize the importance of extracting perceptually relevant image structures [17,5,18] indicating that a model-free interpretation is doomed to fail due to the underconstrained nature of the problem. Relevant models essentially encode the knowledge or expectations of how the data is structured and help augmenting imperfect visual data with intrinsic information, thus making the recovery process more robust.

Two types of volumetric models, generalized cylinders [6] and superquadrics [18] have emerged for such part-level modeling. Lately, superquadric models are gaining popularity in the vision [2,7,8,10,11,16,19,20,21,22,23,25] and robotics community (i.e., grasping [1], modeling kinematic chains [26]) because of their compact representation and robust recovery methods of individual models.

Although rigorous schemes for recovery of volumetric models have been developed, most of them make the assumption that the segmentation problem has been solved by some other means [12,22,21,25]. By trying to avoid, explicitly or implicitly, the phase of classifying and grouping together the image features that form individual volumetric models, the real complexity of the task is obscured. This is due to the difficulty of simultaneous classification (grouping) of image elements and model parameter estimation which has been a major obstacle to successful applications that require reliable extraction of volumetric models from the data. Only recently a method for direct and simultaneous recovery of superquadric models from unsegmented range images has been proposed [16]. Using this part segmentation/recovery method we study in this paper how a hierarchy of part representations can be constructed, much like the part hierarchy proposed by Marr [17].

The paper is organized as follows: in section 2 we discuss the related work. In section 3 we give a brief outline of part model segmentation using *superquadric recovery* in the *recover-and-select* paradigm. In section 4 we describe how different levels of part-hierarchy can be recovered in this segmentation/recovery approach and show some preliminary results. We conclude the paper with a summary and outline the work in progress.

2. Background and Related Work

Most approaches to segmentation in computer vision try to find part boundaries using local image information, such as edges, surface patches and surface normals. Such segmentation methods are inherently sensitive, they are susceptible to noise and details not relevant for the targeted level of representation. Sometimes, the local information for part boundary is entirely absent in the data. The segmentation using such local features is therefore often arbitrary. Essentially, local information cannot decide on the shape of the whole part if the concept of the whole part is not well defined as such. The problem of using part boundaries to define the shape of parts can be circumvented by defining the whole part shape directly. Biederman [5] proposed a set of 35 volumetric primitives, called *geons,* which he obtained by analyzing non-accidental changes on a generalized cylinder. These models could be used like phonemes in a language for describing all possible shapes.

Generalized cylinders are difficult to recover from images, at least in their general form. To make the recovery tractable, Pentland [18] introduced superquadric models which have an explicit and implicit equation of their surfaces. Pentland [19] proposed to recover superquadric models through a coarse search of the entire superquadric parameter space that combined model recovery with segmentation. This method, however, is computationally very expensive.

A much faster method for recovery of deformed superquadric models from range data, formulated as a least-squares minimization of a fitting function, was proposed by Solina and Bajcsy [2,22]. They approached also the problem of segmentation by describing the entire object with a single superquadric and then recursively splitting the models until an adequate description was achieved [2]. The implementation of this segmentation method, however, was not stable. A similar recursive formulation of segmentation using superquadric models was recently proposed by Horikoshi and Suzuki [10].

Pentland [20] later proposed a different segmentation method based on matching 2-D silhouettes (projections of 3-D superquadrics of different shapes, orientations, and scales) to the image. After such segmentation, 3-D superquadric models are fitted to range points of individual part regions. In a related work on recovery of superquadric-like physically based models Pentland pre-segmented range images using simple polynomial shape models [21].

This research on superquadrics was followed by other works using and extending the above superquadric recovery methods (Ferrie *et al.* [7], Gupta and Bajcsy [8]). However, segmentation in those systems still relies on surface [8] and/or contour fitting [7]. Since data points are grouped together on the basis of some other model types and not on the basis of the final superquadric models, this essentially decouples the classification and representation phases, so that the resulting segments do not always correspond well to the final superquadric primitives. When higher-level models are well defined, as is the case with superquadrics, one can attempt to find them directly and simultaneously as shown in [16,23] which successfully employs the representation to guide detection and grouping processes [4,3].

All of the above approaches are geared to recover the part structure only at a single description level. As we argued in the introduction a whole hierarchy of part descriptions is necessary to adequately represent the world for intelligent interaction (scale of representation with respect to the manipulating tools). In this paper we try to extend our previous work on part-level segmentation to incorporate this important feature.

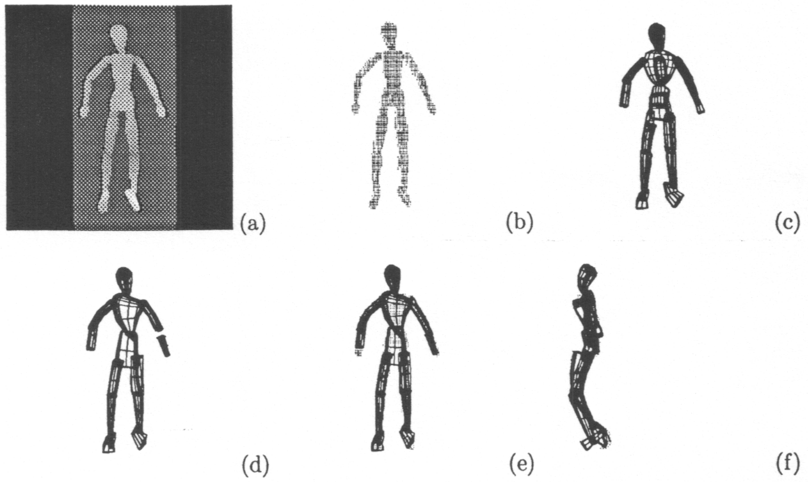


Figure 2: Part-level segmentation of a doll.

3. Part-Level Segmentation

To make the paper self-contained we give in this section a brief summary of the segmentation/recovery method which is described in [16,23]. The part-level segmentation method is a successful combination of a superquadric recovery method [22] and the recover-and-select paradigm [14,15]. The result of this method using deformed superquadrics is shown in Fig. 2. In the range image of a doll[1] (a), (b), a set of seed superquadrics are initiated. Models after first model selection are shown in (c), and in the middle of recovery in (d). The final result is shown in (e), (f), displayed against the input range points.

We will first describe the recover-and-select paradigm and then the integration of superquadrics into the paradigm. The extension of this method for construction of a part-level hierarchy is the topic of section 4.

3.1. Recover-and-Select Paradigm

The recover-and-select paradigm consists of two intertwined stages: model-recovery and model-selection. At the model-recovery stage a redundant set of models (in this particular case, superquadrics) is initiated in the image and allowed to grow, which involves an iterative procedure combining data classification and parameter estimation. All recovered models are passed to the model-selection procedure where only the models resulting in the simplest overall description are selected. By combining model-recovery and model-selection in an iterative scheme a computationally efficient procedure is achieved.

[1] We thank Frank P. Ferrie from McGill University, Montreal, Canada for providing the image of the doll.

3.1.1. Model Recovery

Recovery of parametric models is difficult because one has to solve **two** problems:

1. find image elements that belong to a single parametric model, **and**
2. determine the values of the parameters of the model.

For image elements that have already been classified (segmented) one can determine the parameters of a model by applying standard statistical estimation techniques. Conversely, knowing the parameters of the model, a search for compatible image points can be accomplished by pattern classification methods. In the recover-and-select paradigm these **two** problems are solved simultaneously by an iterative method, conceptually similar to the one described by Besl [4], which combines data classification and model fitting. One of the crucial dilemmas is where to find the initial estimates (seeds) in an image since their selection has a major effect on the success or failure of the overall procedure. To solve this problem we propose to *independently* build *all possible* models from *all* statistically consistent seeds and to use them as hypotheses for the final description. The result of such model-recovery procedure consists for each particular model M_i of a triple (1. the set of data elements n_i that belong to the model M_i, 2. the type of the parametric model and the corresponding set of parameters of the model, 3. the goodness-of-fit value ξ_i to describe the conformity between the data and the model) which are subsequently passed to the model-selection procedure.

3.1.2. Model Selection

The model-recovery procedure provides a redundant representation where several of the models are completely or partially overlapped. Now only those models that produce the simplest description are selected. Intuitively, this reduction in complexity coincides with a general notion of simplicity which has a long history in psychology (Gestalt principles). The formalization of this principle led in information theory to the method of *Minimum Description Length* MDL, which has recently found its way to computer science, including computer vision [13,20].

The task of model selection is thus posed as an optimization problem. The objective function, which is to be maximized in order to produce the "best" description in terms of models, encompasses the information about the individual models and the overlap between them. Maximizing the objective function belongs to the class of problems known as combinatorial optimization (Quadratic Boolean problem). Since the number of possible solutions increases exponentially with the size of the problem, it is usually not tractable to explore them exhaustively. Due to the specific nature of the problem, a reasonable solution can be obtained by a direct application of the *greedy algorithm* which at any individual stage selects the option which is locally optimal. In other words, the models are selected in the sequence that corresponds to the size of their contributions to the objective function, which is equivalent to applying at each stage of the algorithm the *winner takes all* principle.

3.1.3. Model-Recovery and Model-Selection

In order to achieve a computationally efficient procedure the model-recovery and model-selection procedures are combined in an iterative fashion. The recovery of currently active models is interrupted by the model-selection procedure which selects a set of currently optimal models which are then passed back to the model-recovery procedure. This process is repeated until the remaining models are completely recovered. Several trade-offs can be selected in the dynamic combination of these two procedures. For details on the recover-and-select paradigm we refer the reader to [14].

3.2. *Superquadric Models*

Superquadrics are popular volumetric primitives in computer vision [2,7,8,10,11,16,19,20,21, 22,23,25] the reason being that they are convenient part-level models that can be further deformed and glued together to model articulated objects.

Superquadric surface is defined by the following equation

$$F(x,y,z) = \left(\left(\frac{x}{a_1}\right)^{\frac{2}{\epsilon_2}} + \left(\frac{y}{a_2}\right)^{\frac{2}{\epsilon_2}}\right)^{\frac{\epsilon_2}{\epsilon_1}} + \left(\frac{z}{a_3}\right)^{\frac{2}{\epsilon_1}}. \qquad (1)$$

Changing shape parameters ϵ_1 and ϵ_2 provokes the change of the whole shape ranging from parallelepipeds to cylinders and ellipsoids. Modeling capabilities of superquadrics can be enhanced by deforming them in different ways, such as global tapering and bending, i.e. [22] or local deformations for detailed modeling, e.g. [25].

In our part-level segmentation method we used for recovery of superquadric methods the following fitting function:

$$f = \sqrt{a_1 a_2 a_3}\,(F^{\epsilon_1} - 1), \qquad (2)$$

which is based on Eq. (1) for general position. This function has 11 parameters; accounting for size (a_1, a_2, a_3), shape (ϵ_1, ϵ_2), and position in space. For recovery an iterative least-squares method is used to compute the parameters of the model. Global deformations of superquadrics require some additional parameters that can be computed in the same way. The recovery method is described in detail in [22].

3.3. *Superquadrics in the Recover-and-select Paradigm*

Initial seeds are defined as squares in a grid-like pattern laid over the range image. We used small constant-size squares whose size could be adaptively changed depending on the task. A superquadric model is fitted to the data set in each seed. Seeds placed across part boundaries are eliminated because of a poor fit outright or later during model-selection. A decision whether a model should grow further or not, depends on the established similarity between the model and the data. If sufficient similarity is established, the currently estimated parameters, together with the current data set, are accepted and the search for more compatible points is started. While the actual fitting of superquadric models is based on Eq. 2, the comparison and selection of models is based on the approximation of the Euclidean distance[2].

In accordance with the paradigm, a search for more compatible points is performed in the vicinity of each model. This is achieved simply by enlarging the size of a particular model and testing range points included in such larger models. Initially, we enlarged the model

[2]The measure of goodness in Eq. 2 is efficient for fast and robust recovery but varies across the surface of the model and depends on the size of the model.

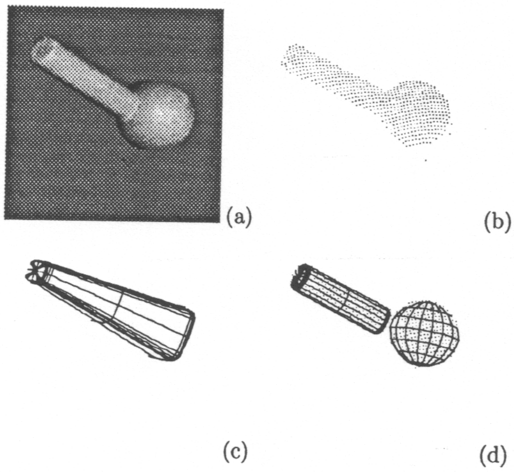


Figure 3: A sphere atop of a cylinder. Depending on the selected level of description, the system can recover from the scene (a-b): a tapered cylinder (c) or a cylinder and a sphere (d).

uniformly in all directions (multiplying all size parameters). Now, we enlarge the models only in those directions that corrupt the model in a lesser degree (if by enlargement points belonging to different part-models are included, those points provoke a very poor fit). On this enlarged set of points, a new superquadric model-recovery procedure is started.

We reported the preliminary results of this part segmentation method in [16,23]. The segmentation method is stable with respect to perturbations since we achieved stable part segmentation of objects even in range images taken from very different viewpoints [11]. Of course the part segmentation depends directly on the primitives chosen. This was even our primary motivation—*what is the best description of a scene given a particular shape language?* Since superquadrics seem to be a quite universal language for part-level shape description [18] which also closely corresponds to Hoffman's notion of parts [9] the obtained segmentation is in general perceptually acceptable. Those parts of the image that cannot be modeled with superquadrics (if the goodness of fit is not acceptable) remain un-modeled—an indication that another model type is required. We are still engaged in improving the described segmentation method and extending it in various directions. Merging of uncalibrated range images from several views, aided by such part-level segmentation, for reverse engineering purposes is one of the possible applications [11].

The segmentation method is being tested on a variety of real and synthetic range images. Processing of examples shown in this article took on the average less than 10 minutes (on a workstation HP-715/50). However, models could be recovered in parallel. The computation of the superquadric fitting function and its derivatives is independent for each range point and could be also done in parallel.

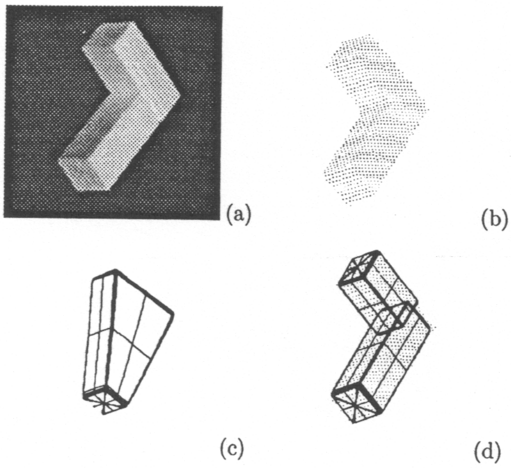


Figure 4: An L-shaped object: (a) intensity image, (b) range points, (c) coarse level description with a tapered superquadric, (d) finer description with two parallelepipeds.

4. Building a Part Hierarchy

In this section we explain how the described part segmentation can be used to recover different levels of part description which was discussed in the introduction. If segmentation is separated from shape recovery of individual part models normally only one level of description in a part hierarchy can be recovered—we call it the basic part level. With a generalization or abstraction mechanism several adjoint parts could be merged into a single part on a coarser scale. Instead of joining smaller parts we propose to directly search for parts on any specific coarseness level. To enable the recovery of part models on a coarser level we are testing the following two techniques:

1. Increasing the threshold for model acceptance in the model-recovery procedure. This enables the superquadric models to grow over part boundaries on a lower level of the coarseness hierarchy.

2. Modifying the fitting function for superquadric recovery. The value of this function is zero for points on the model surface and increases with the square of distance for points away from the surface. The function can be modified so that all points inside a specified distance from the actual surface are also set to zero. By increasing this distance a coarser level of description should be achieved.

In the examples shown (Figs. 3, 4)[3] we obtained two levels in the part description hierarchy. We experiment with both techniques described above and would like to automatically determine the thresholds/distances for different model granularities. Finally, the necessary

[3]These two range images were kindly provided by Marjan Trobina from ETH, Zürich, Switzerland.

coarseness level for description should be determined by a task. We studied the problem of automatic determination of the appropriate scale of observation (for a particular task) in [24].

5. Conclusions

Since the combined recovery/segmentation method was made available only recently the experiments on recovering part hierarchies of different granularity are still under way. Some preliminary results are shown in this article. This approach to different coarseness levels of description is possible only because of a direct and simultaneous part-level segmentation where individual features (range points) are directly evaluated for the final description. We believe that recovering part hierarchies could be especially useful for object recognition and object manipulation since for these tasks only shape/structure recovery up to a discriminating level of detail is required.

There are still several open issues that we will address in our future work. Besides improving the basic part-level segmentation method, the central problem is how to determine different levels in the part description hierarchy and how to control the switching between these levels.

6. Acknowledgments

This research was supported in part by The Ministry for Science and Technology of The Republic of Slovenia (Project P2-1122). We thank Aleš Jaklič, Alenka Macerl, and Marko Rous for advice and help in programming.

7. References

[1] P. K. Allen and P. Michelman. Acquisition and interpretation of 3-D sensor data from touch. *IEEE Trans. on Robotics and Automation*, 6(4):397–404, 1990.

[2] R. Bajcsy and F. Solina. Three dimensional shape representation revisited. In *Proc. First Int. Computer Vision Conf.*, pp. 231–241, London, England, 1987.

[3] R. Bajcsy, F. Solina, and A. Gupta. Segmentation versus object representation – are they separable? In R. Jain and A. Jain, editors, *Analysis and Interpretation of Range Images*, Springer, New York, 1990.

[4] P. J. Besl. *Surfaces in Range Image Understanding.* Springer, 1988.

[5] I. Biederman. Human image understanding: recent research and theory. *Computer Vision, Graphics, and Image Processing,* 32:29–73, 1985.

[6] T. O. Binford. Visual perception by computer. *Proc. of the IEEE Conf. on Systems and Control,* Miami, 1971.

[7] F. P. Ferrie, J. Lagarde, and P. Whaite. Darboux Frames, Snakes, and Super-Quadrics: Geometry from the Bottom Up. *IEEE Trans. on Pattern Analysis and Machine Intelligence,* PAMI-15(8):771–784, 1993.

[8] A. Gupta and R. Bajcsy. Surface and volumetric segmentation of range images using biquadrics and superquadrics. In *Proc. Int. Conf. on Pattern Recognition,* The Hague, pp. 158–162, 1992.

[9] D. D. Hoffman and W. A. Richards. Parts of recognition. *Cognition,* 18:65–96, 1985.

[10] T. Horikoshi and S. Suzuki. 3D parts decomposition from sparse range data using information criterion. In *Proc. IEEE Computer Vision and Pattern Recognition Conf.*, pp. 168–173, 1993.

[11] A. Jaklič, A. Leonardis, and F. Solina. Iterative construction of CAD models from range images. *Proceedings Fourth COST #229 WG.1 and 2 Worskshops on Adaptive Methods and Emergent Techniques for Signal Processing and Communications.* Ljubljana, Slovenia, April 1994.

[12] D. J. Kriegman and J. Ponce. On recognizing and positioning curved 3D objects from image contours. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1990.

[13] Y. G. Leclerc. Constructing simple stable descriptions for image partitioning. *Int. Journal of Computer Vision*, 3:73–102, 1989.

[14] A. Leonardis. *Image Analysis Using Parametric Models: Model-Recovery and Model-Selection Paradigm.* A Dissertation in Computer Science. University of Ljubljana, Faculty of Electrical Engineering and Computer Science. May 1993.

[15] A. Leonardis, A. Gupta, and R. Bajcsy. Segmentation of range images as the search for geometric parametric models. To appear in *Int. Journal on Computer Vision,* 1994.

[16] A. Leonardis, F. Solina, and A. Macerl. A Direct Recovery of Superquadric Models in Range Images Using Recover-and-Select Paradigm. *Proc. Third European Conf. on Computer Vision—ECCV'94*. Stockholm, Sweden, 1994.

[17] D. Marr. *Vision.* San Francisco, Freeman, 1982.

[18] A. P. Pentland. Perceptual organization and the representation of natural form. *Artificial Intelligence*, 28(2):293–331, 1986.

[19] A. P. Pentland. Recognition by parts. In *Proc. First Int. Conf. on Computer Vision*, London, pp. 612–620, 1987.

[20] A. P. Pentland. Automatic extraction of deformable part models. *Int. Journal of Computer Vision,* 4:107–126, 1990.

[21] A. P. Pentland and S. Sclaroff. Closed-form solutions for physically based shape modeling and recognition. *IEEE Trans. on Pattern Analysis and machine Intelligence*, PAMI 13(7):715–729, 1991.

[22] F. Solina and R. Bajcsy. Recovery of parametric models from range images: The case for superquadrics with global deformations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(2):131–147, 1990.

[23] F. Solina, A. Leonardis, and A. Macerl. A direct part-level segmentation of range images using volumetric models. In *Proc. IEEE Int. Conf. on Robotics and Automation*, San Diego, CA, 1994.

[24] F. Solina and A. Leonardis. Selective scene modeling. In *Proc. 11th Int. Conf. on Pattern Recognition*, pp. A:87–90, The Hague, 1992. IAPR.

[25] D. Terzopoulos and D. Metaxas. Dynamic 3D models with local and global deformations: Deformable superquadrics. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI 13(7):703–714, 1991.

[26] R. Volpe and P. Khosla. Manipulator control with superquadric artificial potential functions: Theory and experiments. *IEEE Trans. on Systems, Man, and Cybernetics*, 20(6):1423–1436, 1990.