# Constrained Recovery of Deformable Models from Range Data*

Sven J. Dickinson† Dimitri Metaxas‡ Alex Pentland§

**Abstract**

We present a technique that uses aspects to constrain the fitting of deformable models to range data. Unlike previous work in deformable model recovery from range data, the technique does not depend on accurate initial position and orientation of the model. Furthermore, occlusion is handled at segmentation time and does not complicate the fitting process, as only 3-D points known to belong to a part participate in the fitting of a model to the part. We present the approach in detail and apply it to the recovery of objects from range data.

## 1 Introduction

The recovery of volumetric shape descriptions from range data has drawn much attention in the literature, e.g., [15, 7, 12, 16, 13, 5]. While each of these approaches addresses the problem of recovering a deformable model, superquadric, or set of modes corresponding to a part, many avoid the issue of part segmentation. In some cases, either a single unoccluded object appears in the image or part segmentation is performed manually [13, 16]. In other cases, decomposition of the image into parts is integrated into the fitting process, resulting in a costly global optimization process [15, 12]. Furthermore, in many cases, the fitting process is sensitive to initial placement and orientation of the model. If its initial position is not inside the data or if

*The authors gratefully acknowledge the assistance of Dr. Tim Newman and Sharath Pankanti of the Michigan State University Pattern Recognition and Image Processing (PRIP) Laboratory, in providing both the range image segmentation code as well as the original images (originally captured by Pat Flynn). The authors would also like to thank Gene Amdur at the University of Toronto and Michael Chan at the University of Pennsylvania for their support and assistance in the implementation of this work. The second author was supported by an NSF-IRI-9309917 grant.

†Media Laboratory, Massachusetts Institute of Technology, and Department of Computer Science, University of Toronto.

‡Department of Computer and Information Science, University of Pennsylvania.

§Media Laboratory, Massachusetts Institute of Technology.

its $z$ axis is not closely aligned with the principal axis of the data, a canonical fit may not be achieved.

In this paper, we propose a novel two-step shape recovery process that decouples the tasks of segmenting a range image into parts and fitting deformable models to the parts. In the first step, the range image is segmented into a set of homogeneous regions. Next, the regions are grouped into aspects or views corresponding to a vocabulary of 3-D parts. For unambiguous views, this process yields the qualitative shape of the part, the qualitative orientation of the part through its aspect, and an exact mapping between the regions in the recovered aspect and the surfaces on the part. This information is used in the second stage to provide strong constraints on the fitting of a deformable model to the segmented range data. The 3-D data corresponding to the contours that define the recovered aspects are used to recover a part's global deformations, while the 3-D data corresponding to pixels bounded by aspect faces are used to recover a part's local deformations. Finally, the initial model can be specified in *any* initial size, position, and orientation with correct convergence guaranteed by the constraints.

Following the introduction, we review both our qualitative and quantitative shape representations. Next, we present our qualitative and quantitative shape recovery techniques. Finally, we apply the techniques to a set of range images and close with some conclusions and limitations.

# 2 Related Work

The qualitative and parametric approaches to representing shape have complementary properties. The qualitative representation of part structure has proven [4, 3] useful for segmentation and grouping, while the parametric representations have proven useful for recovering precise descriptions of shape [12, 13, 15, 16]. It is therefore natural to try to combine the strengths of the two approaches, using one for grouping and segmentation, and the other for fitting and description.

Some work has already proceeded along these lines. Using a part-based aspect approach to segmentation based on Dickinson et al. [4], Raja and Jain [14] segment a range image into parts corresponding to geons. In order to determine geon orientation, i.e., end vs. side faces, they fit a superquadric to the segmented part to determine the principal axis of the geon. The technique combines qualitative models for segmentation but does not attempt to recover a precise parametric model. Instead, the static superquadric fitting step is used only as an aid for geon labeling. Metaxas and Dickinson [10] present an approach in which recovered qualitative shape is used to constrain the recovery of a deformable model (global deformations only) from 2-D image data. This paper extends that approach to 3-D data and includes local deformation recovery.

# 3 Object Modeling

Our object representation integrates both qualitative and quantitative shape models. In the following sections, we briefly review these two components and how they have been extended to support shape recovery from range data. Further details can be found in [4, 3] and [16], respectively.

## 3.1 Qualitative Shape Modeling

To demonstrate our approach to shape recovery, we have selected an object representation similar to that used by Biederman [1], in which the Cartesian product of contrastive shape properties gives rise to a set of volumetric parts called geons. For our investigation, we have chosen three properties including cross-section shape, axis shape, and cross-section size variation (Dickinson, Pentland, and Rosenfeld [4]). The values of these properties give rise to a set of ten volumetric parts (a subset of Biederman's geons), shown in Figure 1(a). To construct objects, the volumes are attached to one another with the restriction that any junction of two volumes involves exactly one distinct surface from each volume.

Traditional aspect graph representations of 3-D objects model an entire object with a set of aspects, each defining a topologically distinct view of the object in terms of its visible surfaces (Koenderink and van Doorn [9]). Our approach differs in that we use aspects to represent a (typically small) set of volumetric parts from which each object in our database is constructed, rather than representing an entire object directly. Consequently, our goal is to use aspects to recover the 3-D volumetric parts that make up the object in order to carry out a recognition-by-parts procedure, rather than attempting to use aspects to recognize entire objects. Since the number of qualitatively different volumes is generally small, the number of possible aspects is limited and, more important, *independent* of the number of objects in the database. However, if a volume is occluded from a given 3-D viewpoint, its projected aspect in the image will also be occluded. We accommodate the matching of occluded aspects through the use of a hierarchical representation we call the *aspect hierarchy*.

The aspect hierarchy consists of three levels, including the set of *aspects* that model the chosen volumes, the set of component *faces* of the aspects, and the set of *boundary groups* representing all subsets of contours bounding the faces. Figure 1(b) illustrates a portion of the aspect hierarchy, along with a few of the volumes. The ambiguous mappings between the levels of the aspect hierarchy are captured in a set of conditional probabilities, mapping boundary groups to faces, faces to aspects, and aspects to volumes. These conditional probabilities result from a statistical analysis of a set of images approximating the set of *all* views of *all* the volumes.

The aspect hierarchy was originally introduced as a representation to support 3-D object recognition from 2-D images (Dickinson, Pentland, and Rosenfeld [4, 3]). With faces in the aspect hierarchy now representing 3-D surfaces instead of 2-D projections

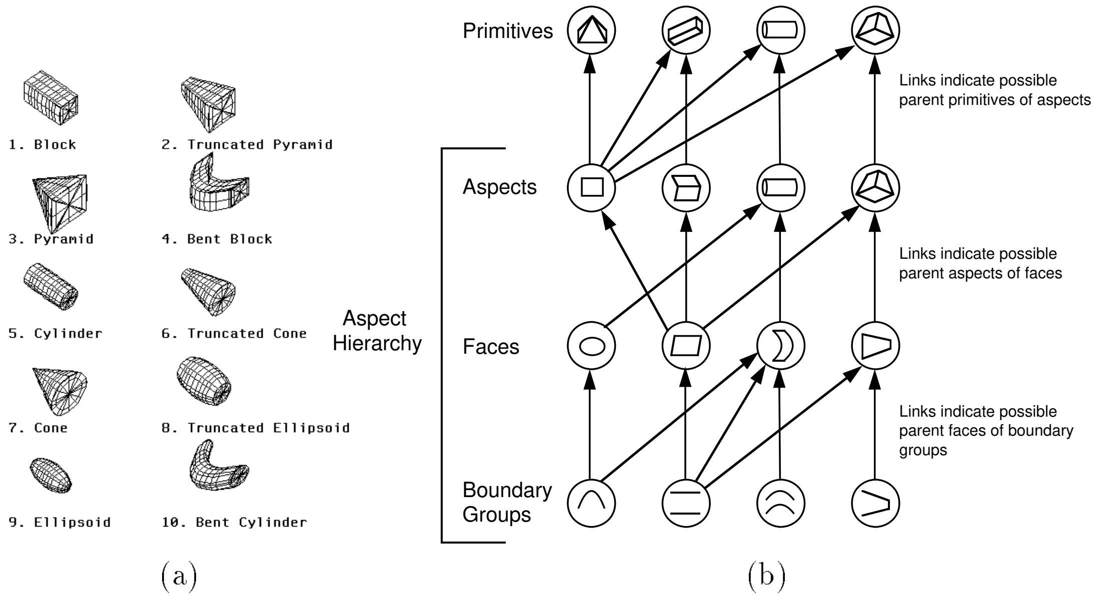


Figure 1: (a) The Ten Volumetric Modeling Parts (b) The Aspect Hierarchy

of 3-D surfaces, much of the ambiguity of the original aspect hierarchy has been removed. Adding face attributes such as Mean and Gaussian curvature can effectively prune many of the mappings from boundary groups to faces, faces to aspects, and aspects to volumes. However, for the experiments presented in this paper, the original aspect hierarchy has been used. For high probability views of volumes, there is little or no ambiguity in the mapping from aspects to volumes, providing a unique labeling of the object. In current work, we are constructing an aspect hierarchy which takes surface curvature into account.

## 3.2 Quantitative Shape Modeling

Geometrically, the models used in this paper [11] are closed surfaces in space whose intrinsic (material) coordinates are $\mathrm{u} = (u, v)$, defined on a domain $\Omega$. The positions of points on the model relative to an inertial frame of reference $\Phi$ in space are given by a vector-valued, time-varying function of u:

$$\mathbf{x}(\mathrm{u}, t) = (x_1(\mathrm{u}, t), x_2(\mathrm{u}, t), x_3(\mathrm{u}, t))^\top \tag{1}$$

where $^\top$ is the transpose operator. We set up a noninertial, model-centered reference frame $\phi$, and express these positions as:

$$\mathbf{x} = \mathbf{c} + \mathbf{R}\mathbf{p}, \tag{2}$$

where $\mathbf{c}(t)$ is the origin of $\phi$ at the center of the model, and the orientation of $\phi$ is given by the rotation matrix $\mathbf{R}(t)$. Thus, $\mathbf{p}(\mathrm{u}, t)$ denotes the canonical positions of points on the model relative to the model frame. We further express $\mathbf{p}$ as the sum of a global reference shape $\mathbf{s}(\mathrm{u}, t)$ (global deformation) and a displacement function $\mathbf{d}(\mathrm{u}, t)$ (local deformation):

$$\mathbf{p} = \mathbf{s} + \mathbf{d}. \tag{3}$$

We define the global reference shape as

$$\mathbf{s} = \mathbf{T}(\mathbf{e}(\mathrm{u}; a_0, a_1, \ldots); b_0, b_1, \ldots). \tag{4}$$

Here, a geometric primitive $\mathbf{e}$, defined parametrically in u and parameterized by the variables $a_i$, is subjected to the *global deformation* $\mathbf{T}$ which depends on the parameters $b_i$. Although generally nonlinear, $\mathbf{e}$ and $\mathbf{T}$ are assumed to be differentiable (so that we may compute the Jacobian of $\mathbf{s}$) and $\mathbf{T}$ may be a composite sequence of primitive deformation functions $\mathbf{T}(\mathbf{e}) = \mathbf{T}_1(\mathbf{T}_2(\ldots \mathbf{T}_n(\mathbf{e})))$. We concatenate the global deformation parameters into the vector

$$\mathbf{q}_s = (a_0, a_1, \ldots, b_0, b_1, \ldots)^\top. \tag{5}$$

Even though our technique for defining $\mathbf{T}$ is independent of the primitive $\mathbf{e} = (e_1, e_2, e_3)^\top$ to which it is applied, we will use superquadric ellipsoid primitives, with linear tapering and bending global deformations [11].

For local deformations, we express the displacement $\mathbf{d}$, as a linear combination of finite element shape functions [8] and local deformation parameters:

$$\mathbf{d} = \mathbf{S}\mathbf{q}_d, \tag{6}$$

where $\mathbf{S}$ is a shape matrix whose entries are the shape functions and

$$\mathbf{q}_d = (\ldots, \mathbf{d}_i^\top, \ldots)^\top \tag{7}$$

is the vector of local deformation parameters [16].

# 4 Kinematics, Dynamics and Generalized Forces

Differentiating (2), the velocity of points on the model is given by:

$$\dot{\mathbf{x}} = [\mathbf{I}\ \mathbf{B}\ \mathbf{RJ}\ \mathbf{RS}]\dot{\mathbf{q}} = \mathbf{L}\dot{\mathbf{q}}, \tag{8}$$

where $\mathbf{J}$ is the Jacobian of the reference shape with respect to the global deformation parameter vector, and $\mathbf{L}$ is the Jacobian matrix that converts $q$-dimensional vectors to 3-D vectors [11]. Here,

$$\mathbf{q} = (\mathbf{q}_c^\top, \mathbf{q}_\theta^\top, \mathbf{q}_s^\top, \mathbf{q}_d^\top)^\top, \tag{9}$$

with $\mathbf{q}_c = \mathbf{c}$ and $\mathbf{q}_\theta = \boldsymbol{\theta}$, serves as the vector of generalized coordinates of the model.

When fitting the model to visual data, our goal is to recover $\mathbf{q}$, the vector of degrees of freedom of the model. We make the kinematic model dynamic based on the use of Lagrangian mechanics and appropriate simplifications suitable for static shape recovery [16, 11]. The dynamic equations of motion then take the form

$$\mathbf{D}\dot{\mathbf{q}} + \mathbf{K}\mathbf{q} = \mathbf{f}_q, \tag{10}$$

where $\mathbf{D}$ is the damping matrix and $\mathbf{K}$ is the stiffness matrix that defines the elastic properties of the model. $\mathbf{f}_q$ are the generalized external forces associated with the components of $\mathbf{q}$ and are computed using

$$\mathbf{f}_q = \int \mathbf{L}^\top \mathbf{f}\, d\mathrm{u}, \tag{11}$$

where $\mathbf{f}(\mathrm{u}, t)$ is the force distribution applied to the model through our physics-based approach to visual estimation. The visual data are converted into traction forces that act on the dynamic model [16] by appropriately assigning data points to points on the model and exerting a force on the model which is proportional to the distance between them.

# 5 Shape Recovery

## 5.1 Qualitative Shape Recovery

From an input range image, we apply Flynn's range image region segmentation algorithm [6]. From the resulting 2-D region label image, we build a *region topology graph*, in which nodes represent regions and arcs specify region adjacencies. Each node (region) encodes the 2-D bounding contour of a region as well as a mask which specifies pixel membership in the region. From the region topology graph, each region is characterized according to the qualitative shapes of its bounding contours (see [2] for details). The result is a *region boundary graph* representation for a region, in which nodes represent bounding contours, and arcs represent pairwise nonaccidental relations between the contours.

Face labeling consists of matching a region boundary graph (and its subgraphs) to the graphs representing the model faces (and boundary groups) in the aspect hierarchy using an interpretation tree search. The result is a *face topology graph* in which each node contains a set of face labels (sorted by decreasing order of probability) associated with a given region.

In an unexpected object recognition domain, in which there is no a priori knowledge of scene content, we search through the space of groupings of regions into aspects. From a set of recovered aspects, we search through the space of volume interpretations that is consistent with the objects in a database (see [4, 3] for details). In an expected object recognition domain, in which we are searching for a particular object or part, we use the aspect hierarchy as an attention mechanism to focus the search for an aspect at appropriate regions in the image [2]. Moving down the aspect hierarchy, target objects map to target volumes which, in turn, map to target aspect predictions which, in turn, map to target face predictions. Those faces in the face topology graph whose labels match the target face prediction provide an ordered (by decreasing probability) set of ranked search positions at which the target aspect prediction can be verified If the mapping from a verified aspect to a target volume is ambiguous, this attention mechanism can be used to drive an active recognition system which moves the cameras to obtain a less ambiguous view of an object's part [2].

## 5.2 Quantitative Shape Recovery

In the dynamic model fitting process, the data are transformed into an externally applied force distribution $\mathbf{f}(\mathrm{u}, t)$. We convert the external forces to generalized forces $\mathbf{f}_q$ which act on the generalized coordinates of the model [16, 11]. We apply forces to the model based on differences between the model's points and the 3-D data. Each of these forces is then converted to a generalized force $\mathbf{f}_q$ that, based on (10), modifies the appropriate generalized coordinate that has to be adapted so that the model fits the data.

Given that our vocabulary of volumes is limited, we devise a systematic way of computing the generalized forces for each volume. The computation depends on the influence of particular parts of the data to model degrees of freedom. Such parts correspond to the various regions making up the aspect used to identify the volume's coarse shape. From the correspondence between the the 3-D points which project to the bounding contours of each region in the recovered aspect and the corresponding points on the model, we use (11) to define forces that will affect the global deformations of the model. Next, from the correspondence between the 3-D points internal to a region and their nearest points on the model, we use (11) to define forces which will affect the local deformations of the model. In the case of occluded volumes, resulting in both occluded aspects and occluded faces, only those portions (boundary groups) of the faces used to define the faces exert external forces on the models.

### 5.2.1 Model Initialization

One of the major limitations of previous deformable model fitting approaches is their dependence on model initialization and prior segmentation [16, 13]. Using the qualitative shape recovery process as a front end, we first segment the data into parts, and for each part, we identify the relevant non-occluded data belonging to the part. In addition, the extracted qualitative volumes explicitly define a mapping between the image faces in their projected aspects and the 3-D surfaces on the quantitative models. Moreover, the extracted volumes can be used to immediately constrain many of the global deformation parameters. For example, from the qualitative shape classes, we know if a volume is bent, tapered, or has an elliptical cross-section. Finally, although the initial model can be specified at any position and orientation, the aspect that a volume encodes defines a qualitative orientation that can be exploited to speed

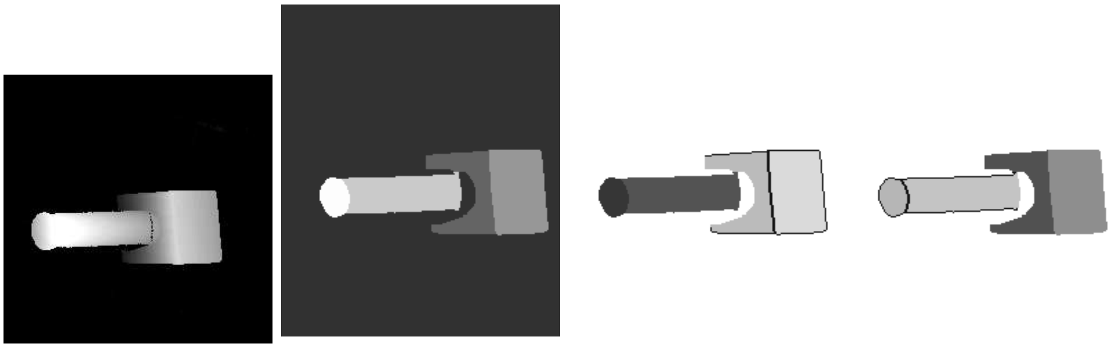

Figure 2: Left to right: range image, region segmentation, recovered qualitative block, and recovered qualitative cylinder

up the model fitting process. Sensitivity of the fitting process to model initialization is also overcome by independently solving for the degrees of freedom of the model. By allowing each face in an aspect to exert forces on only one model degree of freedom at a time, we remove local minima from the fitting process and ensure correct convergence of the model.

# 6 Results and Discussion

We demonstrate our approach by recovering the two parts in the range image shown in Figure 2(a)[1]; the region segmented image is shown in Figure 2(b). For this example, we invoked the expected object recognition mode to first search for the best instance of a block. Figure 2(c) shows the highlighted aspect recovered for the block; only those contours used to infer the block are highlighted in the image. Note that the most probable aspect for the block (containing three faces) was not recovered; however, the next most probably aspect (containing two faces) was recovered and used to locate the block. Figure 2(d) shows the highlighted aspect recovered for the cylinder.

For each of the two recovered qualitative volumes, we now proceed to show the results of using the recovered qualitative shape to constrain the fitting of a deformable model to the original range data. In Figures 3(a-c), we show the initial, an intermediate, and the final frame in the sequence of model deformations taking the initial model to its final shape describing the block. For clarity, only the recovered contours belonging to the block's recovered aspect are shown in the sequence. Next, in Figures 3(d-f), we show the initial, two intermediate, and the final frame in the sequence of model deformations taking the initial model to its final shape describing the cylinder. Finally, in Figure 3(h), we show the two fitted parts together from a different viewpoint.

[1]The image was captured using a Technical Arts Scanner at the Michigan State University's PRIP Laboratory.

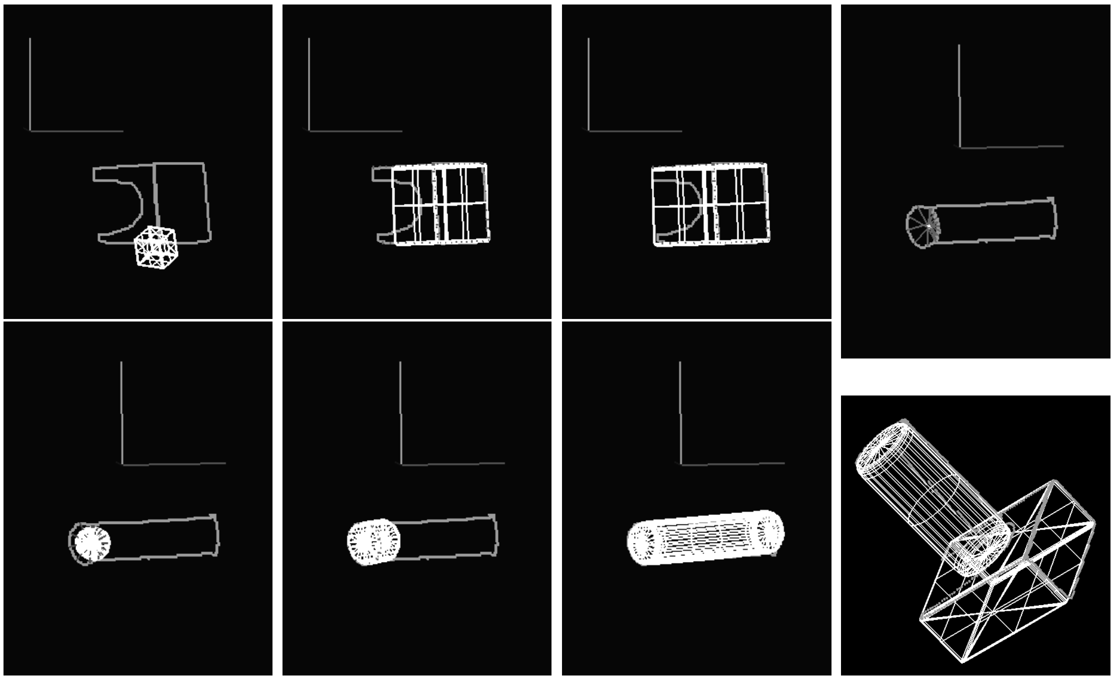

Figure 3: Selected Frames from the Part Fitting Sequence: (a) Initial Block Frame; (b) Intermediate Block Frame; (c) Final Block Frame; (d) Initial Cylinder Frame; (e) Intermediate Cylinder Frame; (f) Second Intermediate Cylinder Frame; (g) Final Cylinder Frame; (h) Recovered Object from a Different Viewpoint.

The fitting constraints provided by the recovered qualitative shape means that the deformable model fitting process is invariant to initial position, orientation, and shape of the model. Unfortunately, relying on the correspondence between recovered image faces and model surfaces means that the recovery process is sensitive to region segmentation errors. However, by only allowing high-scoring volumes to constrain the fitting process, the chances of letting any region segmentation problems affect the fitting process is low. In fact, low-scoring volumes can be used to guide the sensor to acquire a higher-scoring volume.

In conclusion, by first recovering the coarse shape of an object's parts through a part-based aspect matching strategy, we not only address the part segmentation problem prior to fitting, but provide a powerful set of fitting constraints ensuring that convergence of the deformable model is invariant to initial size, position, or orientation of the model. Furthermore, we extend our previous method of shape recovery from 2-D data which recovered only global deformations to shape recovery from 3-D data which recovers both global and local deformations.

# References


[1] I. Biederman. Human image understanding: Recent research and a theory. *CVGIP:IU*,

32:29–73, 1985.

[2] S. Dickinson, H. Christensen, J. Tsotsos, and G. Olofsson. Active object recognition integrating attention and viewpoint control. In *Proceedings, ECCV '94*, Stockholm, May 1994.

[3] S. Dickinson, A. Pentland, and A. Rosenfeld. From volumes to views: An approach to 3-D object recognition. *CVGIP:IU*, 55(2):130–154, 1992.

[4] S. Dickinson, A. Pentland, and A. Rosenfeld. 3-D shape recovery using distributed aspect matching. *IEEE Transactions on PAMI*, 14(2):174–198, 1992.

[5] F. Ferrie, J. Lagarde, and P. Whaite. Darboux frames, snakes, and superquadrics. In *Proceedings, Workshop on Interpretation of 3D Scenes*, pages 170–176. IEEE Computer Society Press, 1989.

[6] P. Flynn. Cad-based computer vision: Modeling and recognition strategies. *Ph.D. thesis*, 1990.

[7] A. Gupta. Surface and volumetric segmentation of 3D objects using parametric shape models. Technical Report MS-CIS-91-45, GRASP LAB 128, University of Pennsylvania, Philadelphia, PA, 1991.

[8] K. Kardestuncer. *Finite Element Handbook*. McGraw–Hill, New York, 1987.

[9] J. Koenderink and A. van Doorn. The internal representation of solid shape with respect to vision. *Biological Cybernetics*, 32:211–216, 1979.

[10] D. Metaxas and S. Dickinson. Integration of quantitative and qualitative techniques for deformable model fitting from orthographic, perspective, and stereo projections. In *Proceedings, Fourth ICCV*, Berlin, May 1993.

[11] D. Metaxas and D. Terzopoulos. Shape and nonrigid motion estimation through physics-based synthesis. *IEEE PAMI, in press*, 1993.

[12] A. Pentland. Automatic extraction of deformable part models. *IJCV*, 4:107–126, 1990.

[13] A. Pentland and S. Sclaroff. Closed-form solutions for physically based shape modeling and recognition. *IEEE PAMI*, 13(7):715–729, 1991.

[14] N. Raja and A. Jain. Recognizing geons from superquadrics fitted to range data. *Image and Vision Computing*, 10(3):179–190, 1992.

[15] F. Solina and R. Bajcsy. Recovery of parametric models from range images: The case for superquadrics with global deformations. *IEEE PAMI*, 12(2):131–146, 1990.

[16] D. Terzopoulos and D. Metaxas. Dynamic 3D models with local and global deformations: Deformable superquadrics. *IEEE PAMI*, 13(7):703–714, 1991.