The Role of Model-Based Segmentation in the Recovery of Volumetric Parts from Range Data[*]

Sven J. Dickinson
Department of Computer Science and
Rutgers Center for Cognitive Science (RuCCS)
Rutgers University
P.O. Box 1179
Piscataway, NJ 08855-1179

Dimitri Metaxas
Department of Computer and
Information Science
University of Pennsylvania
Philadelphia, PA 19104-6389

Alex Pentland
Vision and Modeling Group
Media Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139

November 7, 1996

Abstract

We present a method for segmenting and estimating the shape of 3-D objects from range data. The technique uses model views, or aspects, to constrain the fitting of deformable models to range data. Based on an initial region segmentation of a range image, regions are grouped into aspects corresponding to the volumetric parts that make up an object. The qualitative segmentation of the range image into a set of volumetric parts not only captures the coarse shape of the parts, but qualitatively encodes the orientation of each part through its aspect. Knowledge of a part's coarse shape, its orientation, as well as the mapping between the faces in its aspect and the surfaces on the part provides strong constraints on the fitting of a deformable model (supporting both global and local deformations) to the data. Unlike previous work in physics-based deformable model recovery from range data, the technique does not require pre-segmented data. 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.

[*] A condensed version of this technical report is to appear in the IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI).

1 Introduction

The recovery of volumetric shape descriptions from range data has drawn much attention in the literature, e.g., [28, 14, 24, 31, 30, 25, 10, 11]. 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 [31, 25, 30]. In other cases, decomposition of the image into parts is integrated into the fitting process, resulting in a costly global optimization process [28, 24]. 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 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 three-step shape recovery process which first groups range pixels into homogeneous regions based on surface curvature. In the second step, 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 third and final 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 ensured by the constraints.

Following the introduction, we first present a new view-based qualitative representation for volumetric parts appearing in range images, followed by a review of our quantitative part representation. Next, we present our qualitative and quantitative shape recovery techniques, and show how the use of model-based segmentation can provide very strong constraints on the fitting of part models to range data. Finally, we apply the techniques to a set of range images and close with some conclusions and limitations.

2 Related Work

The first parts representation was due to Binford [3], who suggested the idea of generalized cylinders. Unfortunately, the recovery of this type of representation seems to require elaborate line grouping and reasoning, a difficult and largely unsolved problem. Moreover, because such descriptions are often not unique, it is unclear how they aid in object recognition. The recovery of restricted classes of generalized cylinders was first shown by Agin and Binford [1] and Nevatia and Binford [22], while recent results include the work of Ulupinar and Nevatia [32] and Zerroug and Nevatia [34].

The idea of generalized cylinders has subsequently been elaborated in two very different ways. One variation is due to Biederman [2], who suggested using the Cartesian product of qualitative properties such as tapering, cross-section, etc., in order to create a qualitative taxonomy of generalized cylinders. His theory was that the use of such a qualitative representation could simplify the process of segmenting objects into parts.

Dickinson, Pentland and Rosenfeld [9, 8] have extended Biederman's representation to include intermediate representations and conditional probabilities that guide the grouping process. Using this extended framework, they were able to demonstrate that it is possible to quickly segment 2-D images of objects into their component parts. An adaptation of this approach is used in this paper to group faces from range data into volumetric parts.

Another alternative to Binford's generalized cylinders was suggested by Pentland [23], who used superquadrics with parameterized global deformations. Use of a parameterized implicit function, such as the superquadric, converts the problem of recovering a description into a relatively simple numerical optimization. Using this approach, many researchers, starting with Solina and Bajcsy [29] and Pentland [24], have reported success at recovering superquadric models with global deformations from a variety of data types (Pentland and Sclaroff [25], Terzopoulos and Metaxas [30], Gupta [15], Ferrie et al. [10], Leonardis et al. [17], and Wu and Levine [33]).

At about the same time, Terzopoulos et al. [31] developed physics-based techniques for fitting deformable models with local deformations to visual data. These techniques provide a robust framework for fitting, and offer the possibility for natural extension to moving, dynamic scenes. Consequently, it is natural to apply physics-based techniques to the recovery of deformable superquadrics. During the last few years, such physics-based formulations have become the most popular method for the recovery of deformable superquadric models from range data (Pentland and Sclaroff [25], Terzopoulos and Metaxas [30], and Metaxas and Terzopoulos [20, 21]).

Recently, Metaxas and Terzopoulos [30, 20, 21] have extended this approach by developing a class of deformable models in which both global and local deformations are physics-based (deformable superquadrics are a special case of this class of models). The global deformations capture the salient structure of object parts, while the local deformations capture the object's details. This is the formulation used in this paper.

The qualitative and parametric approaches to representing shape have complementary properties. The qualitative representation of part structure has proven useful for segmentation and grouping, while the parametric representations have proven useful for recovering precise descriptions of shape. 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. [9], Raja and Jain [26] 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.

Dickinson and Metaxas [5] 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, by first introducing a new view-based representation for 3-D data. This representation, called the *range aspect hierarchy*, is similar in structure to the original aspect hierarchy introduced in [7, 9, 8], but with an additional surface curvature attribute recoverable from range data. Just as we used 2-D aspects to constrain 3-D shape recovery from a 2-D image, we can apply the same framework to the extraction of 3-D shape from a range image, including local deformation recovery (underconstrained in shape recovery from 2-D contours). The resulting framework thereby offers a unified approach to 3-D shape recovery and segmentation from range data.

3 Object Modeling

3.1 Qualitative Shape Modeling

In [7], we introduced a hybrid object representation for the recognition of 3-D objects from 2-D images. Objects were constructed from some finite set of object-centered, volumetric parts, while the parts were represented in the image as a finite set of hierarchically-defined viewer-centered aspects. Recently, Raja and Jain [27], extended this idea to the domain of range data where view-based representations were used to model the appearance of a finite set of volumetric parts. Like the aspect definition proposed in [7], Raja and Jain chose to eliminate symmetries from the aspect graph, so that all views having the same component shapes and adjacencies belonged to the same class regardless of which surfaces of the volumetric part the views mapped to. However, unlike our previous representation, Raja and Jain added an additional attribute to an aspect, namely surface shape (based on the signs of the maximum and minimum curvatures). This was facilitated by the use of range data as their sensor image type.

For the model we propose here, we extend our concept of the aspect hierarchy to support range data, and add the surface shape attribute to the face components of an aspect, as proposed by Raja and Jain [27]. As in the case of our original aspect hierarchy, our new view-based representation, called the *range aspect hierarchy*, is composed 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. As in our original aspect hierarchy, each aspect in the range aspect hierarchy consists of one or more faces from the next level down (the face level), along with a specification of how the faces are connected, i.e., for two adjacent faces, those contours of the faces make up the connection. Each face at the face level is defined by the shapes of its projected bounding contours in 2-D (e.g., straight, convex, or concave), as well as a surface shape attribute defined by the signs of the face's maximum and minimum curvatures. The lowest level of the range aspect hierarchy is again called the boundary group level, consisting of all subsets of the bounding contours of the faces. However, unlike the original aspect hierarchy, the surface type attribute of the face from which the boundary group is defined is included in the definition of a boundary group.

Figure 1 illustrates the ten volumetric classes used to demonstrate our approach, while Figure 2 illustrates a portion of the range aspect hierarchy. The ambiguous mappings between the levels of the range 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 range images approximating the set of *all* views of *all* the volumes.

The use of range data and the addition of the surface type attribute to faces and boundary groups means that there is an important distinction between the original aspect hierarchy (for 2-D data) and the range aspect hierarchy (for 3-D data). The addition of the surface shape attribute means a greater distinction between face types, resulting in many more face classes at the face level. For example, the original aspect hierarchy contained a parallelepiped face. However, in the range aspect hierarchy, the surface bounded by parallelepiped in the image could have 5 distinguishable shapes according to the signs of their principal curvatures, e.g., $(0,0)(+,0)(-,0)(-,-)(+,+)$ (note: there is no $(-,-)$ surface in our part vocabulary).

Since the aspects are made up of faces, there will be more aspects at the aspect level, and since boundary groups are components of faces, there will be more boundary groups at the boundary group level.[1] As a result, the ambiguity of the mappings between levels will decrease, resulting in less uncertain inferences between boundary groups and faces, faces and aspects, and aspects and volumes. We trade off the increased size of our model space with a decrease in inferencing ambiguity.

New objects can be added to the object database as long as their constituent parts are in the range aspect hierarchy's part vocabulary. However, if a new part is added to the part vocabulary, the range aspect hierarchy must be recomputed. Due to the simplicity of the

[1] The original aspect hierarchy for 2-D images has exactly 40 aspects and 18 faces. The range aspect hierarchy has exactly 54 aspects and 30 faces.

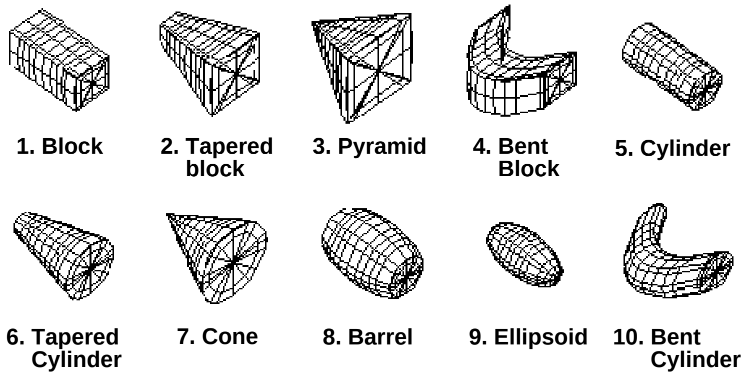


Figure 1: The Ten Modeling Primitives.

volumetric parts, no single part in the original or range aspect hierarchies has more than 12 aspects, while no single aspect has more than 5 faces. In the worst case, the number of additional aspects needed to encode a new part is equal to the number of aspects belonging to the new part, while the number of additional faces is equal to the number of component faces in the new aspects. However, in practice, there is considerable overlap between the aspects of the parts as well as between their component faces. By effectively breaking down the objects' aspect graphs into parts, we can avoid the tremendous complexity of traditional aspect graphs.

3.2 Quantitative Shape Modeling

Geometrically, the deformable models used in this paper are closed 3D surfaces. The time-varying positions of points on the model relative to an inertial frame Φ are given by $\mathbf{x}(\mathrm{u}, t)$, where u are the model's material coordinates defined over a domain Ω, and t is time. We also set up a noninertial, model-centered reference frame ϕ [19], and express these positions as:

$$\mathbf{x} = \mathbf{c} + \mathbf{R}\mathbf{p}, \tag{1}$$

where $\mathbf{c}(t)$ is the origin of ϕ at the center of the model, and the orientation of ϕ is given by the rotation matrix $\mathbf{R}(t)$. We further express $\mathbf{p}$ as the sum of a global reference shape $\mathbf{s}(\mathrm{u}, t)$ and a local displacement function $\mathbf{d}(\mathrm{u}, t)$.

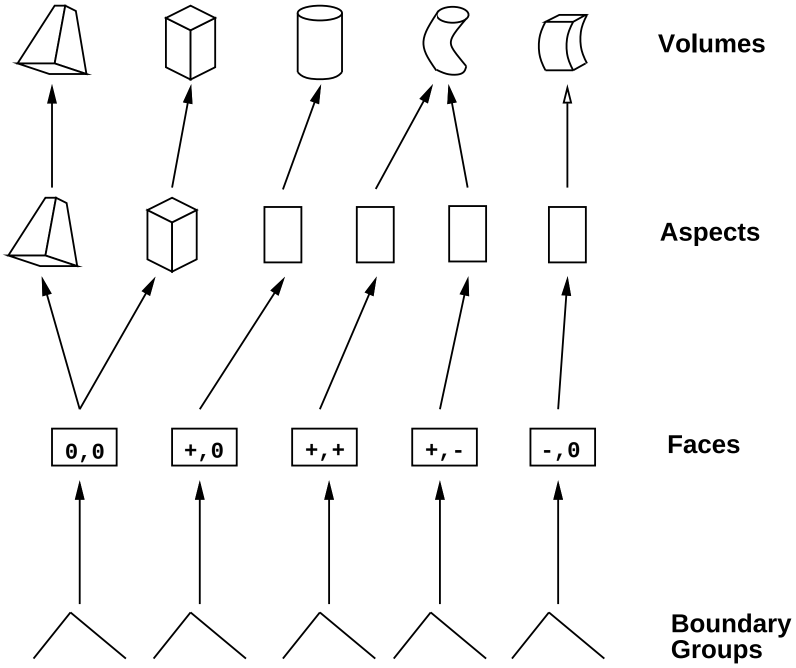


Figure 2: The Range Aspect Hierarchy.

We define the global reference shape as

$$\mathbf{s} = \mathbf{T}(\mathbf{e}(\mathrm{u};\, a_0, a_1, \ldots);\, b_0, b_1, \ldots). \tag{2}$$

Here, a geometric primitive $\mathbf{e}$, with material coordinates u and parameterized by the variables a_i, is subjected to the deformation $\mathbf{T}$ which depends on the parameters b_i. Although generally nonlinear, $\mathbf{e}$ and $\mathbf{T}$ are assumed to be differentiable and $\mathbf{T}$ may be a composite sequence of primitive deformation functions $\mathbf{T}(\mathbf{e}) = \mathbf{T}_1(\mathbf{T}_2(\ldots \mathbf{T}_n(\mathbf{e})))$. For the experiments shown in this paper we use as deformable models, superquadric ellipsoids with linear tapering along principal axes 1 and 2, and bending along principal axis 3 [21]. We then collect the parameters in $\mathbf{s}$ into the vector of global deformation parameters

$$\mathbf{q}_s = (a, a_1, a_2, a_3, \epsilon_1, \epsilon_2, t_1, t_2, b_1, b_2, b_3)^\top, \tag{3}$$

where $a \geq 0$ is a scale parameter, $0 \leq a_1, a_2, a_3 \leq 1$ are aspect ratio parameters, and $\epsilon_1, \epsilon_2 \geq 0$ are "squareness" parameters, $-1 \leq t_1, t_2 \leq 1$ are the tapering parameters in principal axes 1 and 2, respectively; b_1 defines the magnitude of the bending and can be positive or negative; $-1 \leq b_2 \leq 1$ defines the location on axis 3 where bending is applied; and $0 < b_3 \leq 1$ defines the region of influence of bending.

We express the local displacements $\mathbf{d}$ based on the theory of finite elements as

$$\mathbf{d} = \mathbf{S}\mathbf{q}_d, \tag{4}$$

where $\mathbf{S}$ is a shape matrix whose entries are the finite element shape functions and $\mathbf{q}_d$ is the vector of local deformation parameters [30].

We then define

$$\mathbf{q} = (\mathbf{q}_c^\top, \mathbf{q}_\theta^\top, \mathbf{q}_s^\top, \mathbf{q}_d^\top)^\top \tag{5}$$

(with $\mathbf{q}_c = \mathbf{c}$ and $\mathbf{q}_\theta = \boldsymbol{\theta}$), as the vector of generalized coordinates which consists of the model's parameters [21].

3.2.1 Dynamics and Generalized Forces

When fitting the model to visual data, our goal is to recover $\mathbf{q}$. We make the model dynamic based on the Lagrange equations of motion [21]. For static shape reconstruction problems, where we want the model to come to rest as soon as the external forces equilibrate or vanish, we set the mass density to zero [30]. Subsequently, the Lagrange equations of motion simplify to the following first-order system

$$\mathbf{D}\dot{\mathbf{q}} + \mathbf{K}\mathbf{q} = \mathbf{f}_q, \tag{6}$$

where $\mathbf{D}$ is the damping matrix and $\mathbf{K}$ is the stiffness matrix. Finally,

$$\mathbf{f}_q = \int \mathbf{L}^\top \mathbf{f} \, d\mathrm{u} \tag{7}$$

are generalized external forces associated with the components of $\mathbf{q}$, $\mathbf{f}(\mathrm{u}, t)$ is the force distribution applied to the model through our physics-based approach to visual estimation [30], and $\mathbf{L}$ is the Jacobian matrix that converts q-dimensional vectors to 3-D vectors [21].

4 Shape Recovery

Recovering a volumetric description from a range image consists of two steps. First, a qualitative 3-D volume is recovered from the range image. Next, the recovered qualitative volume is used to constrain the fitting of a deformable model to the range data. In this section, we describe each of these steps in greater detail.

4.1 Qualitative Shape Recovery

From an input range image, we apply Flynn's range image region segmentation algorithm [12], which is based on the algorithm of Hoffman and Jain [16]. First, surface normals are computed at each pixel in the image. Next, an initial surface segmentation is formed based on a clustering of the surface normals; similar adjacent patches are merged. Each surface is then classified according to the signs of its maximum and minimum curvatures. One problem with this method, as pointed out in Raja and Jain [27], is that since surfaces are segmented along orientation discontinuities, certain complex surfaces, like the body of the bent cylinder in Figure 1, do not have a unique surface classification. In the case of the bent cylinder, the classification is $(+,+)$ along the outside (where it bends out) and $(+,-)$ along the inside (where it bends in), where (S_{max}, S_{min}) represent the signs of the maximum and minimum curvatures, respectively.

The segmentation algorithm will classify the surface according to which surface type is dominant. Thus, it is essential that the range aspect hierarchy captures all possibilities of single classifications of surface type for the various surface types that make up the chosen volumes. In this case, the bent cylinder will have different aspects depending on which side of the bent cylinder is being viewed. Ideally, we should add a confidence measure to our surface classification. When confidence is low, we would simply ignore the surface type and match solely on the shapes of the face's bounding contours. Although this may introduce additional aspect hypotheses, they will be rank-ordered according to their supporting evidence.

The above surface segmentation and classification steps yield a 2-D region label image where a contiguous region represents a mask specifying which pixels in the original range image belong to the surface represented by the region. In addition, the signs of the maximum and minimum curvatures are encoded for each region. 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. The steps of partitioning the bounding contour and classifying the resulting contours are performed simultaneously using a minimal description length algorithm due to Li [18]. From a set of initial candidate contour breakpoints (derived from a polygonal approximation), the algorithm considers all possible groupings of the inter-breakpoint contours according to a minimum description length measure based on how well lines and elliptical arcs can be fit to the segment groups in terms of the cost of coding the various segments. The result is a *region boundary graph* representation for a region, in which nodes represent bounding contours, and arcs represent relations between the contours, including cotermination, parallelism, and symmetry.[2]

Once we have established a description of each image region which includes both its surface shape and boundary shape (of its 2-D projection), the next step is to match that description against the faces in the range aspect hierarchy using an interpretation tree search (Grimson and Lozano-Pérez [13]). Descriptions that exactly match (both surface shape and boundary) a face in the range aspect hierarchy will be given a single label with probability 1.0. For region boundary graphs that do not match due to occlusion or segmentation errors, we descend to an analysis at the boundary group level and match subgraphs of the region boundary graph along with the surface shape to the boundary groups in the range aspect hierarchy. Each subgraph that matches a boundary group generates a set of possible face interpretations (labels), each with a corresponding probability defined by the non-zero conditional probabilities mapping boundary groups to faces in the range aspect hierarchy. 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, each face in the face topology graph (recall that there may be many at each node) is used to infer a set of aspect hypotheses, using the non-zero conditional probabilities mapping faces to aspects in the range aspect hierarchy. Thus, at each node in the face topology graph, we have a set of aspect hypotheses that can account for that node. We state the problem of shape recovery as a partitioning of the nodes in the face topology graph into groups or clusters, each isomorphic to an aspect of a volumetric part. This can be solved by searching through the various labelings of the face topology graph nodes (choosing one

[2] See Dickinson et al. [9] for a discussion on how parallelism and symmetry are computed.

aspect label per node) until a complete covering of the image is achieved. This search is guided by a heuristic based on the conditional probabilities in the range aspect hierarchy [9, 8].

During the search process, aspect verification, like face matching, is accomplished through the use of an interpretation tree search (Grimson and Lozano-Pérez [13]). Once a set of aspects has been recovered, each aspect is used to infer one or more volume hypotheses based on the non-zero conditional probabilities mapping aspects to volumes in the range aspect hierarchy. This time, we search through the space of volume hypotheses until we find a set of volumes that are consistent with the objects in the database (Dickinson et al. [8]).

In an expected or top-down object recognition domain, in which we are searching for a particular object or part, we use the range aspect hierarchy as an attention mechanism to focus the search for an aspect at appropriate regions in the image. This technique was applied to the top-down recognition of multipart objects in Dickinson et al. [6]. 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 [6]. Finally, it should be noted that for either top-down or bottom up volume recovery, each recovered volume encodes the aspect in which it it viewed; the aspect, in turn, encodes the faces that were used in instantiating the aspect, while each face specifies those contours in the image used to instantiate the face.

4.2 Quantitative Shape Recovery

4.2.1 Simplified Numerical Simulation

Since we want to reconstruct the shape of objects from static visual data, we use (6). Equation (6) is discretized in material coordinates u using nodal finite element basis functions. We carry out the discretization by tessellating the surface of the model into linear triangular elements [21]. Furthermore, for fast interactive response, we employ a first-order Euler method to integrate (6).

4.2.2 Applied Forces

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 [30]. 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 (6), 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 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 (7) 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 (7) 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 regions used to infer the faces exert external global deformation forces on the models.

4.2.3 Model Initialization

One of the major limitations of previous deformable model fitting approaches is their dependence on model initialization and prior segmentation [31, 30, 25]. By first recovering a qualitative volume, we generate a number of strong constraints that are used in the deformable model recovery process. First, the volume's shape class can be used to immediately constrain the model's global deformation parameters before the fitting even begins. For example, if we have recovered a cylinder volume from one of its aspects, then we can initialize the deformable model to have, for example, zero bending, zero tapering, and square $x - z$ cross-sectional shape. Second, we know exactly which contours in the image data should exert forces on the model, since we know exactly which contours were used to recover the qualitative shape of the volume. Third, since the aspect hierarchy encodes a mapping between aspect faces and volume surfaces, our image correspondences between the recovered aspect faces and the projected model faces is explicitly given. Finally, the recovered aspect encodes the qualitative orientation of the volume allowing the orientation of the model to be further constrained. In the examples presented in the next section, however, this last constraint was not used in order to illustrate the lack of dependence of the technique on model initialization.

4.3 Summary of the Algorithm

Our approach to deformable model fitting can be summarized as follows:

1. Segment the range image into a set of homogeneous regions based on surface curvature.

2. Characterize the shapes of the regions and cluster them into local part-based qualitatively-defined aspects.

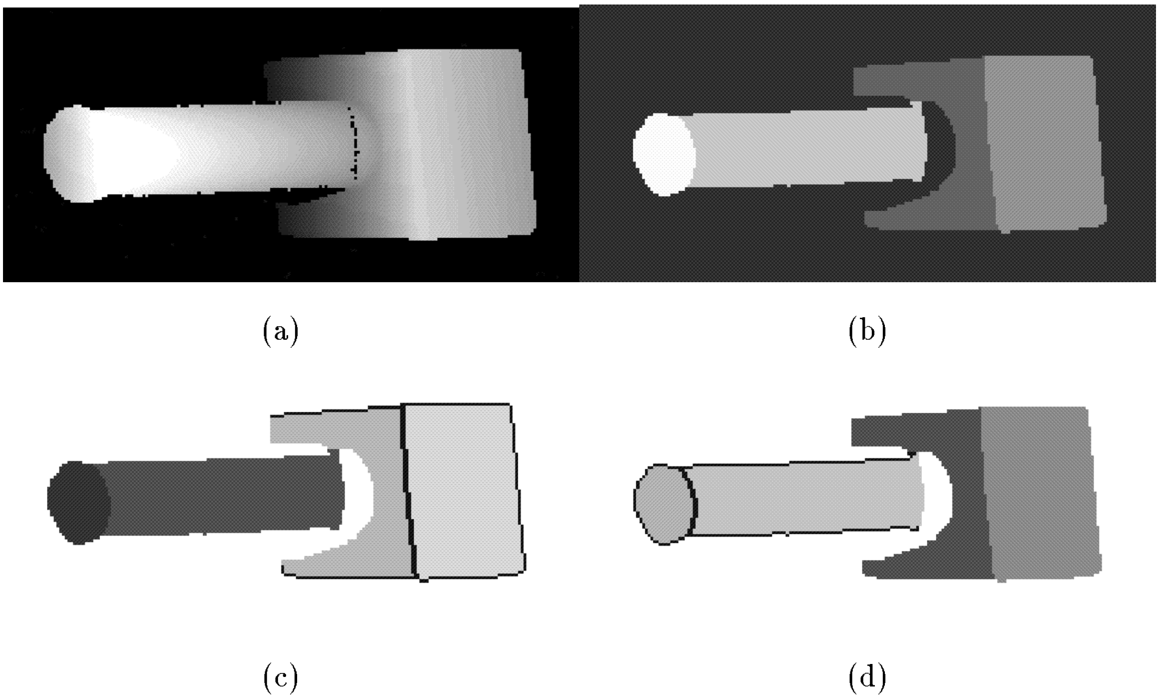
3. Use the information encoded in the qualitative aspects to initialize a deformable model for each part.

4. Fit each deformable model to the range data that correspond to the contours of the recovered aspects, using only global deformations.

5. Improve the fit of the deformable model by including the range data inside the region boundaries, and allow local deformations.

5 Results

We demonstrate our approach to the recovery of volumetric parts from range data by applying it to a set of images taken from Pat Flynn's industrial part range image database at Washington State University. To provide a step-by-step illustration of the approach, we



(a) (b)

(c) (d)

Figure 3: Qualitative Shape Recovery: (a) original range image, (b) region segmented image, (c) recovered qualitative block, (d) recovered qualitative cylinder.

first consider a range image of a scene containing an object consisting of two wooden parts, shown in Figure 3(a). The image was captured using a Technical Arts Scanner at the Michigan State University's PRIP Laboratory. In Figure 3(b), we show the results of applying Pat Flynn's region segmentation algorithm to the image. For this example, we invoked the expected object recognition mode to first search for the best instance of the block volume. Figure 3(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 3(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 Figure 4, we show a sequence of snapshots of the fitting process taking the initial model to its final shape describing the block. Similarly, in Figure 5, we show a sequence of snapshots of the fitting process taking the initial model to its final shape

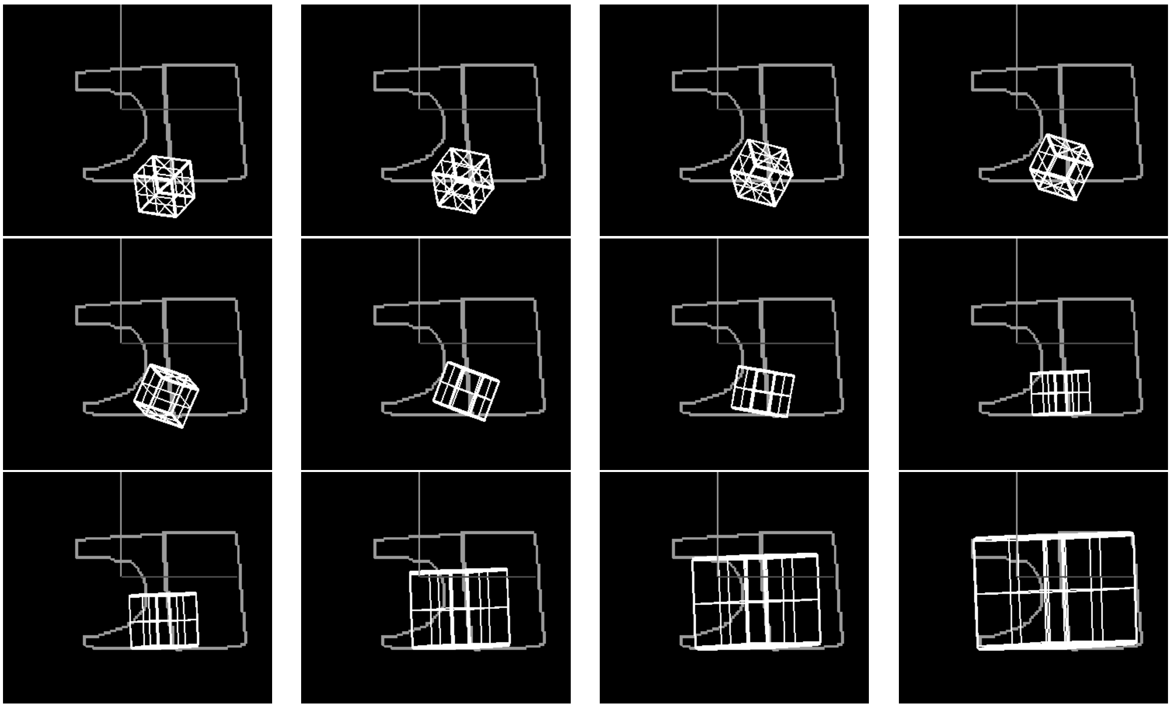


Figure 4: Selected Frames from Block Fitting Sequence (in order from left to right, top to bottom).

describing the cylinder. The bottom-right frame in Figure 5 shows a different view of the entire recovered object, showing the interconnection of the two parts. The system runs on an SGI Indigo 2EX. For this and all following examples, Flynn's region segmentation averages approximately 60 seconds, resulting in a region label image. Qualitative volume extraction averages approximately 10 seconds, while the quantitative volume recovery process (including real-time graphics display) also averages approximately 10 seconds. Clearly, processing time is dominated by the region segmentation step.

We now proceed to apply our technique to a set of synthetic range images containing industrial parts. In Figure 6, we show the results of recovering the volumetric parts belonging to an object composed of two cylinders ("adapter" in Flynn's database). The top row contains the original range images corresponding to four different views of the object, while the second row contains the corresponding region segmented images. The third row contains the fitted volumes, while the last row contains the fitted volumes shown from a different viewpoint. Note that in the images where the intersection of the two volumes is occluded, the fitted volumes do not intersect. This is due to the fact that the fitting process is constrained

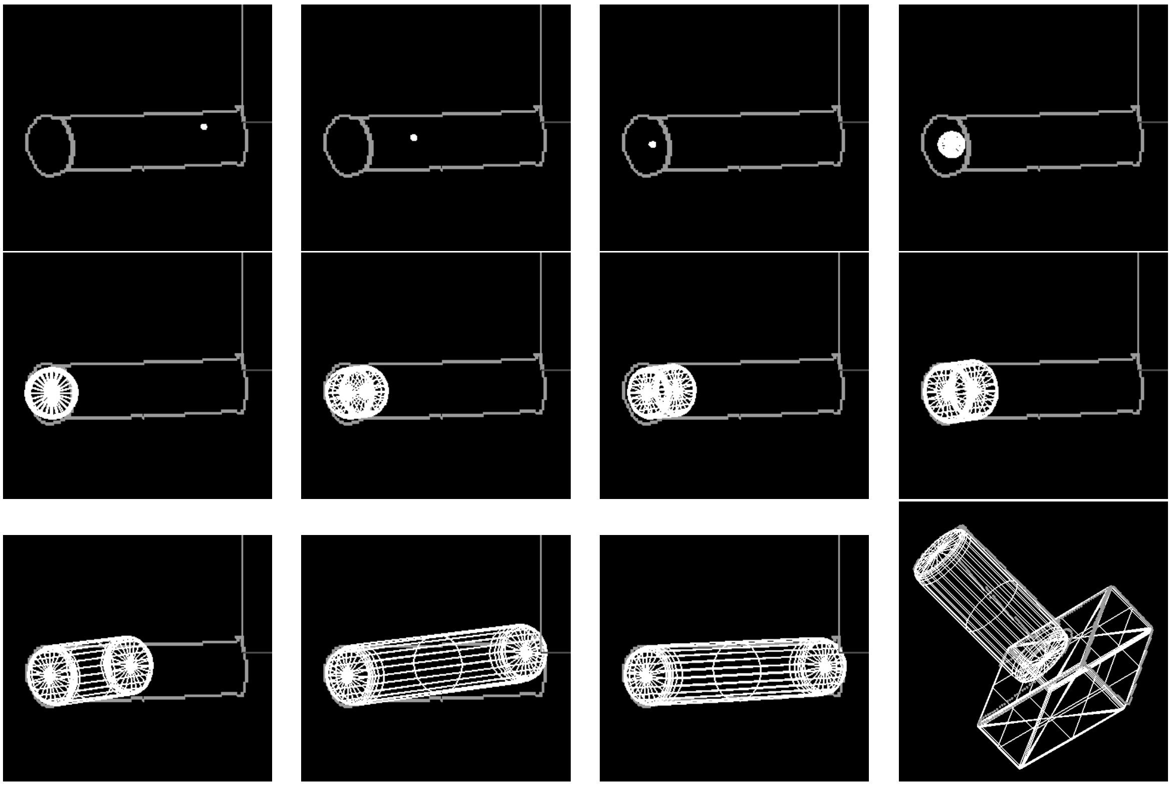


Figure 5: Selected Frames from Cylinder Fitting Sequence (in order from left to right, top to bottom). Note that the first three frames illustrate the solving for the $x - y$ translational degrees of freedom.

by the visible data. In the case of the leftmost view, the length of the smaller cylinder is underestimated due to the fact that part of its length is occluded in the image.

In Figure 7, we show the results of recovering the volumetric parts belonging to an object composed of a block and a cylinder ("column1" in Flynn's database). Note that in the third image, the body of the cylinder was imaged as background. Hence, when a cylinder is fit to the data, there is no data to fit the length of the cylinder and it remains as a disk.

In Figure 9, we show the results of recovering a set of volumetric parts from a sequence of real range images containing one- and two-part industrial objects. In the first row, we show the original images, while in the second row, we show the region segmented images. In the third row, we show the recovered volumes, while in the fourth row, we show the recovered volumes from a different viewpoint.

In Figure 8, we apply our approach to a synthetic range image consisting of a bent cylinder occluding a tapered cylinder. In Figure 8(a) and (b), we show the original range image and region segmented image, while in Figure 8(c) and (d), we show the recovered qualitative volumes. Finally, in Figure 8(e) and (f), we show the recovered 3-D models from two viewpoints.

In Figure 10, we demonstrate one of the limitations of our approach. We apply the technique to two images containing multiple occluded parts. Since we rely on having knowledge of the vocabulary of possible part classes that are visible in the image, the technique breaks down when shapes appear that are not included in the vocabulary. In the first example, we show a scene containing two objects. The object on the left is the familiar cylinder attached to a block, while the object on the right is an angled part composed of a block and a wedge. Two parts are recovered with high confidence (score). The first is the cylinder that is attached to the block. The block was not recovered with high confidence, since the only visible region other than the top face was discarded due to its relatively small size. For the angled object, only a portion of the block part was recovered, and only up to the shadow cast by the cylinder. In the second example, we show an image containing two objects, including a truncated, "L"-shaped object and the object composed of two cylinders. After region segmentation, we have lost all but the top face of the cylinder object. Since a high confidence aspect could not be recovered from the single, flat, elliptical surface, no volumes

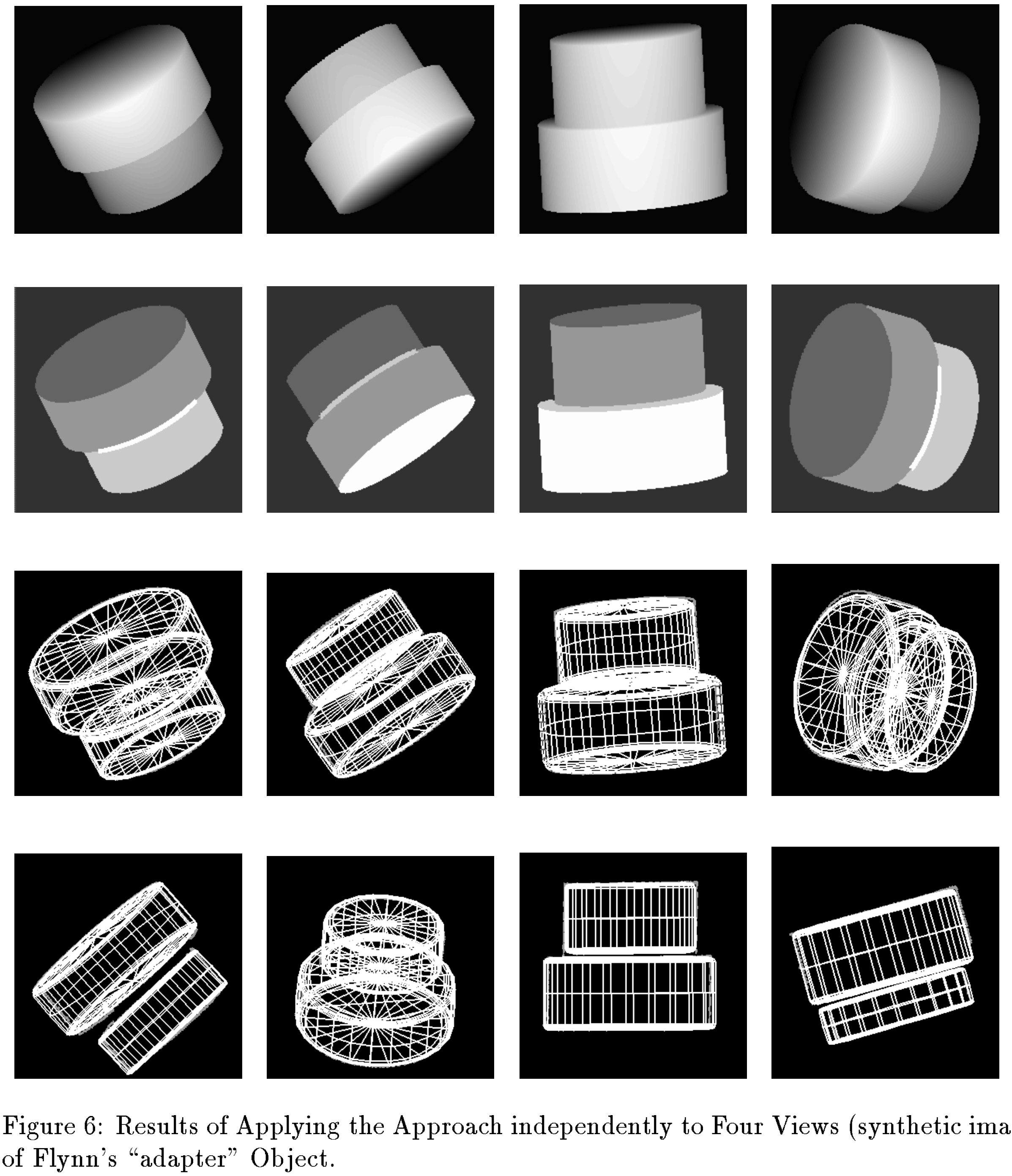


Figure 6: Results of Applying the Approach independently to Four Views (synthetic images) of Flynn's "adapter" Object.

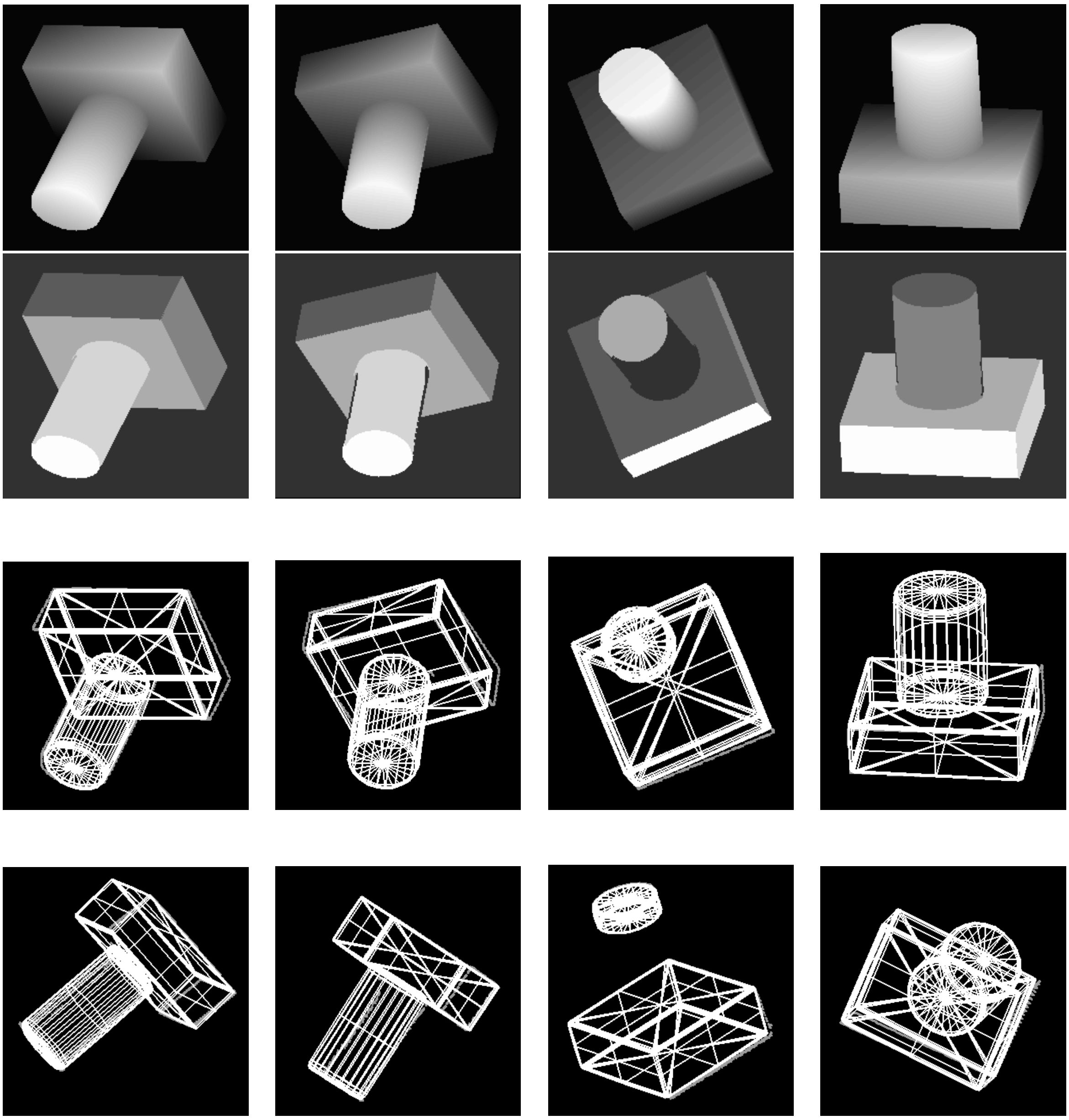
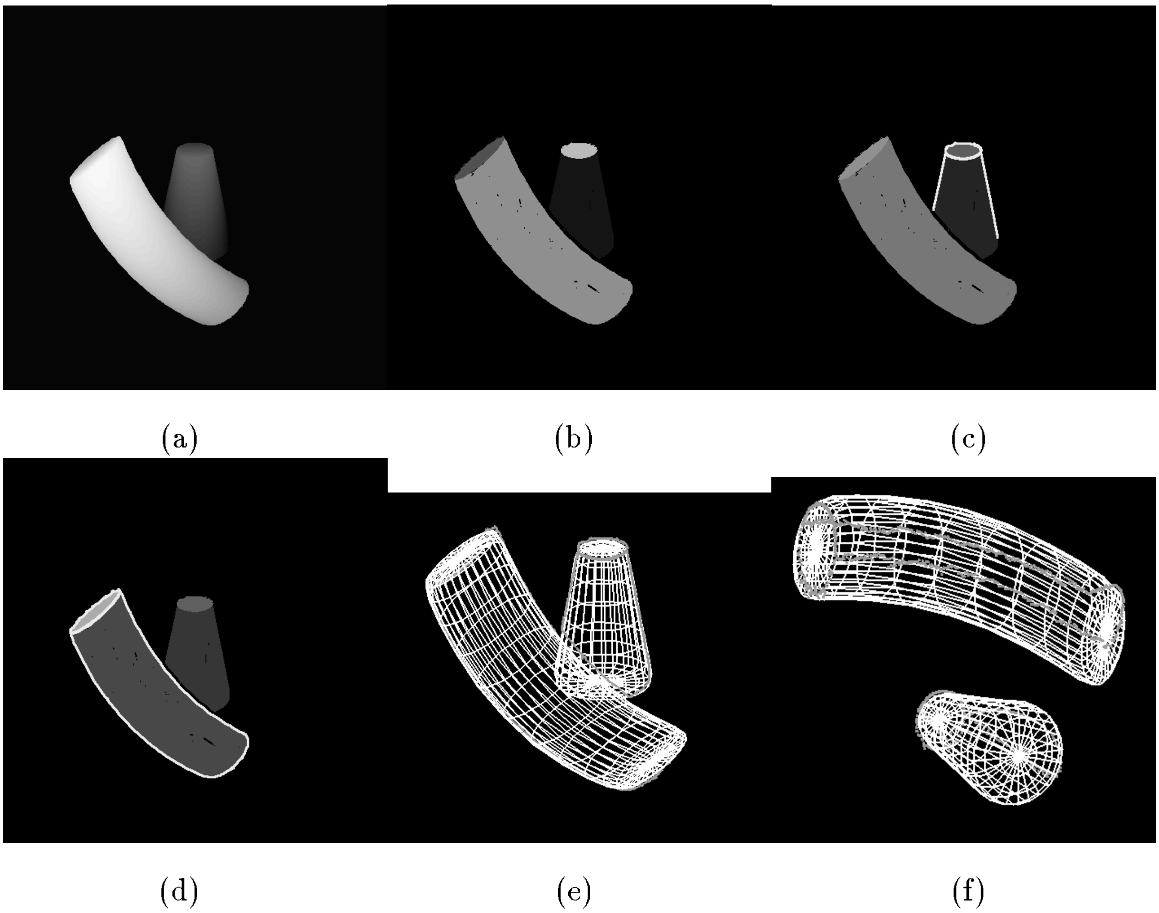


Figure 7: Results of Applying the Approach independently to Four Views (synthetic images) of Flynn's "column1" Object.



(a) (b) (c)

(d) (e) (f)

Figure 8: Results of Model Fitting to a Bent Cylinder Occluding a Tapered Cylinder: (a) original image, (b) region segmented image, (c) recovered tapered cylinder, (d) recovered bent cylinder, (e) fitted models from original viewpoint, (f) fitted models from novel viewpoint.

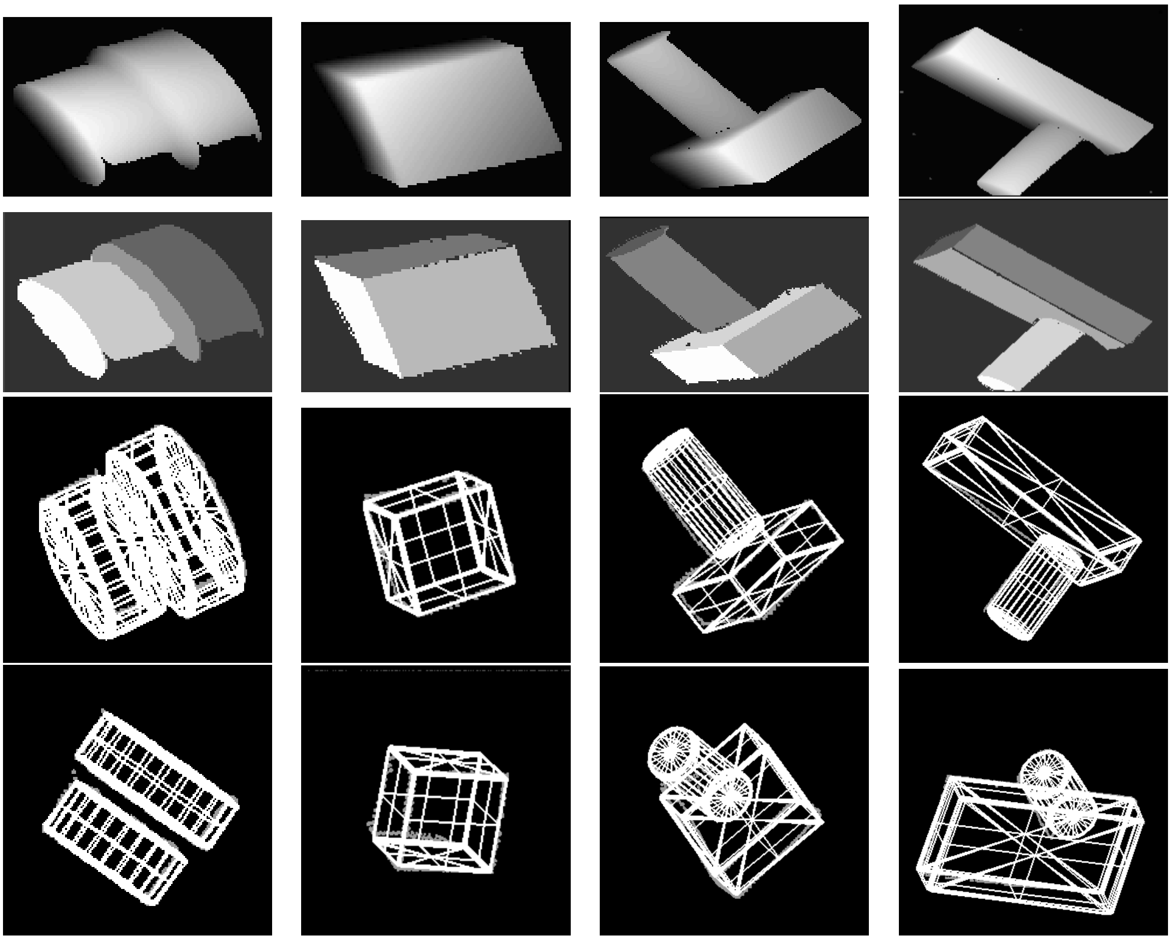


Figure 9: Results of Applying the Approach Independently to Four Views (real images) of Flynn's Industrial Parts.

was recovered for the cylinder object. For the "L"-shaped object, however, the part of the "L" that is consistent with our vocabulary (in this case, the block) was partially recovered.

We use both region segmentation and qualitative shape recovery to first partition the data into chunks, and group those chunks into parts. The resulting parts provide strong constraints on a deformable model fitting procedure that is insensitive to model initialization. The comparison to purely bottom-up recovery methods is clear. By having qualitative models, we encode the knowledge required to segment the scene into parts which, in turn, encode the knowledge required to control the model fitting process. This advantage comes at the expense of requiring that the objects in the scene be composed of parts drawn from our vocabulary. Good part recovery, therefore, is not possible for scenes containing unknown parts.

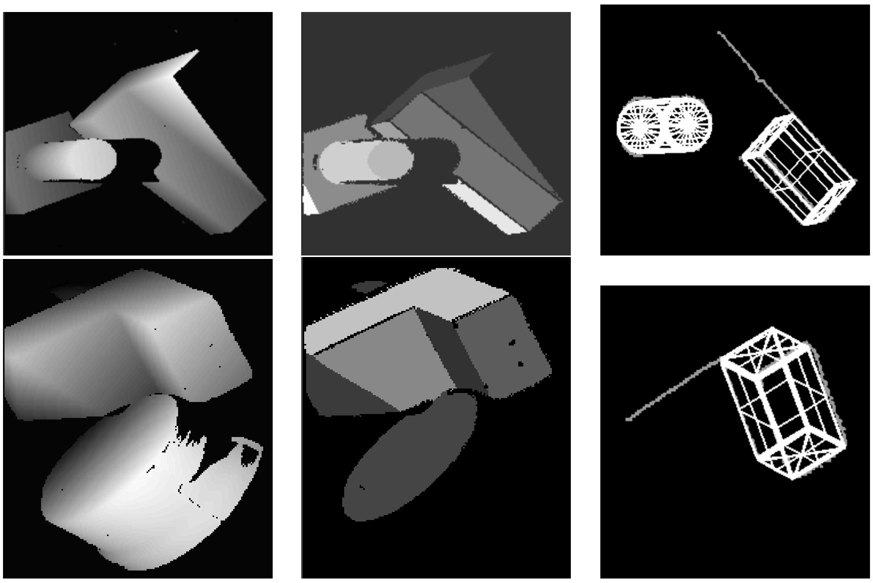


Figure 10: Results of Applying the Approach to Scenes Containing Object Parts Not Found in the Shape Vocabulary.

6 Limitations

The fitting constraints provided by the recovered qualitative shape ensure that the deformable model fitting process is invariant to initial position and orientation, as well as dimensions, degree of curvature, etc. 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. However, filtering out low-scoring volumes means that more regions in the image will be left uninterpreted. In fact, low-scoring volumes can be used to guide the sensor to acquire a higher-scoring volume [4]. We are currently looking at ways in which weaker qualitative information, short of a complete part aspect, can provided the needed constraints for fitting.

Both the qualitative and quantitative shape representation schemes are general. That is, the recovery scheme supports any set of qualitative volumetric shapes that can be mapped to a recoverable viewer-centered aspect hierarchy. Furthermore, any quantitative shape model that can be defined using our physics-based framework can be deformed by image forces. However, it is important to note that choosing one model will constrain the choice of the other, i.e., a quantitative shape model must be chosen such that it accurately models every possible instance of the qualitative shape model.

Finally, the step-by-step procedure for sequentially solving for a model's degrees of freedom during the fitting procedure has been specified for each volume. Ideally, such a procedure should be automated given the properties of the qualitative part. We are currently investigating methods to automate this procedure. Furthermore, we are studying which degrees of freedom can be coupled together during fitting while still maintaining insensitivity to initial size, position, and orientation of the model.

7 Conclusions

Traditional physics-based, deformable shape recovery techniques offer a completely data-driven approach to recovering an object's geometry. Image data points or features exert "forces" on a 3-D model to bring it (or its projection) into alignment with the image data. Unfortunately, much of the previous work on physics-based shape recovery has focused on model fitting at the expense of segmentation. In many cases, these techniques assume a manually segmented scene, or the absence of occlusion, or both. Furthermore, the recovery algorithms often assume a good initialization of the model both in terms of position and orientation or the fit may be incorrect.

We believe that the solution to the problem of recovering 3-D objects from range images lies in the middle ground between data-driven and model-driven approaches. In situations where the domain of objects is known, we propose a scheme whereby a set of local part-based views or aspects can bridge the gap between segmentation and model fitting. Segmenting the image into a set of simple, part-based qualitative view classes provides the needed constraints for physics-bases shape recovery to quickly and robustly converge on a set of volumetric parts.

8 Acknowledgements

The authors would like to thank Professor Pat Flynn at Washington State University for providing both the images and the region segmentation code. We are grateful To Meng Li for providing the minimum description length code for curve segmentation and description. The authors would also like to thank Gene Amdur at the University of Toronto for his support and assistance in the implementation of this work. Finally, we would especially like to thank Michael Chan at the University of Pennsylvania for his assistance in implementing these ideas.

References

[1] G. Agin and T. Binford. Computer description of curved objects. *IEEE Transactions on Computers*, C-25(4):439–449, 1976.

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

[3] T. Binford. Visual perception by computer. In *Proceedings, IEEE Conference on Systems and Control*, Miami, FL, 1971.

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

[5] S. Dickinson and D. Metaxas. Integrating qualitative and quantitative shape recovery. *International Journal of Computer Vision*, 13(3):1–20, 1994.

[6] S. Dickinson, D. Metaxas, and A. Pentland. Constrained recovery of deformable models from range data. In *Proceedings, 2nd International Workshop on Visual Form*, Capri, Italy, May 1994.

[7] S. Dickinson, A. Pentland, and A. Rosenfeld. A representation for qualitative 3-D object recognition integrating object-centered and viewer-centered models. In K. Leibovic, editor, *Vision: A Convergence of Disciplines.* Springer Verlag, New York, 1990.

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

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

[10] F. Ferrie, J. Lagarde, and P. Whaite. Recovery of volumetric descriptions from laser rangefinder images. In *Proceedings, ECCV '90*, pages 387–396, Antibes, France, April 1990.

[11] F. Ferrie, J. Lagarde, and P. Whaite. Darboux frames, snakes, and super-quadrics: Geometry from the bottom up. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(8):771–784, 1993.

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

[13] W. Grimson and T. Lozano-Pérez. Model-based recognition and localization from sparse range or tactile data. *International Journal of Robotics Research*, 3(3):3–35, 1984.

[14] 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.

[15] A. Gupta and R. Bajcsy. Surface and volumetric segmentation of range images using biquadrics and superquadrics. In *Proceedings, 11th IAPR International Conference on Pattern Recognition*, pages 158–162, The Hague, Netherlands, 1992.

[16] R. Hoffman and A. Jain. Segmentation and classification from range images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):608–620, September 1987.

[17] A. Leonardis, F. Solina, and A. Macerl. A direct recovery of superquadric models in range images using recover-and-select paradigm. In *Proceedings, Third European Conference on Computer Vision (Lecture Notes in Computer Science, Vol 800)*, pages 309–318, Stockholm, Sweden, May 1994. Springer-Verlag.

[18] M. Li. Minimum description length based 2-D shape description. Technical Report CVAP114, Computational Vision and Active Perception Lab, Royal Institute of Technology, Stockholm, Sweden, October 1992.

[19] D. Metaxas. Physics-based modeling of nonrigid objects for vision and graphics. *Ph.D. thesis, Dept. of Computer Science, Univ. of Toronto*, 1992.

[20] D. Metaxas and D. Terzopoulos. Constrained deformable superquadrics and nonrigid motion tracking. In *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*, pages 337–343, 1991.

[21] D. Metaxas and D. Terzopoulos. Shape and nonrigid motion estimation through physics-based synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):580–591, June 1993.

[22] R. Nevatia and T. Binford. Description and recognition of curved objects. *Artificial Intelligence*, 8:77–98, 1977.

[23] A. Pentland. Perceptual organization and the representation of natural form. *Artificial Intelligence*, 28:293–331, 1986.

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

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

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

[27] N. Raja and A. Jain. Obtaining generic parts from range images using a multi-view representation. *CVGIP:Image Understanding*, 60(1):44–64, July 1994.

[28] F. Solina. Shape recovery and segmentation with deformable part models. Technical Report MS-CIS-87-111, GRASP LAB 128, University of Pennsylvania, Philadelphia, PA, 1987.

[29] F. Solina and R. Bajcsy. Recovery of parametric models from range images: The case for superquadrics with global deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(2):131–146, 1990.

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

[31] D. Terzopoulos, A. Witkin, and M. Kass. Constraints on deformable models: Recovering 3d shape and nonrigid motion. *Artificial Intelligence*, 36:91–123, 1988.

[32] F. Ulupinar and R. Nevatia. Perception of 3-D surfa¡ces from 2-D contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:3–18, 1993.

[33] K. Wu and M. Levine. Recovering parametric geons from multiview range data. In *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*, pages 159–166, Seattle, WA, June 1994.

[34] M. Zerroug and R. Nevatia. Segmentation and recovery of shgcs from a real intensity image. In *Proceedings, Third European Conference on Computer Vision (Lecture Notes in Computer Science, Vol 800)*, pages 319–330, Stockholm, Sweden, May 1994. Springer-Verlag.