University of Pennsylvania
ScholarlyCommons

Center for Human Modeling and Simulation | Department of Computer & Information Science

1-1-1993

Integration of Quantitative and Qualitative Techniques for Deformable Model Fitting from Orthographic, Perspective, and Stereo Projections

Dimitris Metaxas
University of Pennsylvania, dnm@central.cis.upenn.edu

Sven J. Dickinson
University of Pennsylvania

Copyright 1993 IEEE. Reprinted from *Proceedings of the 4th International Conference on Computer Vision ICCV '93*, pages 641-649.

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Pennsylvania's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/hms/62
For more information, please contact repository@pobox.upenn.edu.

Integration of Quantitative and Qualitative Techniques for Deformable Model Fitting from Orthographic, Perspective, and Stereo Projections

Dimitri Metaxas
Department of CIS
University of Pennsylvania
Philadelphia, PA 19104-6389

Sven J. Dickinson*
Department of CS
University of Toronto
Toronto, Ontario, Canada M5S 1A4

Abstract

In this paper, we synthesize a new approach to 3-D object shape recovery by integrating qualitative shape recovery techniques and quantitative physics-based shape estimation techniques. Specifically, we first use qualitative shape recovery and recognition techniques to provide strong fitting constraints on physics-based deformable model recovery techniques. Secondly, we extend our previously developed technique of fitting deformable models to occluding image contours to the case of image data captured under general orthographic, perspective, and stereo projections.

1 Introduction

Since the introduction of a class of qualitatively-defined volumetric primitives, called *geons* [1], interest has been growing in building 3-D object recognition systems based on qualitative shape. One of the primary motivations in these systems is that, as stated by Biederman [1], the task of recognizing (or identifying) an object should be separated from the task of locating it, i.e., determining its pose. Furthermore, the exact shape of the object need not be recovered to facilitate recognition; a coarse-level description of an object in terms of its parts is not only sufficient to distinguish between different classes of objects, but provides an efficient indexing mechanism for recognition from large object databases. The above systems, however, address only the task of identifying the object. This is in contrast to classical 3-D object recognition systems, in which exact viewpoint is required to verify typically weak object hypotheses, while the object models capture the exact geometry of the object, e.g., [5, 8]. Determining the pose of the object is a critical component of these approaches.

Physics-based modeling [12, 17, 16, 9, 10] provides a very powerful mechanism for quantitatively modeling an object's shape for localization and/or subclass recognition. As opposed to a model-driven recovery process, in which image features are matched to a set of rigid, a priori object models which dictate the *exact* geometry of an object, deformable models offer a less constrained, data-driven recovery process. However, as powerful as these and other active, deformable model recovery techniques are, they have some serious limitations. Their success relies on both the accuracy of initial image segmentation and initial placement of the model given the segmented data. For example, such techniques often assume that the entire bounding contour of a region belongs to the object, a problem when the object is occluded. In addition, such techniques often require a manual segmentation of an object into parts. Clearly, a more robust recovery would require more knowledge of the object's position, orientation, and shape.

In this paper, we propose a two-step recovery process that first recovers the qualitative shape of an object in terms of its parts [3, 2]. If detailed shape or localization is needed to manipulate the object, for example, we then use knowledge of a part's qualitative shape and its orientation (encoded by its aspect) to provide strong constraints in fitting a deformable model to the part. Furthermore, since the qualitative shape recovery technique supports occlusion through a hierarchical aspect representation, it can selectively pass to the model fitting stage only those contours belonging to the object.

*S. Dickinson acknowledges the support of ITRC, IRIS, NSERC, and PRECARN Assoc., Canada.

0-8186-3870-2/93 $3.00 © 1993 IEEE

2 Related Work

Recently, several researchers have proposed various segmentation techniques to partition image or range data, in order to automate the process of fitting superquadric volumetric primitives to the data. Most of those approaches are applied to range data only [15, 4], while Pentland [11] describes a two-stage algorithm to fit superquadrics to image data. In the first stage, he segments the image using a filtering operation to produce a large set of potential object "parts", followed by a quadratic optimization procedure that searches among these part hypotheses to produce a maximum likelihood estimate of the image's part structure. In the second stage, he fits superquadrics to the segmented data using a least squares algorithm. Pentland's approach is only applicable in case of occluding boundary data under simple orthographic projection, as is true of earlier work of Terzopoulos et al. [17], Terzopoulos and Metaxas [16], and Pentland and Sclaroff [12], which address only the problem of model fitting. Taking a different approach, Raja and Jain [13] segment a range image into parts corresponding to geons, and then fit a superquadric to the part to determine geon orientation.

The fundamental difference between our approach and the above approaches is that we use a qualitative segmentation of the image to provide sufficient constraints on our deformable model fitting procedure. In addition, we generalize our deformable model fitting technique to accommodate orthographic, perspective, and stereo projections.

3 Object Modeling

3.1 Qualitative Shape Modeling

In this section, we briefly review the qualitative shape modeling technique described in [3, 2].

3.1.1 Object-Centered Models

Given a database of object models representing the domain of a recognition task, we seek a set of three-dimensional volumetric primitives that, when assembled together, can be used to construct the object models. Many 3-D object recognition systems have successfully employed 3-D volumetric primitives to construct objects. Commonly used classes of volumetric primitives include polyhedra, generalized cylinders, and superquadrics. Whichever set of volumetric modeling primitives is chosen, they will be mapped to a set of viewer-centered aspects.

To demonstrate our approach to object recognition, 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 primitives called geons. For our investigation, we have chosen three properties including cross-section shape, axis shape, and cross-section size variation (Dickinson et al. [3]). The values of these properties give rise to a set of ten primitives (a subset of Biederman's geons). To construct objects, the primitives are attached to one another with the restriction that any junction of two primitives involves exactly one distinct surface from each primitive.

3.1.2 Viewer-Centered Models

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 [6] Our approach differs in that we use aspects to represent a (typically small) set of volumetric primitives 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 primitives 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. The advantage of this approach is that since the number of qualitatively different primitives is generally small, the number of possible aspects is limited and, more important, *independent* of the number of objects in the database. The disadvantage is that if a primitive is occluded from a given 3-D viewpoint, its projected aspect in the image will also be occluded. Thus we must accommodate the matching of occluded aspects, which we accomplish by use of a hierarchical representation we call the *aspect hierarchy*.

The aspect hierarchy consists of three levels, consisting of the set of *aspects* that model the chosen primitives, the set of component *faces* of the aspects, and the set of *boundary groups* representing all subsets of contours bounding the faces. Fig. 1 illustrates a portion of the aspect hierarchy, along with a few of the primitives. 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 primitives. These conditional probabilities result from a statistical analysis of a set of images approximating the set of *all* views of *all* the primitives.

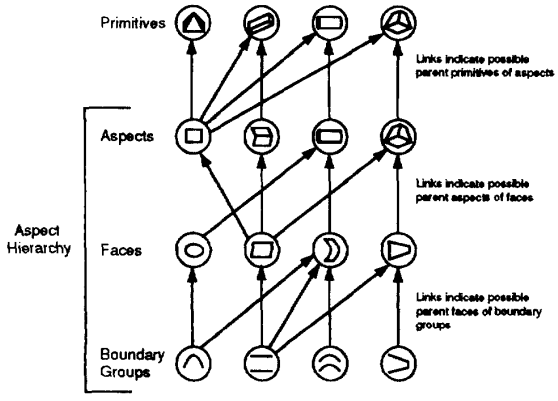

Figure 1: The Aspect Hierarchy

3.2 Quantitative Shape Modeling

In this section we first briefly review the general formulation of deformable models; further detail can be found in [16, 9]. We then extend the formulation to the case of orthographic, perspective, and stereo projections.

3.2.1 Geometry

Geometrically, the models used in this paper are closed surfaces in space whose intrinsic (material) coordinates are $\mathrm{u} = (u, v)$, defined on a domain Ω. The positions of points on the model relative to an inertial frame of reference Φ in space are given by a vector-valued, time-varying function of u: $\mathbf{x}(\mathrm{u}, t) = (x_1(\mathrm{u}, t), x_2(\mathrm{u}, t), x_3(\mathrm{u}, t))^\mathsf{T}$, where $^\mathsf{T}$ is the transpose operator. We set up a noninertial, model-centered reference frame ϕ [9] 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 $\mathbf{R}(t)$. 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 reference shape $\mathbf{s}(\mathrm{u}, t)$ and a displacement function $\mathbf{d}(\mathrm{u}, t)$:

$$\mathbf{p} = \mathbf{s} + \mathbf{d}. \tag{2}$$

Based on the shapes we want to recover, we consider the case of superquadric ellipsoids with linear tapering and bending global deformations [9] and express the reference shape as:

$$\mathbf{s} = \mathbf{T}(\mathbf{a}, \mathbf{b}), \tag{3}$$

where $\mathbf{T}$ is a vector function depending on the superquadric parameters $\mathbf{a}$ and the parameters $\mathbf{b}$ necessary for the definition of the linear tapering and bending deformations [9]. We collect the parameters in $\mathbf{s}$ into the global deformation parameter vector:

$$\mathbf{q}_s = (\mathbf{a}^\mathsf{T}, \mathbf{b}^\mathsf{T})^\mathsf{T}. \tag{4}$$

The above global deformation parameters are adequate for quantitatively describing the ten modeling primitives. We will therefore assume that $\mathbf{d} = \mathbf{0}$.

3.2.2 Kinematics and Dynamics

The velocity of points on the model is given by:

$$\dot{\mathbf{x}} = \dot{\mathbf{c}} + \mathbf{B}\dot{\boldsymbol{\theta}} + \mathbf{R}\dot{\mathbf{s}}, \tag{5}$$

where $\boldsymbol{\theta}$ is the vector of rotational coordinates of the model, and $\mathbf{B} = \partial(\mathbf{R}\mathbf{p})/\partial\boldsymbol{\theta}$. Furthermore, $\dot{\mathbf{s}} = \mathbf{J}\dot{\mathbf{q}}_s$, where $\mathbf{J}$ is the Jacobian of the deformable superquadric model with respect to the global degrees of freedom $\mathbf{q}_s$ [9]. We can therefore write:

$$\dot{\mathbf{x}} = [\mathbf{I}\ \mathbf{B}\ \mathbf{R}\mathbf{J}]\dot{\mathbf{q}} = \mathbf{L}\dot{\mathbf{q}}, \tag{6}$$

where $\mathbf{L}$ is the Jacobian of the superquadric model, $\mathbf{q} = (\mathbf{q}_c^\mathsf{T}, \mathbf{q}_\theta^\mathsf{T}, \mathbf{q}_s^\mathsf{T})^\mathsf{T}$, with $\mathbf{q}_c = \mathbf{c}$ and $\mathbf{q}_\theta = \boldsymbol{\theta}$.

When fitting the model to visual data, our goal is to recover $\mathbf{q}$, the vector of degrees of freedom of the model. Our approach carries out the coordinate fitting procedure in a physically-based way. We make our model dynamic in $\mathbf{q}$ by introducing mass, damping, and a deformation strain energy. This allows us, through the apparatus of Lagrangian dynamics, to arrive at a set of equations of motion governing the behavior of our model under the action of externally applied forces. In the absence of local deformations, the Lagrange equations of motion take the form [16]:

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{D}\dot{\mathbf{q}} = \mathbf{g}_q + \mathbf{f}_q, \tag{7}$$

where $\mathbf{M}$ and $\mathbf{D}$ are the mass and damping matrices, respectively, $\mathbf{g}_q$ are inertial forces arising from the dynamic coupling between the local and global degrees of freedom, and $\mathbf{f}_q(\mathrm{u}, t)$ are the generalized external forces associated with the degrees of freedom of the model. The generalized external forces will be discussed in detail in Section 4.2.2.

3.2.3 Orthographic Projection

In the case of orthographic projection, the points on the model $\mathbf{x} = (x, y, z)$ project to the image points x_p and y_p as follows:

$$x_p = x, \qquad y_p = y. \tag{8}$$

By taking the derivative of the above equation (8) with respect to time, we arrive at the following formulas:

$$\dot{x}_p = \dot{x}, \qquad \dot{y}_p = \dot{y}. \tag{9}$$

Rewriting (9) in matrix form and using (6), we arrive at the following matrix equations:

$$\begin{bmatrix} \dot{x}_p \\ \dot{y}_p \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{L}\dot{\mathbf{q}}. \tag{10}$$

If we rewrite (10) in compact form, we get

$$\begin{bmatrix} \dot{x}_p \\ \dot{y}_p \end{bmatrix} = \mathbf{L}_o\dot{\mathbf{q}}, \tag{11}$$

where

$$\mathbf{L}_o = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{L}. \tag{12}$$

3.2.4 Perspective Projection

In the case of perspective projection, points on the model $\mathbf{x} = (x, y, z)$ project into image points, x_p and y_p, based on the formula:

$$x_p = \frac{x}{z}f, \qquad y_p = \frac{y}{z}f, \tag{13}$$

where f is the focal length.

By taking the derivative of the above equation (13) with respect to time, we arrive at the following formulas:

$$\dot{x}_p = \dot{x}\frac{f}{z} - \frac{x}{z^2}f\dot{z}, \qquad \dot{y}_p = \dot{y}\frac{f}{z} - \frac{y}{z^2}f\dot{z}. \tag{14}$$

Rewriting (14) in matrix form and using (6), we arrive at the following matrix equations

$$\begin{bmatrix} \dot{x}_p \\ \dot{y}_p \end{bmatrix} = \begin{bmatrix} f/z & 0 & -x/z^2 f \\ 0 & f/z & -y/z^2 f \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = \tag{15}$$

$$\begin{bmatrix} f/z & 0 & -x/z^2 f \\ 0 & f/z & -y/z^2 f \end{bmatrix} \mathbf{L}\dot{\mathbf{q}}. \tag{16}$$

If we rewrite (16) in compact form, we get

$$\begin{bmatrix} \dot{x}_p \\ \dot{y}_p \end{bmatrix} = \mathbf{L}_p\dot{\mathbf{q}}, \tag{17}$$

where

$$\mathbf{L}_p = \begin{bmatrix} f/z & 0 & -x/z^2 f \\ 0 & f/z & -y/z^2 f \end{bmatrix} \mathbf{L}. \tag{18}$$

The above two Jacobian matrices, $\mathbf{L}_o$ and $\mathbf{L}_p$, will be used in the calculation of the generalized external forces $\mathbf{f}_q$ from two dimensional external forces $\mathbf{f}$ that the data exert on the model.

3.2.5 Stereo Projection

In the case of stereo projection, we assume two parallel cameras, each under perspective projection, resulting in two images, L and R. The model points $\mathbf{x}$ project on each of the images based on (13) and the corresponding Jacobian matrices $\mathbf{L}_{pL}$ and $\mathbf{L}_{pR}$ are calculated using (18).

To recover the exact location of the model frame $\mathbf{c}$, we apply the following procedure:

- We first independently fit the model to the left and right image data. This results in two model instances, m_L and m_R, one per image, having the same scale.
- Choosing one of the images, say R, we project the locations of the left and right model frames, m_{L_c} and m_{R_c}, into R. Let the locations of these projected model centers be $\mathbf{c}_L$ and $\mathbf{c}_R$ respectively.
- We then map the difference in the x coordinates[1] of $\mathbf{c}_L$ and $\mathbf{c}_R$ into a force that modifies $\mathbf{m}_L$ and $\mathbf{m}_R$ in the direction of $\mathbf{m}_L$ and $\mathbf{m}_R$, respectively, according to the following formula:

$$\dot{\mathbf{m}}_k = s|c_{Lx} - c_{Rx}|\frac{\mathbf{m}_k}{||\mathbf{m}_k||}, \tag{19}$$

  where $k = L$ or $k = R$, $s = 1$ if $c_{Lx} < c_{Rx}$, and $s = -1$ otherwise.
- Once $\mathbf{c}_L = \mathbf{c}_R$, we first sum the forces that the left and right image data exert on the model. From their sum, we then compute the generalized force f_{q_a} that corresponds to the scaling parameter a of the deformable model [16], and using (7), we modify a.

4 Shape Recovery

4.1 Qualitative Shape Recovery

Qualitative shape recovery consists of the following three steps, resulting in a graph representation of the image in which nodes represent recovered qualitative 3-D primitives, and arcs represent hypothesized connections between the primitives; details of the complete recovery process, including algorithms to handle various segmentation errors, can be found in Dickinson et al. [3]. In the following subsections, we briefly review the approach to recovering qualitative shape.

[1] Since the two cameras are parallel, the projections of the two model frame centers differ only in the x direction.

4.1.1 Face Recovery

The first step to recovering a set of faces is a region segmentation of the input image. We begin by applying Saint-Marc and Medioni's edge-preserving adaptive smoothing filter to the image [14], followed by a morphological gradient operator [7]. A hysteresis thresholding operation is then applied to produce a binary edge image from which a connected components analysis yields a set of regions. The resulting regions are captured in a *region topology graph* in which nodes represent regions and arcs specify region adjacency.

Next, the bounding contours of the regions are classified according to their shape. Each region is represented by a graph in which nodes represent bounding contours (parsed at curvature discontinuities), and arcs represent nonaccidental relations between the contours, e.g., parallelism, cotermination, and symmetry. A graph representing a given region is then compared to the faces in the aspect hierarchy (also represented as graphs). If a match occurs, a single face label with probability 1.0 is assigned to the image region. If, due to occlusion or segmentation errors, no match occurs, then subgraphs of the graph are matched to the boundary groups in the aspect hierarchy. Each matching boundary group can be used to infer one or more face hypotheses, each with a corresponding probability. The face labeling process results in a *face topology graph*, in which nodes represent image regions, and arcs represent region adjacencies. Furthermore, each node has one or more face labels associated with it.

4.1.2 Aspect Recovery

Given a graph representation of the faces in the image, there are two approaches to labeling, or recovering, aspects. In an unexpected object recognition task, we search for a complete and consistent covering of the face topology graph in terms of aspects [3]. Using the aspect hierarchy, each face label at each node in the face graph gives rise to a set of possible aspect hypotheses for that node. We search through the space of aspect labelings of the nodes in the face graph, and apply a heuristic based on the probabilities in the aspect hierarchy. In an expected, or top-down, object recognition task, we can use knowledge of the target object to constrain the search process [2].

4.1.3 Primitive Recovery

Given a recovered aspect, we can use the aspect hierarchy to generate a set of primitive hypotheses for that aspect, each with a corresponding probability. As in the case of aspect labeling, in an unexpected recognition framework, we search through the space of primitive labelings of the aspects in the image, while in an expected recognition framework, we use object knowledge to constrain the primitive interpretation of a given aspect [2].

In the case of stereo projection, we independently apply the qualitative shape recovery process to the left and right images. The correspondence problem then consists of matching qualitative primitive descriptions in the two images. To simplify this process, a pair of primitives represents a correspondence if: (i) the primitives have the same label, (ii) their aspects have the same label, and (iii) for each pair of corresponding faces in their aspects, there exists an epipolar line such that both faces intersect this line. Matching of qualitative primitives from multiple images is beyond the scope of this paper and will not be further discussed here.

4.2 Quantitative Shape Recovery

4.2.1 Simplified Numerical Simulation

In computer vision applications [16], we can simplify the equations while preserving useful dynamics by setting the mass density $\mu(\mathrm{u})$ to zero to obtain:

$$\mathbf{D}\dot{\mathbf{q}} = \mathbf{f}_q. \tag{20}$$

These equations yield a model which has no inertia and comes to rest as soon as all the applied forces vanish or equilibrate. Equation (20) 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. Furthermore, for fast interactive response, we employ a first-order Euler method to integrate (20).

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 [16]. We apply forces to the model based on differences between the model's projection in the image and the image data. Each of these forces corresponds to the appropriate generalized coordinate that has to be adapted so that the model fits the data. Given that our vocabulary of primitives is limited, we devise a systematic way of computing the generalized forces for each primitive. The computation depends on the influence of particular parts of the projected image on the model degrees of freedom. Such parts

correspond to the image faces (grouped to form an aspect) provided by the qualitative shape extraction. In the case of occluded primitives, 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.

For each of the three projection models, we compute the generalized forces $\mathbf{f}_q$ from 2D image forces f, using the following formula:

$$\mathbf{f}_q^\top = \int \mathbf{f}^\top \mathbf{L}_k \, du = (\mathbf{f}_{q_c}^\top, \mathbf{f}_{q_\theta}^\top, \mathbf{f}_{q_s}^\top), \tag{21}$$

where $k = o$ or $k = p$, depending on whether we assume orthographic or perspective projection, respectively. For orthographic projection, we assign forces from image data points to points on the model that lie on a particular region of the model defined by the qualitative shape recovery. For the case of perspective projection, we assign forces from image data points to points on the model that, in addition to satisfying the above property, are near occluding boundaries, thus satisfying the following formula:

$$|\mathbf{i} \cdot \mathbf{n}| < \tau, \tag{22}$$

where $\mathbf{n}$ is the unit normal at any model point, $\mathbf{i}$ is the unit vector from the focal point to a point on the model, and τ is a small threshold.

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 [17, 16, 12]. Using the qualitative shape recovery process as a front end, we first segment the image into parts, and for each part, we identify the relevant non-occluded contour data belonging to the part. In addition, the extracted qualitative primitives explicitly define a mapping between the image faces in their projected aspects and the 3-D surfaces on the quantitative models. Finally, the aspect that a primitive encodes defines a qualitative orientation that is exploited during model fitting, as will be demonstrated in Section 5.

5 Experiments

To illustrate the shape recovery approach, consider the real image of a toy table lamp, as shown in Fig. 2; the results of the bottom-up qualitative shape recovery algorithm are also shown in Fig. 2. At the top, the image window contains the contours extracted from the image, along with the face numbers. To the left is a

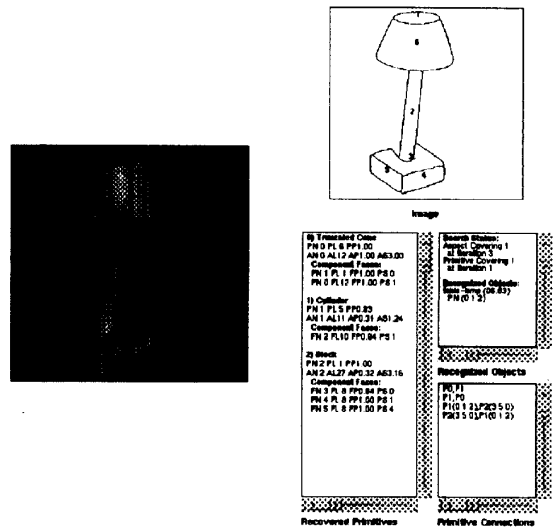

Figure 2: Original Image and Recovered Qualitative Primitives

window describing the recovered primitives (primitive covering). The mnemonics PN, PL, PP, and PS, refer to primitive number (simply an enumeration of the primitives in the covering), primitive label (see [3]), and primitive probability, respectively. The mnemonics AN, AL, AP, and AS refer to the aspect number (an enumeration), aspect label (see [3]), aspect probability, and aspect score (how well aspect was verified), respectively. The mnemonics FN, FL, FP, and PS refer to face number (in image window), face label (see [3]), face probability, and corresponding primitive attachment surface (see [3]), respectively, for each component face of the aspect.

To illustrate the fitting stage, consider the contours belonging to the lamp shade (truncated cone). Having determined during the qualitative shape recovery stage that we are trying to fit a deformable superquadric to a truncated cone, we can immediately fix some of the parameters in the model. In addition, the qualitative shape recovery stage provides us with a mapping between faces in the image and physical surfaces on the model. For example, we know that the elliptical face (FN 1) maps to the top of the truncated cone, while the body face (FN 0) maps to the side of the truncated cone. For the case of the truncated cone, we will begin with a cylinder model (superquadric) and will compute the forces that will deform the cylinder into the truncated cone appearing in the image. Assuming an *orthographic* projection and that the x and y dimensions are equal, we compute the following forces:

1. The cylinder is initially oriented with its z axis

orthogonal to the image plane. The first step involves computing the centroid of the elliptical image face (known to correspond to the top of the cylinder). The distance between the centroid and the projected center of the cylinder top is converted to a force which translates the model cylinder. Fig. 3(a) shows the image contours corresponding to the lamp shade and the cylinder following application of this force. Fig. 3(b) shows a different view of the image plane, providing a better view of the model cylinder.

2. The distance between the two image points corresponding to the extrema of the principal axis of the elliptical image face and two points that lie on a diameter of the top of the cylinder is converted to a force affecting the x and y dimensions with respect to the model cylinder. Figs. 3(c) and 3(d) show the image and the cylinder following application of this force.

3. The distance between the projected model contour corresponding to the top of the cylinder and the elliptical image face corresponds to a force affecting the orientation of the cylinder. Figs. 3(e) and 3(f) show the image and the cylinder following application of this force. This concludes the application of forces arising from the elliptical image face, i.e., top of the truncated cone.

4. Next, we focus on the image face corresponding to the body of the truncated cone to complete the fitting process. The distance between the points along the bottom rim of the body face and the projected bottom rim of the cylinder corresponds to a force affecting the length of the cylinder in the z direction. Figs. 3(g) and 3(h) show the image and the cylinder following application of this force.

5. Finally, the distance between points on the sides of the body face and the sides of the cylinder corresponds to a force which tapers the cylinder to complete the fit. Figs. 3(i) and 3(j) show the image and the tapered cylinder following application of this force. The result of fitting all three parts of the lamp is shown in Figs. 4 and 5.

For the case of perspective projection, we apply our shape recovery technique to the image in Fig. 6. A top-down search for the best three instances of a qualitative block primitive yields the three primitives shown in Fig. 7. Note that due to a large shadow edge that resulted in the undersegmented regions 12 on the triangular face of the wedge, the shape was misclassified as a block since region 12 was classified as having

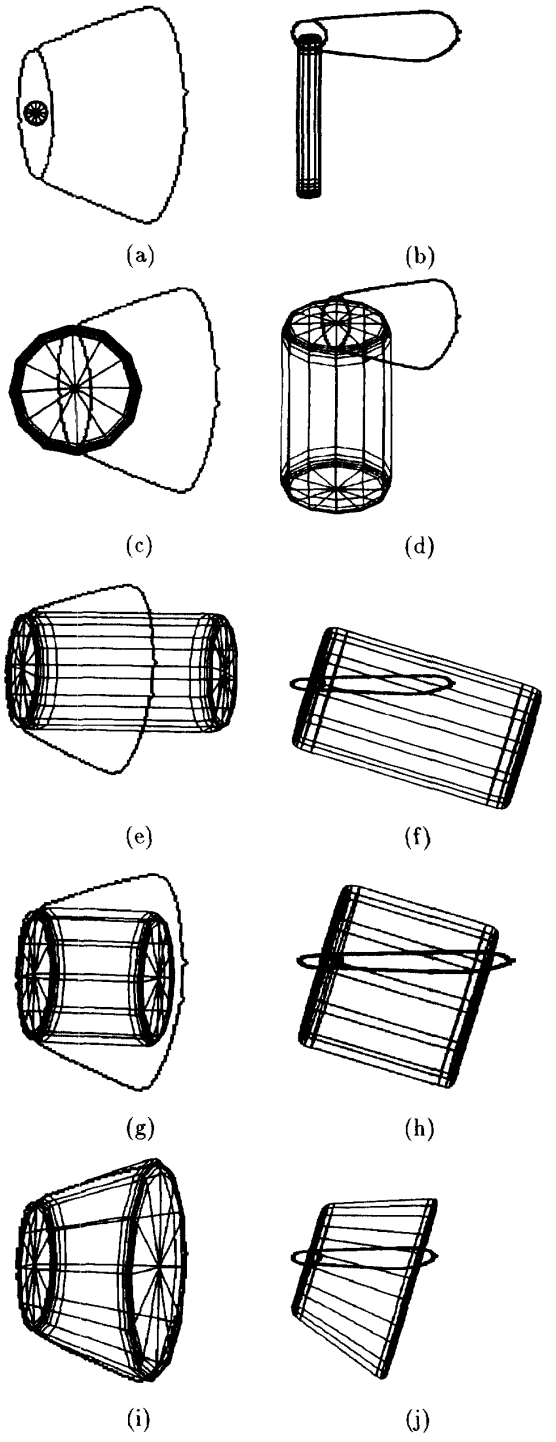

Figure 3: Quantitative Shape Recovery for Lamp Shade

Figure 4: Front View of Final Recovery of Table Lamp (Note that depth information is lost in orthographic projection.)

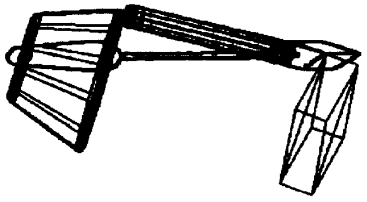

Figure 5: Side View of Final Recovery of Table Lamp (Note that depth information is lost in orthographic projection.)

opposites sides parallel. If we apply the quantitative recovery process to these three blocks, we obtain the models depicted in Figures 8.

Finally, we apply the shape recovery technique to the stereo pair shown in Fig. 9. The results of the qualitative shape recovery are shown in Fig. 10. Following the scaling step, the projection of the final model into the two images is shown in Fig. 11.

6 Conclusion

In this paper, we presented a new approach to 3-D object shape estimation based on the idea that the processes of recognizing an object and locating it are decoupled, and that recognition *does not* require accurate localization. The qualitative shape recovery component of the approach captures the coarse shape of objects composed of volumetric primitives *without* solving for exact viewpoint and *without* a precise geo-

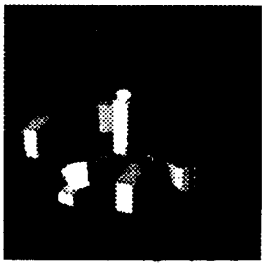

Figure 6: Image of Blocks on a Table

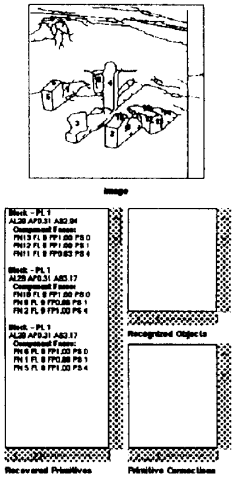

Figure 7: The Best Three Instances of a Qualitative Block

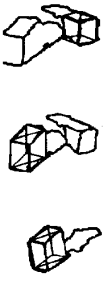

Figure 8: Models Fitted to Three Blocks

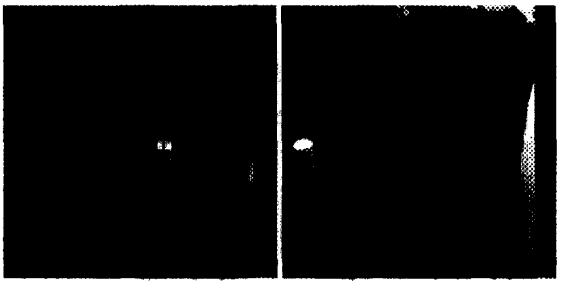

Figure 9: Left and Right Stereo Images of a Cylinder

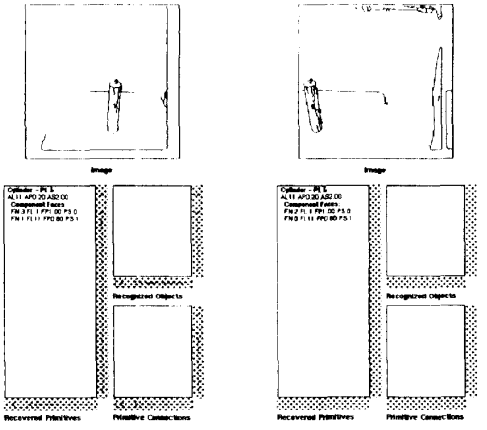

Figure 10: Qualitative Shape Recovery of Left and Right Images

metric verification of image features, which is sufficient for many tasks. If, however, we need to accurately locate (in order to manipulate) the object once it's been identified, or we need to extract a more detailed shape description in order to distinguish between subclasses of an object, then we can apply the quantitative shape recovery component using the constraints provided by the qualitative shape recovery component.

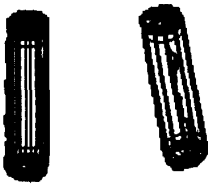

Figure 11: Unifying the Left and Right Models to Determine Scale and Depth

References

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

[2] S. Dickinson, A. Pentland, and A. Rosenfeld. From volumes to views: An approach to 3-D object recognition. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 55(2), 1992.

[3] 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.

[4] 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.

[5] D. Huttenlocher. Three-dimensional recognition of solid objects from a two-dimensional image. Technical Report 1045, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1988.

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

[7] J. Lee, R. Haralick, and L. Shapiro. Morphologic edge detection. *IEEE Journal of Robotics and Automation*, RA-3(2):142–155, 1987.

[8] D. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers, Norwell, MA, 1985.

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

[10] D. Metaxas and D. Terzopoulos. Shape and nonrigid motion estimation through physics-based synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, in press*, 1993.

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

[12] 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.

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

[14] P. Saint-Marc and G. Medioni. Adaptive smoothing: A general tool for early vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):514–529, 1991.

[15] 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.

[16] 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.

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