The Institute For Research In Cognitive Science

Physics-Based Object Pose and Shape Estimation from Multiple Views

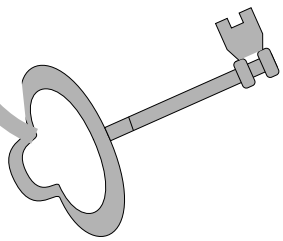
by

Michael Chan
Dimitri Metaxas

University of Pennsylvania
3401 Walnut Street, Suite 400C
Philadelphia, PA 19104-6228

November 1994

Site of the NSF Science and Technology Center for Research in Cognitive Science



University of Pennsylvania
Founded by Benjamin Franklin in 1740

IRCS Report 94-23

To appear in Proc. of the 12th International Conference on Pattern Recognition 1994.

Physics-Based Object Pose and Shape Estimation from Multiple Views

Michael Chan and Dimitri Metaxas
Dept. of Computer and Information Science
University of Pennsylvania, Philadelphia, PA 19104, U.S.A.

Abstract

This paper presents a new algorithm for object pose and shape estimation from multiple views. Using a qualitative shape recovery scheme we first segment the image into parts which belong to a vocabulary of primitives. Based on the additional constraints provided by the qualitative shapes we extend our physics-based framework to allow object pose and shape estimation from stereo images where the two cameras have arbitrary relative orientations. We then generalize our algorithm to integrate measurements from multiple views. To recover more complex objects we generalize the definition for the global bending deformation. We also present an algorithm for model discretization which evenly tessellates the model surface. We demonstrate the usefulness of our technique in experiments involving real images from of a variety of object shapes which may be partially occluded.

1 Introduction

The performance of most physics-based shape estimation techniques depends on the accuracy of the initial segmentation and the initial placement of the model given the segmented data [9, 6, 7]. In order to address the above limitations, a new approach to shape recovery from 2D images which integrates qualitative shape recovery [3] and quantitative physics-based estimation techniques [5] has been recently proposed [4]. Through qualitative shape recovery we can extract the qualitative shapes of objects that are composed of primitives which belong to a fixed vocabulary of shapes based on aspect matching. Furthermore, the qualitative shape recovery handles occlusion through a hierarchical aspect representation. We then use the qualitative correspondence of edge segments in the images and related contours on the models to provide strong fitting constraints to our physics-based estimation technique [5]. The "tokens" for matching features in the images are groups of edge segments (curved or straight) corresponding to an aspect of a part (as opposed to points or lines). Since the qualitative and quantitative shape recovery processes are both model-based, the approach is robust to noise and occlusion.

In this paper, extending the paradigm introduced in [4], we present a new algorithm for shape estimation from multiple views, including stereo as the simpler case. This estimation process can handle more complex scenarios where significantly different or incomplete (but consistent) sets of edge segments from the same object are extracted from multiple images taken from cameras with arbitrary relative orientations. Moreover, to be able to model and recover a larger variety of complex objects we propose a new definition for the global bending deformation, the parameters of which can be decoupled during recovery. We also present a new efficient algorithm which approximately tessellates the model surface uniformly in the 3D Euclidean space, allowing a more robust recovery as the model deforms to fit the data.

2 Dynamic deformable models

In this section, we review and extend the deformable models definition developed in [5].

2.1 Model kinematics

The positions of points on the model relative to a world coordinate frame of reference $\mathbf{\Phi}$ are $\mathbf{x}(\mathrm{u},t) = (x(\mathrm{u},t), y(\mathrm{u},t), z(\mathrm{u},t))^T$, where ^{T} denotes transposition and u are the model's material coordinates. We express the position of a point as $\mathbf{x} = \mathbf{c} + \mathbf{Rp}$, where $\mathbf{c}(t)$ is the origin of model reference frame ϕ located at the center of the model, $\mathbf{R}(t)$ is the rotation matrix that gives the orientation of ϕ relative to $\mathbf{\Phi}$ and $\mathbf{p}(\mathrm{u},t)$ gives the positions of points on the model relative to ϕ. We further write $\mathbf{p} = \mathbf{s} + \mathbf{d}$, as the sum of a reference shape $\mathbf{s}(\mathrm{u},t)$ and a displacement $\mathbf{d}(\mathrm{u},t)$. We express the reference shape as $\mathbf{s} = \mathbf{T}(\mathbf{e}(\mathrm{u};\, a_0, a_1, \ldots); b_0, b_1, \ldots)$, where $\mathbf{T}$ defines a *global deformation* (depending on the parameters $b_i(t)$), which transforms a geometric primitive $\mathbf{e}$ defined parametrically in u and parameterized by the variables $a_i(t)$.

We concatenate the global deformation parameters into the vector $\mathbf{q}_s = (a_0, a_1, \ldots, b_0, b_1, \ldots)^T$.

To illustrate our approach in this paper, we use a deformable superquadric ellipsoid [8] with global bending deformation as a reference shape. In the following section we define a bending deformation that ensures constant curvature along the major axis of bending.

2.2 Constant curvature bending

We define the bending transformation $\mathbf{s} = \mathbf{T}_b(\mathbf{e}; b_0)$ about the y-axis of a primitive $\mathbf{e}$ as

$$\begin{aligned} s_x &= cos(b_0(e_z - e_{z_0})) \cdot (e_x - \frac{1}{b_0}) + \frac{1}{b_0}, \\ s_y &= e_y, \\ s_z &= -sin(b_0(e_z - e_{z_0})) \cdot (e_x - \frac{1}{b_0}) + e_{z_0}, \end{aligned} \tag{1}$$

where b_0 is the radius of curvature, and e_{z_0} an arbitrary offset of the center of bending along the z-axis.

We generalize this deformation to deal with bending about an arbitrary axis in the xy-plane by introducing the following additional transformation

$$\mathbf{s} = \mathbf{J}_b{}^{-1}(\mathbf{T}_b(\mathbf{J}_b \cdot \mathbf{e})), \tag{2}$$

where $\mathbf{T}_b$ is defined above and matrix $\mathbf{J}_b$ is defined as

$$\mathbf{J}_b = \begin{bmatrix} cos(b_1) & -sin(b_1) & 0 \\ sin(b_1) & cos(b_1) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3}$$

Here, b_1 is the angle between the y-axis and the axis about which bending takes place. The introduction of $\mathbf{J}_b$ allows us to decouple the recovery of the rotation and bending parameters during model fitting. We can also define a piecewise bending deformation which we will demonstrate in the experiment section.

2.3 Repositioning the model frame ϕ

We move the model reference frame from its center to a different point on the model surface depending on which qualitative shape we want to recover as described in section 5.1. For a cylindrical primitive, for example, we place the model reference frame at the center of the circular end. For a rectangular primitive, we place it at one of its corner instead. These choices are made so that the recovery of the rotation and deformation parameters can be decoupled.

2.4 Dynamics and generalized forces

The velocity of points on the model is given by,

$$\dot{\mathbf{x}} = \mathbf{L}\dot{\mathbf{q}}, \tag{4}$$

where $\mathbf{L}$ is the Jacobian matrix [5] and $\mathbf{q} = (\mathbf{q}_c^T, \mathbf{q}_\theta^T, \mathbf{q}_s^T, \mathbf{q}_d^T)^T$ are the generalized coordinates of the model. Here $\mathbf{q}_c = \mathbf{c}$, $\mathbf{q}_\theta$ is the model's rotational degrees of freedom (expressed as a quaternion), $\mathbf{q}_s$ and $\mathbf{q}_d$ [1] represent the global and local deformations respectively. >From Lagrangian mechanics, after appropriate simplifications justified for static shape reconstruction [5], the dynamic equations of motion of our model take the form

$$\mathbf{D}\dot{\mathbf{q}} + \mathbf{K}\mathbf{q} = \mathbf{f}_q, \quad \mathbf{f}_q = \int \mathbf{L}^T \mathbf{f}\, d\mathrm{u}, \tag{5}$$

where $\mathbf{D}$ is the damping matrix, $\mathbf{K}$ is the stiffness matrix, and $\mathbf{f}_q$ are generalized external forces computed from the force distribution $\mathbf{f}(\mathrm{u})$ derived from the visual data. These forces are applied to the model through our physics-based approach to visual estimation [5].

3 Jacobian computation for generalized perspective projection

To allow shape and pose estimation in a world coordinate frame from images taken from a camera with a different frame of reference, the Jacobian matrix $\mathbf{L}$ used in (5) needs to be modified appropriately.

Let $\mathbf{x} = (x, y, z)^T$ denote the location of a point j w.r.t the world coordinate frame. Then we can write

$$\mathbf{x} = \mathbf{c}_c + \mathbf{R}_c \mathbf{x}_c, \tag{6}$$

where $\mathbf{c}_c$ and $\mathbf{R}_c$ are respectively the translation and rotation of the camera frame w.r.t. the world coordinate frame, and $\mathbf{x}_c = (x_c, y_c, z_c)^T$ is the position of the point j w.r.t to the camera coordinate frame.

Under perspective projection, the point $\mathbf{x}_c$ projects into an image point $\mathbf{x}_p = (x_p, x_p)^T$ according to

$$x_p = \frac{x_c}{z_c} f, \quad y_p = \frac{y_c}{z_c} f, \tag{7}$$

where f is the focal length of the camera. By taking the time derivative of (7) we get $\dot{\mathbf{x}}_p = \mathbf{H}\dot{\mathbf{x}}_c$ where

$$\mathbf{H} = \begin{bmatrix} f/z_c & 0 & -x_c/z_c^2 f \\ 0 & f/z_c & -y_c/z_c^2 f \end{bmatrix}. \tag{8}$$

Based on (4), (6) and (8), we obtain

$$\mathbf{x}_p = \mathbf{H}(\mathbf{R}_c^{-1}\dot{\mathbf{x}}) = \mathbf{H}\mathbf{R}_c^{-1}(\mathbf{L}\dot{\mathbf{q}}) = \mathbf{L}_p\dot{\mathbf{q}}. \tag{9}$$

By replacing the Jacobian matrix $\mathbf{L}$ in (5) by $\mathbf{L}_p = \mathbf{H}\mathbf{R}_c^{-1}\mathbf{L}$, two dimensional image forces $\mathbf{f}$ can be appropriately converted into generalized forces $\mathbf{f}_q$ measured in the world coordinate frame.

[1] With $\mathbf{d} = \mathbf{S}\mathbf{q}_d$ where $\mathbf{S}$ is the shape matrix.

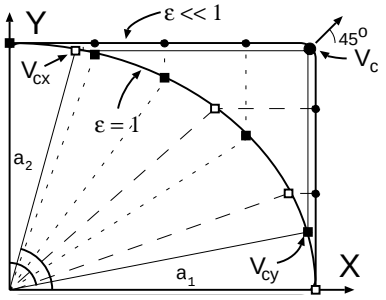


Figure 1: Illustration of the resampling scheme in the v-coordinate of the surface mesh

4 Model discretization

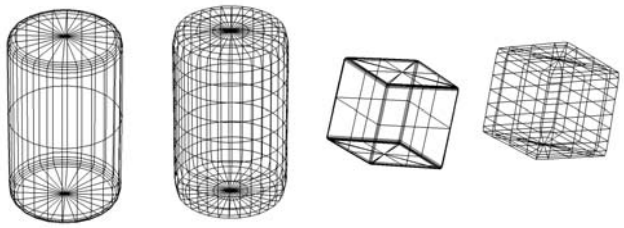
To ensure that the forces are evenly applied to the 3D model surface during model fitting, we need to have a uniform discretization of the model surface in the 3D Euclidean space. Here we present a technique to discretize a superquadric surface dynamically that maintains approximate uniform tessellation. We focus on the counteracting effect of the squareness parameters ϵ_1 and ϵ_2 which causes the nonuniform tessellation of the model surface when their values approach zero. Although the material coordinates u $= (u, v)$ of the superquadric surface are defined for $-\frac{\pi}{2} < u < \frac{\pi}{2}$ and $0 < v < 2\pi$, it suffices due to symmetry to illustrate our method for the uniform tessellation of the first quadrant of the xy plane, with corresponding material coordinate $v \in [0, \frac{1}{2}\pi]$.

The points generated by uniform sampling in the material coordinate space are quite uniformly distributed on the model surface when $\epsilon_2 = 1$. By projecting these model points along the x and y axes onto a model surface where $\epsilon_2 << 1$, their positions remain relatively "spread out" for $0 < v < v_{cx}$ and $v_{cy} < v < \frac{\pi}{2}$, respectively. The values of v_{cx} and v_{cy} are obtained by projecting along the axes the point on the model surface in the xy plane where the normal vector to the surface is at 45^o from each axis. Mathematically, these are given by $v_{cx} = sin^{-1}(sin^{\epsilon_2} v_c), v_{cy} = cos^{-1}(cos^{\epsilon_2} v_c)$ where $v_c = tan^{-1}[(\frac{a_2}{a_1})^{\frac{1}{\epsilon_2 - 2}}]$. These are illustrated in Fig. 1.

Based on the above, instead of uniformly sampling v in its entire range $[0, \frac{1}{2}\pi]$, we sample v uniformly in each of the above two intervals and we transform the chosen values of v so that the x or y values of the superquadrics with $\epsilon_2 = 1$ and $\epsilon_2 << 1$ are equal. If v_k is the material coordinate chosen, the transformed value is given by:

$$v'_k = \begin{cases} sin^{-1}(sin^{\frac{1}{\epsilon_2}}(v_k)), & 0 < v_k < v_{cx} \\ cos^{-1}(cos^{\frac{1}{\epsilon_2}}(v_k)), & v_{cy} < v_k < \frac{\pi}{2} \end{cases}, \quad (10)$$

where these formulas result in approximately uniform tessellation in the first quadrant of the xy plane. Fig. 2 illustrates the improved tessellation algorithm.



(a) (b) (c) (b)

Figure 2: Retessellation of a cylinder (a) before (b) after; and a block (c) before (d) after.

5 Shape and pose estimation

5.1 Qualitative shape matching

In [3] a qualitative approach to the representation, recovery, and recognition of 3D objects from a single 2D image was presented. Using the qualitative shape recovery process as a front end, we first segment the image into parts [2] using an aspect matching paradigm. Each recovered qualitative part defines: (i) the relevant non-occluded contour data belonging to the part, (ii) a mapping between the image faces in their projected aspects and the 3D surfaces on the quantitative models, and (iii) a qualitative orientation (that the aspect encodes) which is exploited during model fitting (see [4] for details.). The mapping of certain edge segments corresponding to the *occluding contour* on the model surfaces in (ii) needs to be updated continuously during the fitting process [4].

For stereo reconstruction, the qualitative shape recovery process is independently applied to the left and right images. The correspondence problem then consists of matching qualitative primitive descriptions in the two images. A pair of primitives in two different images are considered a match 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 (not restricted only to parallel geometry) such that both faces intersect this line. For reconstruction from multiple views, the process is similar except that correspondences have to be established for each pair of images.

5.2 Quantitative multi-view integration

Based on the above qualitative shape recovery, correspondences are established between edge segments in the images and the corresponding subsets of nodes on the model surface. If $e_{j,i}$ is the jth edge point in the ith image (i = L or R for stereo case), and $\mathcal{M}_{j,i}$ is the subset of model nodes in correspondence with the edge segment that $e_{j,i}$ belongs to, then the 2D force exerted by that edge point on the projected model (based on a shortest distance criterion)

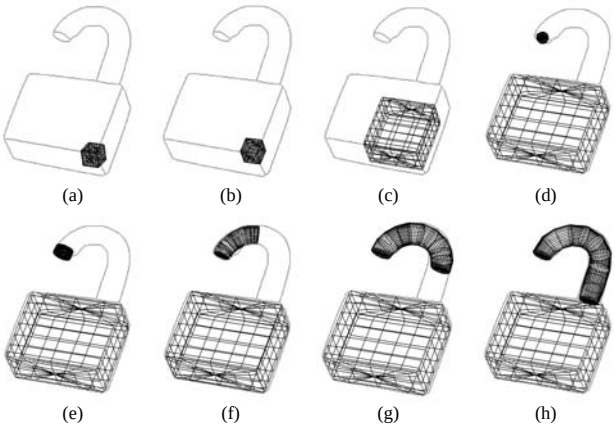


Figure 3: Two models fitted to the image of a lock.

is given by

$$\mathbf{f}_{j,i} = \beta \ \min_{k \in \mathcal{M}_{j,i}} \left(\mathbf{P}_i (\mathbf{R}_{c_i}^{-1} (\mathbf{x}_k - \mathbf{c}_{c_i})) - \mathbf{e}_{j,i} \right) \quad (11)$$

where $\mathbf{x}_k$ is the position of kth model node, β controls the magnitude of the force and $\mathbf{P}_i$ is the perspective projection operator w.r.t. the ith image.

The generalized force exerted on the model can be computed by replacing the integral in (5) by the following summation:

$$\mathbf{f}_q = \sum_j \mathbf{L}_{m_{j,1}}^T f_{j,1} + \ldots + \sum_j \mathbf{L}_{m_{j,n}}^T f_{j,n}, \quad (12)$$

where $m_{j,i}$ is the model node at which forces are exerted by the $e_{j,i}$ and $\mathbf{L}_{m_{j,i}}$ is the Jacobian matrix evaluated at $m_{j,i}$. [2] This net force will appropriately deform, position and orient our model so as to recover the shape and pose of the underlying object.

This new algorithm is significantly simpler and more general than the stereo algorithm previously proposed in [4] for the case of parallel cameras which made use of two models instead of one. By virtue of generalization in section 3, we can now integrate possibly incomplete measurements from more than 2 cameras which are not necessary parallel.

6 Experiments

We show the results of our technique in a series of experiments on model recovery from single and stereo images as well as from multiple views.

The experiment shown in Fig. 3 demonstrates the use of our framework to estimate object shapes under perspective projection from a single image of a lock. Using qualitative shape recovery process OPTICA [3, 4] a box like primitive and a bent cylinder primitive are recovered. Figs. 3(a-d) and (e-h) show the quantitative fitting stages of the 2 models to the lower and upper parts of the lock, respectively. Note that the employment of the new tessellation method is especially important here, since bending deformation is used in the recovery process.

[2] In case of occlusion, resulting in both occluded aspects and occluded faces, only data points belonging to the unoccluded portions (boundary groups) of the faces exert external forces on the models (as demonstrated in section 6.).

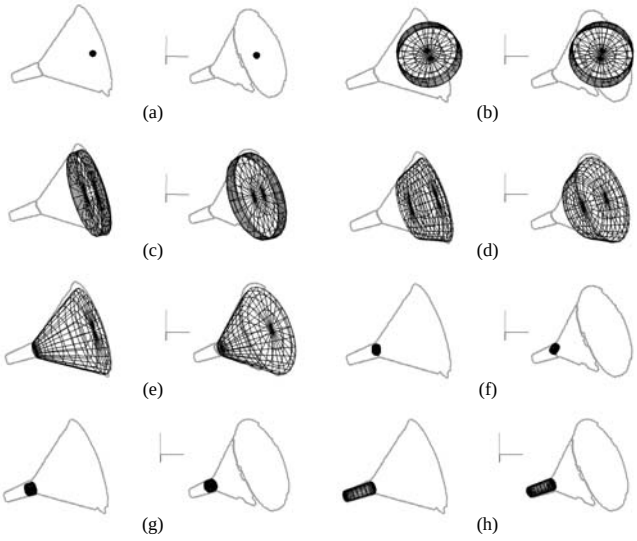


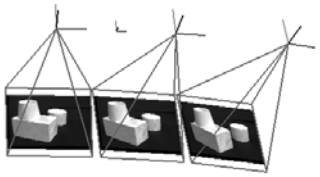
Figure 4: Fitting deformable part models to stereo images of a funnel (the wire frame models are the corresponding projections of the 3D model).

In the second experiment, images of a funnel were taken from 2 views (stereo), where the relative displacements and orientations of the 2 cameras were known. The qualitative shape recovery process performed in both images decomposed the funnel into a tapered cylinder and short cylinder (note that one of the edge segments is missing because of low contrast in the left image). Fig. 4(a-h) show the intermediate stages of the fitting process. Each 3D model is fitted simultaneously to both images.

In the last experiment, images of a scene with 3 objects were taken from 3 different viewpoints (see Fig. 5a). 3 object parts were recovered with the qualitative shape recovery process as shown in Fig. 5b. Some edge segments were missing because of occlusion, poor contrast in the original images or because they were not used to deduce the corresponding shape primitives. Fig. 5c shows the projections of the fitted models overlaid on the respective images. Finally, Fig. 6 shows the shape and pose of the recovered 3D models. This experiment demonstrates how information from incomplete and partially occluded image data from multiple views can be integrated in our model-based approach. This experiment demonstrates how information from incomplete and partially occluded image data

(a) (b) (c)

Figure 5: (a) original images of a scene with 3 objects from 3 different viewpoints, (b) edge segments extracted, (c) projection of fitted 3D model overlaid on the original images.



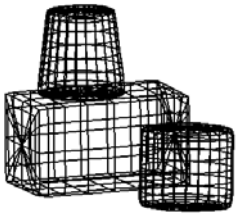


Figure 6: Recovered 3D shape and pose of objects in a scene from 3 different views.

from multiple views can be integrated in our model-based approach.

7 Conclusion

This paper presented a new algorithm for object pose and shape estimation from multiple views including the simpler case of stereo. Our model-based approach assumed the existence of objects that can be decomposed into parts belonging to a vocabulary of primitives. We used qualitative shape recovery techniques to segment images and provide necessary constraints for our physics-based quantitative fitting process. Our new stereo and multiple view integration algorithm allowed object pose and shape estimation in case of cameras with arbitrary relative orientation using a single model. We also defined a new global deformation for bending and we developed a new algorithm for uniform 3D model discretization and hence improved the accuracy of data force computation.

Acknowledgments

We would like to thank S. Dickinson for providing his qualitative shape recovery system OPTICA.

References

[1] A. Barr. Global and local deformations of solid primitives. *Computer Graphics*, 18:21–30, 1984.

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

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

[4] D. Metaxas and S. Dickinson. Integration of quantitative and qualitative techniques for deformable model fitting from orthographic, perspective, and stereo projections. In *Proc. IEEE 4th International Conference on Computer Vision*, pages 641–649, 1993.

[5] 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, 1993.

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

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

[8] F. Solina and R. Bajcsy. Recovery of Parametric Models from Range Images: The Case for Superquadrics with Global Deformations. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(2):131–146, 1990.

[9] D. Terzopoulos, A. Witkin, and M. Kass. Constraints on deformable models: Recovering 3D shape and nonrigid motion. *Artificial Intelligence*, 36(1):91–123, 1988.