To appear in the Proceedings of the Third Asian Conference on Computer Vision (ACCV), Jan 8-10, 1998, Hong Kong p.p. II-455-462.

3-D Reconstruction of Multipart Self-Occluding Objects

Nebojsa Jojic[†], Jin Gu[‡], Helen C. Shen[‡] and Thomas Huang[†]

[†]Beckman Institute, University of Illinois at Urbana-Champaign
405 N. Mathews Ave, Urbana, Illinois 61801, USA
{jojic,huang}@ifp.uiuc.edu

[‡]Department of Computer Science
Hong Kong University of Science and Technology, Hong Kong
{csgjx,helens}@cs.ust.hk

Abstract. In this paper we present a method for reconstruction of multipart objects from several arbitrary views using deformable superquadrics as the models of the object's parts. Two visual cues are used: occluding contours and stereo (possibly aided by projected patterns). The object can be relatively complex and can exhibit numerous self occlusions from some or all views. Our preliminary experiments on a human body and a tailor's mannequin show that the reconstruction is more complete than in purely stereo or structured light based methods and more precise than the reconstruction from occluding contours only.

1 Introduction

In this paper we study the problem of 3D shape reconstruction of objects consisting of several parts which may partially or completely occlude each other from some views.

The usual stereo or structured light methods can not give a complete surface estimate in such cases, as either the lighting source or the cameras may not see all the parts of the object. Moreover, the correspondence problem is not a trivial task in using stereo or structured light methods.

To overcome these difficulties, we model the object with several deformable superquadrics and use occluding contours and stereo to govern part positioning, orientation and deformation. Image contours provide a crude surface estimate which guides stereo matching process. In turn the 3D points provided by the stereo cue can further refine the surface estimate and improve contour fitting.

Compared with related work [5][6], we propose a faster force assignment algorithm based on chamfer images, and we can reconstruct multiple objects and multipart self-occluding objects, which can be rigid or may not be capable to perform prescribed set of movements as required in [3]. We can use arbitrary camera configuration, unlike the case in [7] where parallel projection and coplanar viewing direction are assumed. Moreover, the whole scheme avoids the problem of merging of reconstructed surface patches from different views as in [7].

2 Deformable Superquadric Geometry and Mechanics

In this section, we describe briefly the deformable model we adopt as the 3D part model. The details of this model can be found in [6].

The deformable model is represented as a sum of a reference shape

$$\mathbf{s}(u,v) = \left[a_1 C_u^{\epsilon_1} C_v^{\epsilon_2}, a_2 C_u^{\epsilon_1} S_v^{\epsilon_2}, a_3 S_u^{\epsilon_1}\right], \tag{1}$$

and a displacement function $\mathbf{d}(u,v) = \mathbf{S}(u,v)\mathbf{q_d}$, where u and v are material coordinates; $S_w^\epsilon = sgn(\sin w) \mid \sin w \mid^\epsilon$; $C_w^\epsilon = sgn(\cos w) \mid \cos w \mid^\epsilon$; a_1, a_2, a_3, ϵ_1, ϵ_2 are global deformation parameters (stored in $\mathbf{q_s}$); $\mathbf{q_d}$ contains nodal variables (displacements at the nodes sampled over material coordinates), and $\mathbf{S}(u,v)$ is the shape matrix containing basis functions in the finite element representation of the continuous displacement function.

In addition to global and local deformations of the parts, a rigid transformation of each part is allowed. It is defined by translational and rotational degrees of freedom $\mathbf{q}_c$ and $\mathbf{q}_\theta$. Collecting all the possible degrees of freedom for a part together, we can describe the state of a deformable model in terms of a time-varying vector $\mathbf{q} = (\mathbf{q}_c^T, \mathbf{q}_\theta^T, \mathbf{q}_s^T, \mathbf{q}_d^T)^T$. Under the external forces $\mathbf{f}$ the model will move and deform according to the following equation:

$$\mathbf{C}\dot{\mathbf{q}} + \mathbf{K}\mathbf{q} = \mathbf{f}_\mathbf{q}, \tag{2}$$

where $\mathbf{C}$ and $\mathbf{K}$ are the damping and stiffness matrices, respectively, $\mathbf{f_q}$ are the generalized external forces associated with the degrees of freedom of the model. These forces are related to the external forces derived from image data [6].

The image forces we use in this paper can be written as $\mathbf{f} = \mathbf{f}_{contour} + \mathbf{f}_{3D}$, where the first component of the force deforms the superquadric to have similar occluding contours as the imaged object and the second component governs fitting of the model to the range data provided by the stereo cue.

3 Contour Force Computation

We first find model nodes residing on the occluding contour. With respect to camera l, such nodes $\mathbf{P}_i$ should satisfy $\left|\mathbf{N}_i \cdot (\mathbf{P}_i - \mathbf{O}_l)\right| \leq \epsilon$, where $\mathbf{O}_l$ is the optical center of the camera, $\mathbf{P}_i$ is the position vector of the model node i, and $\mathbf{N}_i$ is the surface normal at this node(refer to Fig. 1). Then $\mathbf{P}_i's$ are projected onto the image plane by projection operator $\prod_l$ acquired by camera calibration, i.e. $\mathbf{p}_i^I = \prod_l \mathbf{P}_i$, where the superscript I denotes the image plane points.

For each image contour point $\mathbf{c}_k^I$, the closest model projection $\mathbf{p}_i^I$ is found. Let $\mathbf{P}_i$ be its corresponding 3D point on the model, and $\mathbf{C}_k$ be a 3D point that is projected to $\mathbf{c}_k^I$, i.e. $\mathbf{c}_k^I = \prod_l \mathbf{C}_k$. As can be seen in Fig. 1, such $\mathbf{C}_k$ should lie on the line formed by $\mathbf{O}_l$ and $\mathbf{c}_k^I$. To find a single direction for the force to bring $\mathbf{p}_i^I$ to $\mathbf{c}_k^I$ we utilize the principle of minimal action and compute $\mathbf{C}_k$ as follows:

$$\mathbf{C}_k = arg\Big(min_{\mathbf{C}_k} ||\mathbf{C}_k - \mathbf{P}_i||\Big), \qquad given \quad \mathbf{c}_k^I = \prod_l \mathbf{C}_k. \tag{3}$$

The contour force acting on the model point $\mathbf{P}_i$ is defined as:

$$\mathbf{f}_{c_i} = k_c(\mathbf{C}_k - \mathbf{P}_i), \tag{4}$$

where k_c is a scaling constant. In the case of a pinhole camera, the solution to the Eq. 3 is simply the orthogonal projection of $\mathbf{P}_i$ onto the line $\mathbf{O}_l\mathbf{c}_k^I$. In general, $\mathbf{C}_k$ could be found independent of camera model using Eq. 3.

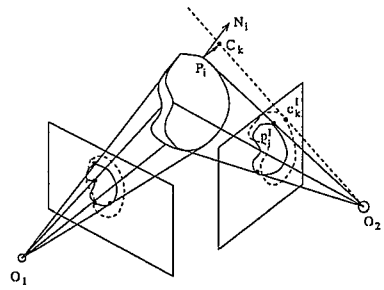


Fig. 1. Contour Forces

4 Contour Force Assignment Using Chamfer Image

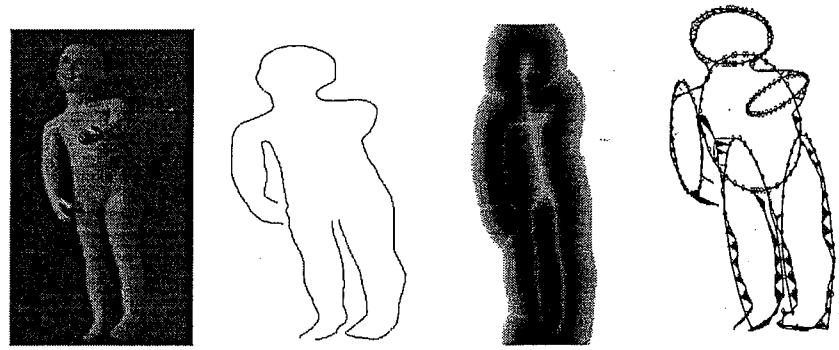
To avoid a computationally expensive search for the closest model point for each contour point, we developed an algorithm for force assignment based on the chamfer image of the occluding contours with an additional index matrix containing the index of the closest contour point for each pixel in the image.

For each image, occluding contours are stored as a set of ordered lists of contour points (each list corresponds to a contour segment). The chamfer image, containing distances to the closest contour point for each image pixel, is computed using a slight modification of the well known two-pass algorithm[1]. The modification consists of adding the index matrix $\mathbf{I}$ to track the closest contour point for each pixel. When a model point $\mathbf{P}_i$ is projected to a point $\mathbf{p}_i^I = (x_i^I, y_i^I)^T$ in the image plane, its closest contour point $\mathbf{c}_k$ can be retrieved immediately using the index k stored in the index matrix $\mathbf{I}$, i.e. $k = \mathbf{I}(\lfloor x_i^I \rfloor, \lfloor y_i^I \rfloor)$. Now the force acting to bring $\mathbf{p}_i^I$ to $\mathbf{c}_k^I$ can be computed using Eq. 4.

However, this force assignment does not guarantee a good final fit, as the model points are "choosing" the data points, instead of vice versa. The problem becomes apparent when the model parts are small or relatively far away from the image contours, causing the model to be attracted only by parts of the image contours. It is also possible that all the model points are attracted to a nearly straight segment of the contour causing the superquadric to further reduce in width instead of growing. Finally, two superquadrics can be attracted to the same part of the occluding contour, even though one of them is much closer to this part than the other. We add a post-processing step to deal with these problems, by making the force assignment more similar to the usual approach to fitting, where the data points "choose" which model points they will attract.

After the first assignment based on the chamfer images we go through all the image contour points and check if they have been assigned to model points. Suppose the contour point $\mathbf{c}_k^I$ has been assigned to a model point projection $\mathbf{p}_i^I(k)$, the closest assigned contour point with a smaller index k_1 has been assigned to $\mathbf{p}_i^I(k_1)$, and the closest assigned contour point with a larger index k_2 has been assigned to $\mathbf{p}_i^I(k_2)$. Any contour point with index l satisfying $k_1 < l < k$ or

$k < l < k2$, has not yet been assigned to the model points. If any of the points $\mathbf{p}_i^I(k_1)$, $\mathbf{p}_i^I(k_2)$ is closer to the contour point $\mathbf{c}_k^I$ than its assigned point $\mathbf{p}_i^I(k)$, $\mathbf{p}_i^I(k)$ is discarded, the closer one of $\mathbf{p}_i^I(k_1)$, $\mathbf{p}_i^I(k_2)$ is assigned to $\mathbf{c}_k^I$ instead. After this, each unassigned contour point is assigned to the closer of the model points to which its neighboring contour points have been assigned. An example of the force assignment in case of complex image contours and six superquadrics is shown in Fig. 2. The whole process of force assignment and computation can



(a)Intensity Image (b) Image Contours (c)Chamfer Image (d)Force Assignment

Fig. 2. Force Assignment

be done in $O(N_m) + O(N_c)$ time, where N_m and N_c are numbers of model and contour points, while the ordinary brute force search for the closest model point for each of the contour points is done in $O(N_m \times N_c)$ time.

However, in the beginning of the fitting process, when the superquadrics are crudely positioned, oriented and sized, the described scheme can cause some errors in force assignment. To avoid that, we do the force assignment based on the brute force search in the first few time steps in solving Eq. 2. After the superquadrics global parameters have reached more reasonable values, force assignment can be based on the algorithm described above. Finally, we note that force assignment need not be done in each time step of integration of the motion equation. It can be done only every few iterations.

5 Forces Based on Stereo Aided by Structured Light

Providing that the correspondence between the features in two images are available, it is possible to reconstruct a number of 3D points on the object's surface using triangulation techniques. In case of the lack of feature points, a structured lighting source can be used to impose extra feature points on object surface.

Correspondence establishment in stereo is known to be a difficult problem. In this paper, we make use of the surface estimated by occluding contours based

on the deformable superquadric models to assist the matching of feature points from two images. A match between two features is considered good only if it satisfies both the epipolar constraint and a constraint on the distance from the point, which is reconstructed by the match under consideration, to the surface estimated using occluding contours.

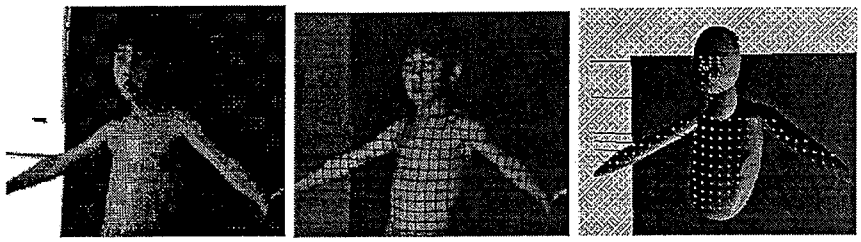


Fig. 3. Structured light provides feature points: (a)The intensity image of a human object; (b)The object illuminated by structured light; (c)The reconstructed feature points overlapped with an estimate from contours

Once the feature correspondences have been established and a number of surface points $\mathbf{R}_k$ have been reconstructed by triangulation, the part models can be further deformed to fit these points by applying forces:

$$\mathbf{f}_{3D_i} = k_{3D}(\mathbf{R}_k - \mathbf{P}_i), \tag{5}$$

where $\mathbf{P}_i$ is the model node closest to the point $\mathbf{R}_k$ and k_{3D} is scaling constant. Brute force search is used in finding force assignment because the surface estimated from occluding contours usually provides such a good initial guess that the force assignment for fitting the 3D points $\mathbf{R}_k$ hardly varies over time.

The forces derived from stereo cue not only refine the surface reconstruction, but also may reposition superquadrics slightly so that a better contour force assignment can be achieved. Therefore, the two visual cues assist each other. They also complement each other, as occluding contours provide estimates at the parts where stereo becomes unreliable, while stereo refines the estimate at the rest of the surface.

6 Force Assignment Based on Fuzzy Clustering

We experimented with weighted force assignment based on fuzzy clustering proposed by Kakadiaris[4]. Instead of using the nearest neighbor force as described in the previous sections, each data point $\mathbf{r}_i$ should attract points on several models with forces scaled by a weight $p_{r_i,j}$ equal to the probability that this point is correctly associated with the superquadric j. Hence the force that the data point applies to the nearest point $\mathbf{x}_i^j$ on superquadric j is:

$$\mathbf{F}_{r_i}^{j} = p_{r_i,j} k(\mathbf{r}_i - \mathbf{x}_i^j). \tag{6}$$

The probability $p_{r_i,j}$ is computed by fuzzy classification of data points based on their proximity to the model points:

$$p_{r_i,j} = \frac{\frac{1}{||\mathbf{r}_i - \mathbf{x}_i^j||}}{\sum_k \frac{1}{||\mathbf{r}_i - \mathbf{x}_i^k||}}. \tag{7}$$

This scheme allows a single data point to attract several superquadrics, which is desirable in reconstruction of multipart objects, as it serves the purpose of gluing the parts (for example, the arms in Fig. 4, or legs in Fig. 5).

The weighted assignment method can be directly applied to the forces $\mathbf{f}_{3D}$ and to the contour forces $\mathbf{f}_c$ described in previous sections, if the brute force search method is used in the search for the closest point on a model part.

To use fuzzy clustering with the fast force assignment algorithm proposed in Section 4, we modify the postprocessing step of that algorithm in the following way. Instead of examining the assignments of a contour point's immediate neighbors, a segment of certain length on the contour can be studied. All the model points from different parts that were originally assigned to some of the points on that contour segment, can be attracted by all the points in the observed contour segment, including the ones that were not originally assigned to any of the model points. The assignment is valid with the probability in Eq. 7, and the forces are later computed by scaling the right hand side of the Eq. 4 by this weight.

7 Experimental Results and Conclusions

We performed preliminary experiments on human-like objects with two CCD cameras and a structured lighting source in between to project a stripe pattern on the object surface. Both ordinary intensity images and images under structured lighting are taken.

In the first experiment, we imaged a real human(Fig. 5) with two cameras in the above configuration. Five deformable superquadrics (for arms, torso, neck and head) are manually positioned in the virtual 3D space so that their projections onto image plane lie relatively near the image contours (see the right figure of Fig. 4(a)). Fig. 4 illustrates several steps in integrating Eq. 2. The left column shows the model parts smoothly shaded or texture mapped and the right column shows the fitting of the model parts to the contours in one of the images.

In the second experiment, we tried to reconstruct a more difficult object - a doll with such a posture that the body parts occlude each other(see Fig. 2(a)). The object was positioned on the turning table and a total of eight camera views were used in reconstruction. Six superquadrics were initialized and in Fig. 5(b) we show the final result after fitting to both contours and the 3D points reconstructed by stereo cue. The reconstructed surface is compared to the reconstruction using only stereo correspondences obtained by manual matching to demonstrate how efficiently our scheme deals with self occlusion.

The experiments show that the contour based estimate of the surface, even by only two cameras can efficiently assist stereo matching. In the first experiment, for example, for the 80 feature points detected in the image, 74 correct matches are found purely guided by the contour based estimate of 3D surface. Usually,

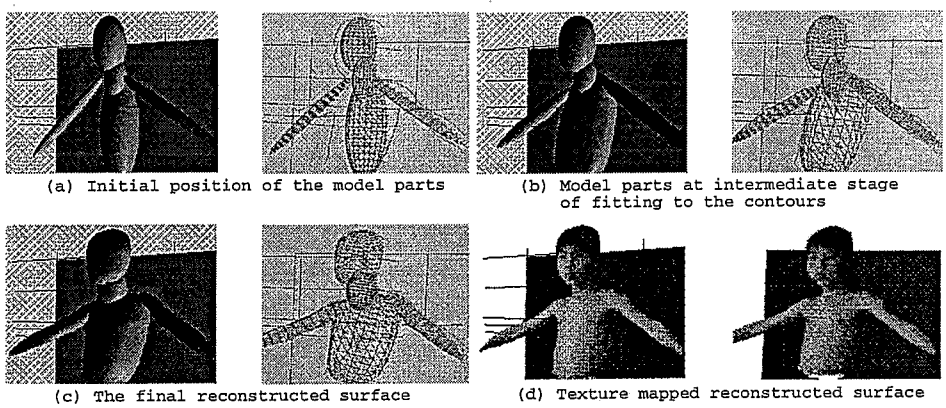


Fig. 4. Reconstruction of a human upper body

the remaining false matches are corrected as well during further refinement of the surface estimate using both cues. As can be seen from the Fig. 5(a), the stereo aided by structured light can not make estimates near the occluding contours due to absence of matched feature points there, but the combined two cues can provide a rather complete estimate of those parts that were not visible to both cameras.

The advantage of our approach to 3D reconstruction is that it does not require special camera configuration. Also, no prescribed set of movements is needed as in [3] which makes it possible to reconstruct rigid multipart object such as the doll in Fig. 2(a). This may be useful in applications such as mannequin manufacturing in the textile industry. Other applications include dressing humans in a virtual garment designed using CAD systems. For example, in [2], a virtual T-shirt is shown draped over the body reconstructed in Fig. 4(d).

In the future, improving the "gluing" of object parts together is one of the tasks we are going to work on. Additional cues, such as orientation and curvatures of the projected stripes, shading, may be added in the reconstruction algorithm. In some applications, it may be beneficial to develop an automatic part initialization algorithm based on color and texture.

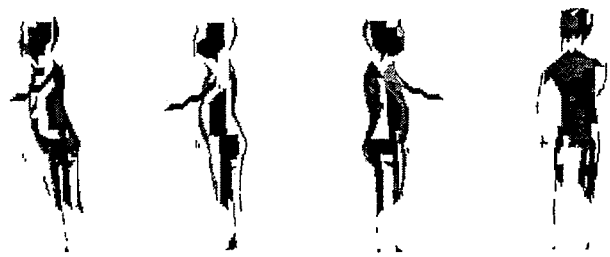
Acknowledgements

This research project is supported by a grant from The Industry and Technology Development Council, (grant no. AF/122/96). The authors from University of Illinois were also supported by Army Research Laboratory under Cooperative Agreement No. DAAL01-96-2-0003. We would also like to thank Roderick Yuen, the son of Dr. Matthew Yuen, the project manager of the grant, for his help in our image acquisition.

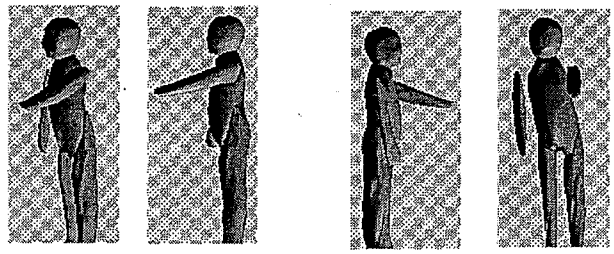
The first author would like to thank Ioannis Kakadiaris for very useful discussions.

References

1. H. G. Barrow, J. M. Tenenbaum, R. C. Bollers, and H. C. Wolf. Parametric corre-



(a) Reconstruction by structured light only.



(b) Reconstruction based on integration of occluding contours and stereo aided by structured light.

Fig. 5. Reconstruction of a Mannequin

spondence and chamfer matching: Two new techniques for image matching. *Vision-7*, pages 659–63, 1979.

2. N. Jojic and T. S. Huang. On analysis of cloth drape range data. In *these proceedings (ACCV '98)*.
3. I. Kakadiaris and D. Metaxas. Model-based estimation of 3d human motion with occlusion based on active multi-viewpoint selection. In *Proceedings 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 81–7, 1996.
4. I. A. Kakadiaris. *Motion-Based Part Segmentation, Shape and Motion Estimation of Complex Multi-Part Objects: Application To Human Body Tracking.* PhD thesis, University of Pennsylvania, Philadelphia, PA, 1996.
5. D. Metaxas and D. Terzopoulos. Shape and nonrigid motion estimation through physics-based synthesis. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 15(6):580–91, 1993.
6. 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):703–14, 1991.
7. Y. F. Wang and J. K. Aggarwal. Integration of active and passive sensing techniques for representing three-dimensional objects. *IEEE Transactions on Robotics and Automation*, 5(4):460–71, 1989.

This article was processed using the LaTeX macro package with LLNCS style