Shape and Pose Parameter Estimation of 3D Multi-part Objects

Yonemoto, Satoshi
Department of Intelligent Systems, Kyushu University

Tsuruta, Naoyuki
Department of Intelligent Systems, Kyushu University

Taniguchi, Rin-ichiro
Department of Intelligent Systems, Kyushu University

出版情報：Proceedings of Asian Conference on Computer Vision, ACCV'98, pp.479-486, 1998-06
バージョン：
権利関係：

Shape and Pose Parameter Estimation of 3D Multi-part Objects

Satoshi YONEMOTO, Naoyuki TSURUTA, Rin-ichiro TANIGUCHI

Department of Intelligent Systems , Kyushu University
6-1, Kasuga-koen, Kasuga, Fukuoka 816 JAPAN
yonemoto,tsuruta,rin@is.kyushu-u.ac.jp

Abstract. This paper presents an analysis-by-image-synthesis framework of shape and pose estimation of 3D multi-part objects, whose purpose is to map objects in the real world into virtual environments. In general, complex 3D multi-part objects cause serious self-occlusion and non-rigid motion. To deal with the occlusion among them, we employ both multiple calibrated cameras and time-varying sequences, since there is enough information to estimate the parameters in the sensory data. In our framework, to minimize the error between the selected measurements and the estimated model parameters, we proceed model fitting process based on proper gradient-based minimization.

1 Introduction

We have been developing a system to construct easily virtual environments by means of seamless fusion of information obtained by observing or measuring complex real world objects. As one of the actual applications, we consider 3D animation in which various kinds of natural creatures appear. Therefore, in our approach, animals are intended for the main subject, which have scarcely been considered in the past studies because of its complexity.

For designing this system, the following issue is quite important.

Required information to map the real world objects into a virtual environment consists of two kinds of information: one is a priori object model, or a priori knowledge of the objects; the other is the result of a posteriori observation, or measured object parameters. To construct an efficient or easy-to-use mapping system, the most important point is where we establish the boundary between a priori object model and a posteriori observation.

To simplify a posteriori observation, we need a precise object model in advance, which causes the difficulty in constructing the object model. On the contrary, when we assume a simple a priori object model, a posteriori observation becomes difficult, i.e., we have to solve very difficult CV problems.

For this problem, our approach adopted here is as follows:

A priori object model: each object consists of parts represented in deformable superquadrics[6], which are connected via articulation points one another.

A posteriori observation: estimation of parameters of each deformable superquadrics.

We give a priori object model, the approximated shape of parts and their connection structure, to the system interactively, referring to the initial frames of image sequences, taken by multiple cameras. To simplify this process, we have constructed a GUI-based model description system, whose details we will be omitted here.

A posteriori observation means here recovering 3D shape and pose parameters from images, which is an ill-posed problem without proper constraints as a priori object model. Generally, there are two types of approaches to solve this problem:

- direct 3D matching between a 3D geometric model and the measurements acquired by 3D shape recovering techniques such as *shape from X* techniques.
- 3D estimation from 2D matching between the 2D projected model data and the 2D measurements.

In the former approach, there is no method to reliably recover 3D shapes yet. Therefore, we have adopted the latter approach. Although it often takes huge computation time, once we get camera parameters, we can solve it as a direct problem under the projection geometry. Our approach is based on analysis-by-synthesis approach[12], and has the following two advantages:

- Non-rigid multi-part objects, which are essential in the "real world," can be handled by using deformable superquadrics models, while most of existing approaches for 3D shape and pose estimation assume their rigidness.
- A multiple camera system is employed to handle self-occlusion among the parts of the object. While a multiple viewpoint image analysis is getting popular to solve the ill-posed problem[4] [8][9] [7][11], no method solving self-occlusion problem properly is proposed, except for [10].

This paper shows our framework for shape and pose parameter estimation of 3D multi-part objects focussing on model description and parameter estimation. Some experimental results are shown.

2 Framework for Shape and Pose Parameter Estimation

2.1 Principles of our Analysis-by-Image-Synthesis

The general principle of analysis-by-image-synthesis is described as follows:

1. Project a model into the same dimension as the measurements to detect the error between the projected model data and the measurement space.
2. Minimize the error by means of adapting the model parameters to the observed data.

To estimate the model parameters from the measurements is generally an ill-posed problem because of the lack of the accuracy and of the noisy data. Therefore, 2. should be solved by using proper optimization techniques.

In our analysis-by-image-synthesis, the models are represented in 3D parameterized geometric models, and the measurements are real 2D image data, which is actually expressed in image features. In this sense, the principle is generally called appearance-based matching. As we consider the object models consist of multi-parts, the problem is not simple, i.e., the model fitting process for each part may interfere one another. The interfere, in the projection, is observed as self-occlusion. However, since the self-occlusion relationship can be acquired from multi-part model structure in advance, we can select only a set of corresponding pairs of the visible model sample data and the measured points.

2.2 Outline of 3D Tracking Multi-part Objects

Based on the above framework, an outline of 3D tracking multi-part objects is:

1. Acquire an initial model, using initial multi-viewpoint frames: Build the model structure of the object and then estimate initial model parameters by means of semi-automatic modeling[1].
2. Track image features from the previous frame to the current frame in every viewpoint[2].
3. Estimate model parameters of every part.
4. Refine the model parameters in accordance with the result of 3.
5. Iterate 2-4 in the succeeding frames.

3 Modeling Multi-part objects

3.1 Interactive 3D Modeling Tool

For adapting the system for various objects, as is the above mentioned, it requires a priori models of them. In usual systems, however, only system developers can make their models, which causes the lack of system flexibility. Therefore, it is better that users can participate the modeling process, or that they can give desirable models a priori, and simultaneously determine initial model parameters.

On the other hand, since the modeling process requires a great deal of skill, the users' intervention should be minimized from the viewpoint of "easy addition of a priori knowledge". Therefore, we are developing a GUI-based interactive 3D shape modeling tool so as to alleviate the burden imposed on the users as little as possible.

[1] We employ an interactive 3D shape modeling tool.

[2] We assume that it should be able to achieve this process by means of using reasonable low-level feature detection and its tracking method, which is omitted in this paper.

3.2 Object Model Description

In our method, any 3D model description can be introduced only if it is represented in parameterized-form. As a 3D parametric model here, we consider deformable superquadrics(we call DSQ). Although the various types of DSQ are designed so far[1] [5][6][13], we employ the one developed in [6], which has an advantage in the sense that it can be represented by a small number of parameters and can represent deformations such as tapering and bending.

DSQ Geometry When (η, ω) is a material coordinate system, a point on SQ $\mathbf{e}$ is[1]:

$$\mathbf{e}(\eta,\omega) = \begin{pmatrix} e_1(\eta,\omega) \\ e_2(\eta,\omega) \\ e_3(\eta,\omega) \end{pmatrix} = a \begin{pmatrix} a_1 \cdot C_\eta^{\epsilon_1} \cdot C_\omega^{\epsilon_2} \\ a_2 \cdot C_\eta^{\epsilon_1} \cdot S_\omega^{\epsilon_2} \\ a_3 \cdot S_\eta^{\epsilon_1} \end{pmatrix}, \tag{1}$$

where $-\frac{\pi}{2} \leq \eta \leq \frac{\pi}{2}$, and $-\pi \leq \omega < \pi$, and a, a_1, a_2, a_3 are scale parameters, and ϵ_1, ϵ_2 are squareness parameters, and where $C_w^\epsilon = \mathrm{sign}(\cos w)|\cos w|^\epsilon$, and $S_w^\epsilon = \mathrm{sign}(\sin w)|\sin w|^\epsilon$.

Using $\mathbf{e}$ on SQ, a point on DSQ $\mathbf{s}$ is expressed as [6]:

$$\mathbf{s} = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} = \begin{pmatrix} (\frac{t_1 e_3}{a a_3} + 1)e_1 + b_1 \cos(\frac{e_3 + b_2}{a a_3}\pi b_3) \\ (\frac{t_2 e_3}{a a_3} + 1)e_2 \\ e_3 \end{pmatrix}, \tag{2}$$

where t_1, t_2 are tapering parameters, and b_1, b_2, b_3 are bending parameters.

From the above definitions, the shape and pose parameters of each part can be expressed as:

$$\mathbf{q} = (a, a_1, a_2, a_3, \epsilon_1, \epsilon_2, t_1, t_2, b_1, b_2, b_3, r_1, r_2, r_3, c_1, c_2, c_3)^T \tag{3}$$

where r_1, r_2, r_3 are rotation parameters, and c_1, c_2, c_3 are translation parameters for each part.

Multi-part Geometry In our method, multi-part object models consist of the 3D deformable parts(DSQ) and the structure is represented in a hierarchical tree structure(See Section 4.2).

4 Parameter Estimation of Multi-part Objects

4.1 Acquiring Initial Model

Using the modeling tool, we can build model structure and make registration of part i on the initial multi-viewpoint frames $f_v(0)$ $(v = 1, \cdots, V; i = 1, \cdots, N)$ where V is the maximum number of viewpoint and N is the maximum number of constituent parts. The registration gives:

1. Initial shape and pose parameters.
2. The correspondence between the initial model data and the initial measurements for tracking in the succeeding frames, to determine to which model sample on a part the measurement corresponds.

4.2 Estimation Strategy

In general, multi-part objects essentially include some constraints based on their structure, which are useful to reduce the search space. In a proposed method, we use a simple top-down strategy under the structural constraints to reduce the computation time. We impose the following structure constraints on objects: 1) Usual objects can be expressed as hierarchical tree structures of parts; 2) these constituent parts are connected via articulation points one another. As a result, pose parameters of a part have an influence on its descendant parts, and consequently, it leads to the quasi-optimal solution.

1. Set $i = root$.
2. Estimate the parameters of $part_i$.
3. If it has child parts, then estimate their parameters recursively(goto 2).
4. If not, stop the recursion.

4.3 Parameter Estimation of Each Part

Model Fitting Problem In our analysis-by-image-synthesis approach, the parameter estimation can be reduced into the model fitting of the measurements based on a proper numerical analysis. When the error between the model and the measurements is represented by d_j, in general, the objective function E is defined as the follows:

$$E = metric \sum_j (w_j \rho(d_j)) \tag{4}$$

In this paper, we simply used $\rho(x) = x^2$, $w_j = 1$ and $metric$ is min. Therefore, we define the error between the model sample points and the measured points as the following equation, and minimize it:

$$E_i(s) = \sum_{v=1}^{V} \sum_{j \in {}^vP^i} \left(({}^vU_{m_j}^{\ i}(s) - {}^vU_j^i)^2 + ({}^vV_{m_j}^{\ i}(s) - {}^vV_j^i)^2 \right), \tag{5}$$

where ${}^vP^i$ is the number of corresponding pairs on $part_i$ from viewpoint v and s is a computation time to minimize.

Details of Parameter Estimation

Step 1 Determine the correspondence between the model points and given measured points from each viewpoint frames $f_v(t)$ $(v = 1, \cdots, V)$:

(1.1) Select visible model sample points:

Since the object model consists of multi-parts, the projection of the model includes two kinds of the occlusion: 1) self-occlusion caused by a part itself, 2) occlusion caused by the other parts.

To deal with such occlusion, we have to delete model sample points do not appear on the projection.

The selection process proceeds as follows:

1. Making the projection for all parts by using a *Z-buffer* technique.
2. Collect model sample points on $part_i$ from the projection.

In this way, points which should be measured can be selected in accordance with model-based approach(top-down processing).

(1.2) Select good measured points for tracking:
Among the measured points on $part_i$ at previous frame, select the *trackable* measured points. In addition, to deal with model sample points newly appeared in (1.1), extract other good measured points to track in the place of lost measured points. This is performed completely in a bottom-up way.

(1.3) Determine corresponding pairs:

1. Delete the corresponding pairs for the model sample points deleted because of new occlusion in (1.1), and non-trackable measured points in (1.2).
2. Get new corresponding pairs between model sample points newly appeared because of disocclusion in (1.1) and newly found good measured points to track in (1.2). the correspondence is determined by 2D Euclidean distance similarity.

We have acquired the following information at the current frames, using the results which are obtained at the previous frame:

1) The correspondence between the model sample points and selected measured points at previous frame $f_v(t-1)$:

$$\{({}^vU_{m_j}^i(t-1), {}^vV_{m_j}^i(t-1))^T, ({}^vU_j^i(t-1), {}^vV_j^i(t-1))^T\},$$

2) The correspondence between tracked measured points:

$$\{({}^vU_j^i(t), {}^vV_j^i(t))^T, ({}^vU_j^i(t-1), {}^vV_j^i(t-1))^T\}$$

Using the above two consequences, we can get the initial correspondence(hypothesis) between the model sample points $({}^vU_{m_j}^i(t), {}^vV_{m_j}^i(t))^T$ and selected measured points $({}^vU_j^i(t), {}^vV_j^i(t))^T$ at current frame $f_v(t)$, which should be estimated:

$$\{({}^vU_{m_j}^i(t), {}^vV_{m_j}^i(t))^T, ({}^vU_j^i(t), {}^vV_j^i(t))^T\}.$$

Step 2 Estimate the parameters iteratively:

(3.1) Tuning each parameter α_k to converge toward the orientation of gradient-descent:

$$\alpha_k(s+1) = \alpha_k(s) - \beta(s+1)\frac{\partial E_i(s)}{\partial \alpha_k(s)}, \tag{6}$$

where $k \in \{a, \cdots, b_3, r_1, \cdots, c_3\}$ and $\beta(s)$ is the step size.

(3.2) Compute the objective function $E_i(s+1)$ for the following corresponding pairs between the model sample points(should be updated) and selected measured points(in the following formula, the flow number t is omitted. s means the iteration time):

$$\{({}^v\mathrm{U}_{\mathrm{m_j}}^{\mathrm{i}}(s+1), {}^v\mathrm{V}_{\mathrm{m_j}}^{\mathrm{i}}(s+1))^{\mathrm{T}}, ({}^v\mathrm{U}_{\mathrm{j}}^{\mathrm{i}}, {}^v\mathrm{V}_{\mathrm{j}}^{\mathrm{i}})^{\mathrm{T}}\},$$

where $j \in {}^vP^i$; $v = 1, \cdots, V$.

(3.3) Go to (3.1) unless the function converges.

5 Experimental Results

The proposed method is performed in the following sequence: First, to acquire a sequence of multi-viewpoint frames, real cameras are mutually calibrated, and then virtual cameras to make model views are set up in the same configuration. Then, using the interactive modeling tool, we make a model structure of a object and get initial model parameters. Finally, to extract the object, based on the initial model(and the result of registration), we estimate model parameters from frame to frame. In the experiment, to evaluate the basic performance of the proposed method, we have applied it to synthesized image data. We have created the synthetic data rendered by employing three virtual cameras orthogonally calibrated. We have used an object having serious occlusion, i.e., one consists of three parts connected by two joints, like an arm. We have used a steepest descent method as gradient-based minimization in estimating their parameters. Fig.1 shows the input model views acquired from virtual cameras and reconstruction of the estimated model views from the same viewpoint. Fig.2 shows the transition of the number of the corresponding pairs on middle part. Each bar represents the number of the corresponding pairs(the black bars are newly appeared, the grays are even lost and the whites are tracked). These results show that, in case we must depend on an unstable feature tracking method, using this model-based approach, we can select only the proper corresponding pairs between the model data and the measurement.

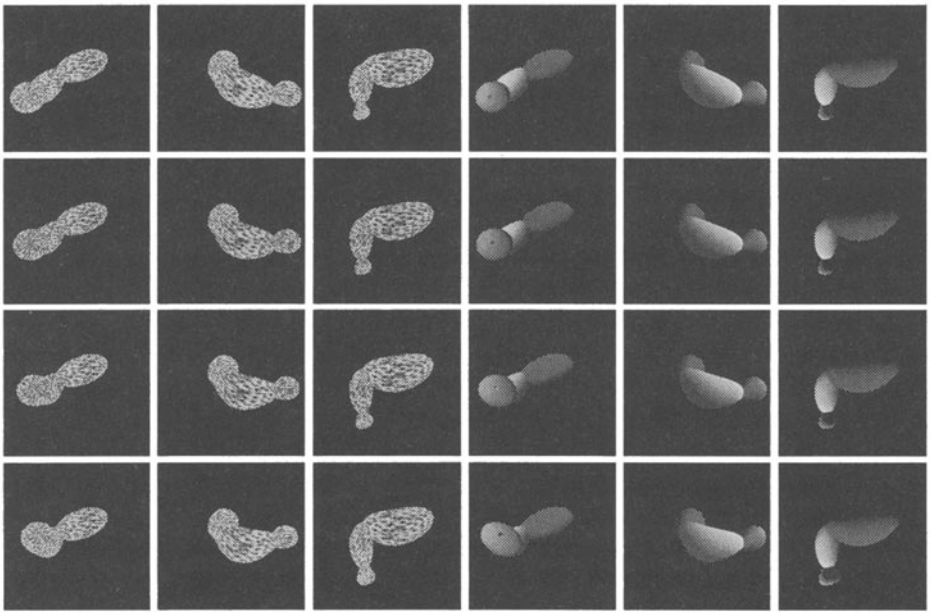

Fig. 1. (Left): Input views. (Right): Reconstruction of estimated model views from the same viewpoint. Each row indicates time-varying sequences at one time interval, and each column indicates multi-viewpoint frames(front,left,up).

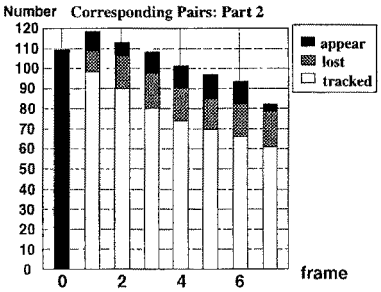

Fig. 2. the transition of the number of the corresponding pairs(on middle part).

6 Conclusions

To map efficiently the real world objects into a virtual environment, we have adopted two kinds of information: one is a priori object model, which consists of deformable parts; the other is a posteriori observation, which means recovering 3D shape and pose parameters from images. And we have explained the advantages in our approach is that non-rigid multi-part objects can be handled, and a multiple cameras are employed to deal with self-occlusion among their parts. The experiments using synthetic data have proved the basic performance of the proposed method. In future work, we plan precise experiments using real image data. Moreover, to acquire more precise model, we try to introduce local deformation into a model. To develop a real-time system is also a future work.

References

1. A.H. Barr, *Global and Local Deformations of Solid Primitives*, Computer Graphics(*Proc.*SIGGRAPH'84), Vol.18, No.3, pp21-29, 1984.
2. A. Pentland and S. Sclaroff, *Closed-Form Solutions for Physically Based Shape Modeling and recognition*, PAMI-13, No.7, pp715-729, 1991.
3. A. Pentland and B. Horowitz, *Recovery of Nonrigid Motion and Structure*, PAMI-13, No.7, pp730-741, 1991.
4. D.G. Lowe, *Fitting Parameterized Three-Dimensional Models to Images*, PAMI-13, No.5, 1991.
5. D. Terzopoulos and D. Metaxas, *Dynamic 3D Models with Local and Global Deformations: Deformable Superquadrics*, PAMI-13, No.7, pp703-714, 1991.
6. D. Metaxas and D. Terzopoulos, *Shape and Nonrigid Motion Estimation through Physics-Based Synthesis*, PAMI-15, No.6, pp580-591, 1993.
7. D.M.Gavrila, L.S.Davis, *3D model-based tracking of Humans in action: a multiview approach*, CVPR, pp73-80, 1996.
8. I.A. Kakadiaris and D. Metaxas, *3D Human Body Model Acquisition from Multiple Views*, ICCV, pp618-623, 1995.
9. I.A. Kakadiaris and D. Metaxas, *Model-Based Estimation of 3D Human Motion with Occlusion Based on Active Multi-Viewpoint Selection*, CVPR, pp81-87, 1996.
10. J.M. Rehg, T. Kanade, *Model-Based Tracking of Self-Occluding Articulated Objects*, ICCV, pp612-617, 1995.
11. J.Ohya, F.Kishino, *Human posture estimation from multiple images using genetic algorithm*, 12th ICPR, pp750-753, 1994.
12. R. Koch, *Dynamic 3-D Scene Analysis through Synthesis Feedback Control*, PAMI-15, No.6, pp556-568, 1993.
13. F. Solina and R. Bajcsy, *Recovery of Parametric Models from Range Images: The Case for Superquadics with Global Deformations*, PAMI-12, No.2, pp131-146, 1990.