MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11 - 13, 2002, Nara- ken New Public Hall, Nara, Japan

3—5

Non-contact Virtual Clay Modeling Interface using Multi-viewpoint Images

Etsuko Ueda Hirotaka Kodama Yoshio Matsumoto Tsukasa Ogasawara *
Graduate School of Information Science
Nara Institute of Science and Technology

Abstract

This paper proposes a "non-contact virtual clay modeling interface." We developed a prototype of a three-dimensional modeling system that allows the designer to deform the "virtual clay" by his or her hand movements. In our proposing method, the user's hand movement is observed using multiple cameras to estimate its position and orientation. The virtual clay is modeled by Catmull-Clark subdivision surface. Using the estimated hand position and orientation, the virtual clay is deformed based on the free-form deformation technique.

1 Introduction

In the product design using a 3D CAD system, it is difficult to input a 3D free-form with a conventional pointing device such as a mouse or a trackball. There are many studies on virtual clay modeling interfaces, which performs the deformation of the flexible objects by designer's hand in virtual world to make a CAD system more intuitive.

Many of the conventional virtual clay modeling systems adopt data gloves or haptic devices as interface devices to measure the hand movements. Matsumiya et al. developed a 3D modeling interface, which can deform the virtual clay by hand, using VR technology and implicit surface[1]. In order to immerse oneself into the virtual environment, the user wore a data glove and a head mounted display with 3D trackers. McDonnell et al. developed a real-time sculpting system using a force-feedback device(PHANToM). The clay was modeled by the subdivision solid, and was deformed based on the physical behavior[2]. Debunne et al. also developed a deformation system of the elastic object using PHANToM[3].

However, these contact type devices utilized in these researches reduce the designer's maneuverability. In addition, the data gloves and force-feedback devices are quite expensive. Therefore, we consider that non-contact type devices are suitable for the interface of a practical 3D modeling system. In this paper, we propose a non-contact virtual clay modeling interface using camera images. The aim of our research is to develop a user interface of a 3D design system which can be used practically. Our proposing method has two advantages. The first is that it provides a non-contact interface. The second is that our system uses only normal CCD cameras that are commonly used and are not expensive.

This paper is organized as follows. Section 2 describes the hand representation and presents the outlines of the hand position and orientation estimation algorithm. Section 3 describes the modeling of the virtual clay and the deformation method. Section 4 describes the implementation of the virtual clay modeling interface. Finally, we conclude this paper and describe future works in section 5.

2 Hand Posture Estimation

We have proposed a hand pose estimation method using multi-viewpoint silhouette images[4]. The hand position and its orientation can be estimated simultaneously by this method. We have used the multi-viewpoint camera system for sensing the user's hand where the influence of the self-occlusion is smaller than that in the monocular or stereo camera systems. The basic idea of our method is fitting a 3-dimensional hand model into an observed data. Our method uses two kinds of hand models as follows.

Voxel model The voxel model is a hand pose representation as the occupied volume in 3D space. By integrating multiple silhouette images obtained from multi-viewpont camera system, the hand pose is reconstructed as the voxel model. This model is utilized as the observational data of the hand. Figure 1 shows an example of the voxel model.

Skeletal model The skeletal model is a parametric hand pose representation based on the joint angles. It consists of the links and the skin surface data as triangle patches. The skeletal model also represents the position where the hand exists in terms of the vertex coordinates of the triangular patches. Figure 2 shows the skeletal model with skin surface.

In this paper, we do not treat the hand shape whose fingers are bent. Thus, it is assumed that the hand pose does not change, and all fingers are kept straight. The algorithm which estimates the hand position and orientation using 3D model fitting is as follows. As shown in Figure 3, the center of gravity of the skeletal model is used as the origin of the coordinate system

*Address: 8916-5 Takayama-cho, Ikoma, Nara 630-0192 Japan. E-mail: {etsuko-u, hirota-k, yoshio, ogasawar}@is.aist-nara.ac.jp

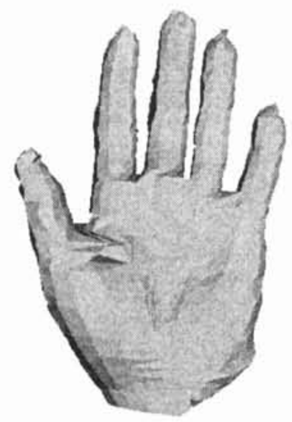


Figure 1: Voxel model   Figure 2: Skeletal model

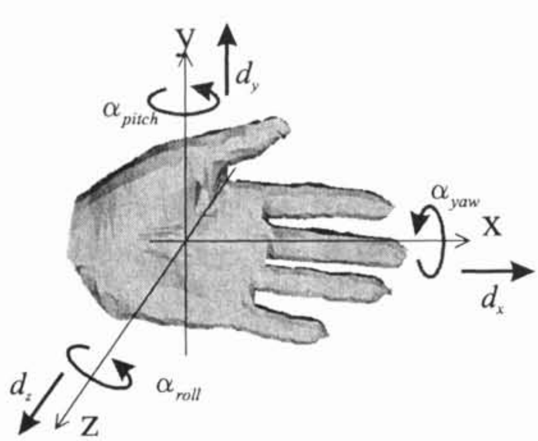


Figure 3: Coordinate system

in the fittig process. The direction of each axis is the same as the world coordinate system.

Firstly, the position of the center of gravity of the skeletal model is moved to the center of gravity of the voxel model. In this state, the patch of the skeletal model which locates outside the voxel model is defined as the "o-patch." For all o-patches, the force vector perpendicular to each o-patch is generated. Next, the moments are calculated using the force vector and the distance vector from joint position to each o-patch. From the summation of these forces and these moments for whole hand, the accelerations and the angular accelerations around each axis are calculated by Equation (1),(2).

$$m\ddot{\boldsymbol{x}} = \boldsymbol{F} \tag{1}$$

$$\boldsymbol{I}\ddot{\boldsymbol{\theta}} = \boldsymbol{N} \tag{2}$$

where $\ddot{\boldsymbol{x}}$ is an acceleration, $\boldsymbol{F}$ is the summation of the forces, m is a mass of the skeletal model, $\ddot{\boldsymbol{\theta}}$ is an angular acceleration, $\boldsymbol{N}$ is the summation of the moments, and $\boldsymbol{I}$ is the inertia matrix.

Both the translation and the rotation of the hand are performed simultaneously based on $\ddot{\boldsymbol{x}}$ and $\ddot{\boldsymbol{\theta}}$. In Figure 3, $\boldsymbol{\alpha}$ is a rotation angle and $\boldsymbol{d}$ is a translation distance. The position and orientation of the skeletal model are recursively changed, until the skeletal model is completely included in the voxel model.

3 Virtual Clay Model

3.1 Representation of Virtual Clay using Subdivision Surface

In this modeling system, we use the surface model based on the subdivision suface as the "virtual clay". This surface model is one of the parametric representation of the 3D object. There are two advantages in using the subdivision surface compared with conventional parametric representations. One is that the subdivision surface can define smooth surface of a polyhedron using only one surface. In order to represent a whole surface of a polyhedron, a conventional parametric surface, such as B-spline or NURBS, requires smooth connection of multiple surfaces. When the connection of multiple surfaces is performed, a procedure for keeping continuity is required. Therefore, it is difficult to smoothly deform the polyhedron in real-time. The other advantage is that the subdivision surface can change the smoothness of the surface by changing the number of the subdivision process.

We describe the polyhedron representation of the 3D object by the Catmull-Clark subdivision surface[5]. It represents the polyhedron by the quadrate patches, each of which is called the "face." In the initial state, the object is represented using the coarse faces, which is called the "initial polyhedron." The subdivision process generates three kinds of the vertices from initial polyhedron as follows:

1. Face points($\boldsymbol{f}$): For each face, the face point is its centroid.

2. Edge points($\boldsymbol{e}$): For each edge, the edge point is the weighed average

$$\boldsymbol{e} = \frac{\boldsymbol{a}_{avg} + \boldsymbol{m}}{2} \tag{3}$$

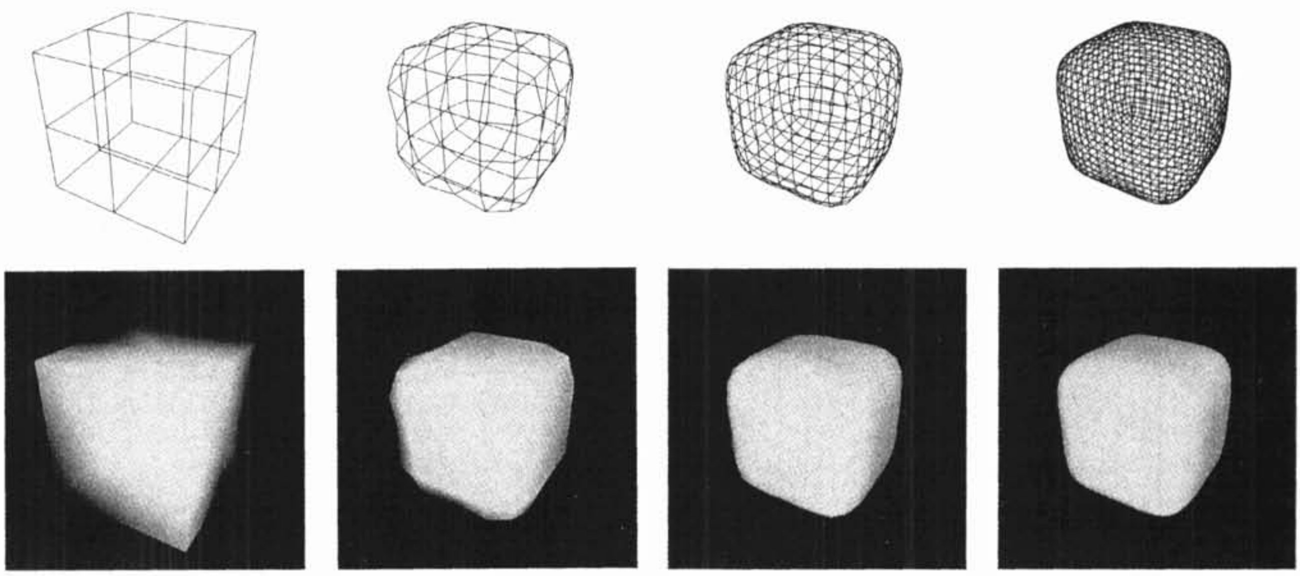
, where $\boldsymbol{a}_{avg}$ is the average of the centroids of faces that contain the edge, $\boldsymbol{m}$ is the midpoint of the edge.

3. Vertex points($\boldsymbol{v}$): For each vertex $\boldsymbol{p}$, the vertex point is the weighed average

$$\boldsymbol{v} = \frac{\boldsymbol{a}_{avg} + 4\boldsymbol{m}_{avg} + \boldsymbol{p}}{6} \tag{4}$$

where $\boldsymbol{a}_{avg}$ is the average of the centroids of faces that contain the vertex, $\boldsymbol{m}_{avg}$ is the average of the midpoints of the edge that contain the vertex, $\boldsymbol{p}$ is the position of the vertex.

The new polyhedron is then assembled as follows. Face points are connected to the edge points of edges that define the face, and edge points are connected to the vertex points of the vertices that define the edge. A patch is subdivided into four new patches by these processes that are performed recursively.



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

Figure 4: Subdivision surface: (a) the initial polyhedron, (b) after one subdivision step, (c) after two subdivision steps, (c) after two subdivision steps

The subdivision process that is described above can be represented as a matrix-vector multiplication.

$$\boldsymbol{d} = \boldsymbol{A}\boldsymbol{p} \tag{5}$$

where $\boldsymbol{p} = [p_1, p_2, \ldots, p_n]^T$ is the set of the vertex positions of the initial polyhedron, $\boldsymbol{d} = [d_1, d_2, \ldots, d_m]^T$ is the set of the vertex positions of the subdivided polyhedron, and $n < m$. $\boldsymbol{p}$ is called the "control points" and $\boldsymbol{d}$ is called the "data points". $\boldsymbol{A}$ is a $m \times n$ weighting matrix which represents the rule of the subdivision. Figure 4 shows the examples of the subdivided cube.

3.2 Direct Manipulation of Virtual Clay

When the positions of data points are moved by the estimated hand position, the whole surface shape of the object is recalculated. The deformation method that can perform a smooth deformation by Hsu's algorithm[6] is as follows. When the set of data points $\boldsymbol{d}$ is moved to $\boldsymbol{d}_{new}$ by the user's operation, the control points $\boldsymbol{p}_{new}$ which can generate the data points $\boldsymbol{d}_{new}$ is obtained as

$$\boldsymbol{p}_{new} = \boldsymbol{A}^{+}\boldsymbol{d}_{new} \tag{6}$$

where a pseudoinverse matrix $\boldsymbol{A}^{+}$ which is calculated using the least squares method.

$$\boldsymbol{A}^{+} = \boldsymbol{A}^{T}(\boldsymbol{A}\boldsymbol{A}^{T})^{-1} \tag{7}$$

Using $\boldsymbol{p}_{new}$, the whole surface of the 3D object $\boldsymbol{d}_{mod}$ is recalculated by Equation (5).

$$\boldsymbol{d}_{mod} = \boldsymbol{A}\boldsymbol{p}_{new} \tag{8}$$

$\boldsymbol{d}_{mod}$ obtained by Equation (8) is relatively smooth, in comparison with $\boldsymbol{d}_{new}$ obtained by user's operation.

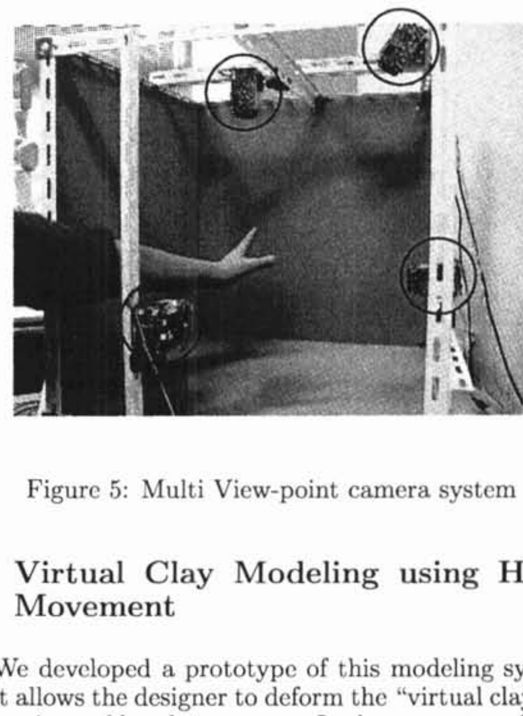


Figure 5: Multi View-point camera system

4 Virtual Clay Modeling using Hand Movement

We developed a prototype of this modeling system that allows the designer to deform the "virtual clay" by the estimated hand movement. In the present modeling system that we developed, it is required that the fingers are required to be stretched. Therefore, this system treats the hand as a spatula.

Figure 5 shows a hand pose observation system which uses the four cameras. The specification of our modeling system is:

- Workspace size: 60 cm × 60 cm × 60 cm
- Number of the CCD camera: 4
- Number of the control point of virtual clay: 28

- Number of the data point of virtual clay: 364
- Processing frequency: approx 3Hz
- CPU: Dual PentiumIII 1.0GHz

Using the estimated hand position and orientation, the user pushes in the data points from outside of the virtual clay. Then, the whole surface of the virtual clay is recalculated from the new positions of the data points using FFD. Figure 6 shows the example of the sequential deformation of the virtual clay.

5 Conclusion

In this paper, we proposed the non-contact virtual clay modeling interface using multi-viewpoint images, and developed a prototype system. The user's hand is utilized as the input device. The hand position and orientation in the real world is estimated using multi-viewpoint camera system and is directly utilized as the hand position and orientation in the virtual world to deform the virtual clay. The virtual clay was modeled by the Catmull-Clark subdivision surface. The developed interface can deform the virtual clay by directly manipulating the data points. At present, this system cannot perform the smooth deformation of the virtual clay, because the processing speed is not fast enough. In order to perform the complicated deformation of the virtual clay using estimation of hand shape, the processing speed should be improved.

As future works, we are going to deform the virtual clay using the finger movements. Our goal is to build a natural interface for a 3D modeling system using this proposed method.

References

[1] Masatoshi Matsumiya, Haruo Takemura, Naokazu Yokoya. "An Immersive Modeling System for 3D Free-form Design Using Implicit Surfaces". In *Proc. of Intn' Conf. Siggraph 2000*, pp. 1989–1996, 2000.

[2] Kevin T. McDonnell, Hong Qin. "Dynamic Sculpting and Animation of Free-form Subdivision Solids". In *Proc. of Intn' Conf. Computer Animation 2000*, pp. 126–133, 2000.

[3] Gilles Debunne, Mathieu Desbru, Marie-Paule Cani, Alan H. Barr. "Dynamic Real-Time Deformations using Space & Time Adaptive Sampling". In *Proc. of Intn' Conf. Siggraph 2001*, pp. 31–36, 2001.

[4] Etsuko Ueda, Yoshio Matsumoto, Masakazu Imai, Tsukasa Ogasawara. "Hand Pose Estimation Using Multi-Viewpoint Silhouette Images". In *Proc. of Intn' Conf. Intelligent Robots and Systems 2001*, pp. 1989–1996, 2001.

[5] Edward Catmull, James Clark. "Recursively Generated B-spline Surfaces on Arbitrary Topological Meshes". *Computer Aided Design*, Vol. 10, No. 6, pp. 350–355, 1978.

[6] William M Hsu, John F. Hughes, Henry Kaufman. "Direct Manipulation of Free-Form Deformations". In *Proc. of Intn' Conf. Siggraph 1992*, pp. 177–184, 1992.

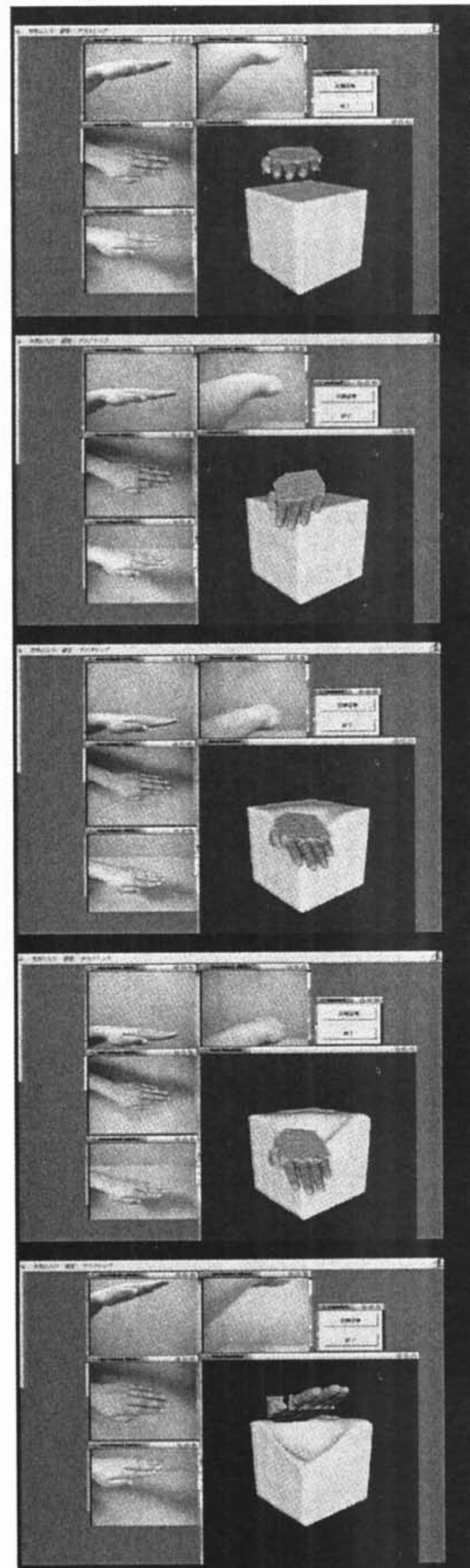


Figure 6: Deformation of virtual clay