A 3D Vision-based Interaction Method with a Simplified Hand Model

Minkyung LEE, Sehwan KIM, and Woontack WOO

GIST U-VR Lab., {mlee,skim,wwoo}@gist.ac.kr

Abstract In this paper, we propose a 3D vision-based interaction method with a simplified hand model. The proposed method consists of three main techniques; 1) disparity estimation using segmented hand images from a background scene to get 3D information in real-time, 2) structural modeling of user's hand with the anatomical structure to guarantee various interaction modes, and 3) model-to-model collision detection between augmented virtual objects and user's fingertips to make collision detection better. The proposed method, which provides a solution for occlusion problems, can supply more realistic interactive AR environment to a user by exploiting 3D information of the user's hand. The simplified hand model, constructed only with the hand images, allows a user to freely interact with augmented virtual objects. Thus, various kinds of realistic/immersive interaction modes are available. The proposed method can be applied to design various applications, such as education, entertainment, etc., with natural and accurate interaction in AR environment.

1 Introduction

Augmented Reality (AR) is a technology which provides improved immersion and reality through seamless connection between real and virtual worlds to a user [1]. Especially, interaction techniques have played an important role in AR. Vision-based interaction using user's hands is essential in allowing natural interaction with augmented virtual objects for a user.

In the case of vision-based interface using user's bared hand, we should solve the occlusion problem, in which augmented virtual objects occlude user's hands. It allows a user to have more natural AR view. Additionally, we should consider collision detection based on user's hand model to allow users to interact with augmented virtual objects accurately.

Skeletal models are one class of the 3D hand models [2]. It constructs a hand model using the reduced set of equivalent joint angle parameter with segment length instead of using all parameters of a volumetric model. A user's hand with several bones is complex to be modeled. However, we can have an estimated hand model with the simplification of anatomical structure using serial links [3]. It can be considered as a proper method for modeling user's hand in real-time for an AR system. Geometric structure of user's fingers and palms can be represented with cylinders and rectangular parallelepipeds, respectively [4][5]. The geometric structures are described with small number of parameters of simple primitives. A number of simple models, however, constitute the hand model, total complexity is still too high. Additionally, it is hard to detect all finger joints and all joint angles with previous skeletal models. Thus, a novel way for a simplified user's hand model should be proposed for vision-based interaction.

In this paper, we propose a 3D vision-based interaction method with a simplified hand model. First, we apply a background training method to image sequences obtained from a stereo camera [6]. Information only about markers and user's hand can be extracted by applying background subtraction to sequences. By applying binary images of markers to the subtracted foreground, we can easily get the subtracted marker area. The remained foreground excluding markers is automatically considered as a user's hand. The silhouette of user's hand is extracted from the segmented image. The user's hand model comes from the mapping between the obtained skeleton and known hand model. Model-model collision between the hand model and augmented virtual objects will be detected with more accurate interaction results.

This paper is organized as follows. Overall algorithm and each component are described in Chapter 2, and implementation and experimental results are shown in Chapter 3. Finally, conclusion and future work are described in Chapter 4.

2 Interaction with a Simplified Hand Model

As shown in Figure 2, the proposed algorithm consists of three modules 1) disparity estimation for images subtracted from background, 2) simplified structural modeling of user's hand with the anatomical structure, and 3) model-model collision detection between augmented virtual objects and user's hand model.

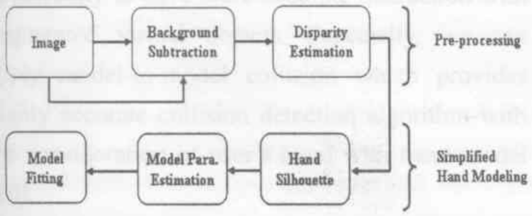


Figure 1. Procedure for simplified hand modeling

The first component of the proposed algorithm, disparity estimation module, is to provide natural and realistic AR-based interaction to a user. Overall procedure of disparity estimation is shown in Figure 2.

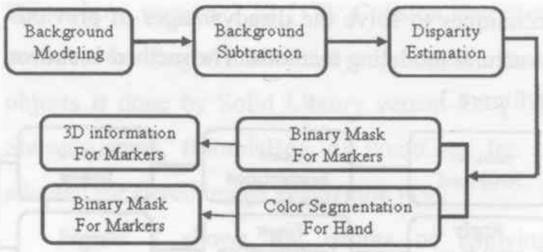


Figure 2. 3D information acquisition for a usage

With a stereo camera at a fixed position, which obtains user's environmental information, we can easily apply a background subtraction algorithm to the image. Thus, we obtain the extracted image which includes only markers and user's hand. Such background subtraction algorithm helps to get faster and more accurate disparity map [6]. The background subtraction method included in the proposed algorithm is divided into two phases; background modeling and foreground subtraction from the modeled background image.

In the phase of background modeling, we train background and make a reference image in RGB and in normalized RGB space, respectively. Then, in the foreground subtraction phase, we subtract the current image from the reference image. Finally we can have more accurate and faster algorithm by applying SAD (Sum of Absolute Difference)-based disparity estimation to the subtracted image [7]. In the case of markers, 3D information of markers can be easily obtained by applying binary image of markers, which is provided by ARToolKit [8].

In general, very high dimensional parameters are required to construct complex structures, such as user's hand, using simple geometrical structures. It is not so easy to measure all these high dimensional parameters only with computer vision techniques. We propose a simplified skeletal modeling method based on computer vision techniques to solve the disadvantages of previous structural modeling methods. The method is shown in Figure 3.

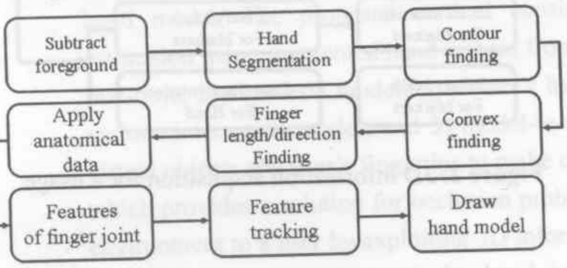


Figure 3. Conceptual diagram for skeletal hand modeling

3D information of user's hand are easily obtained by using the silhouette of a subtracted hand image. Convexes and concaves are obtained from user's hand silhouette. Each finger length and finger direction are calculated by connecting the center point between two concaves and the fingertips, shown in Figure 4.

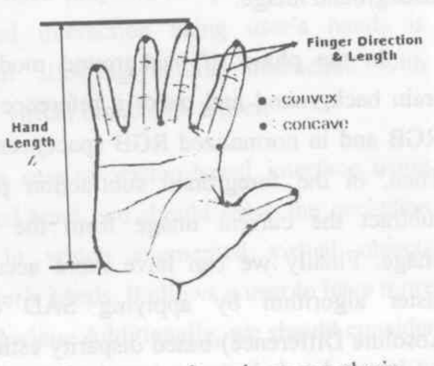
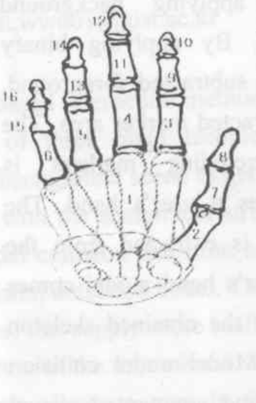


Figure 4. Parameters from image analysis

We refer statistically calculated anatomical data to calculate relative link size between each joint to generate simple and general vision-based hand model. Anatomical data of each joint are given in a normalized form related to the user's hand length, as shown in Figure 5 [9]. With the data, we can estimate each joint from the user's hand silhouette.

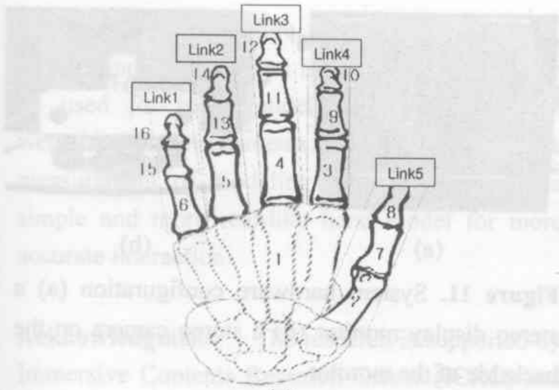


Link No.	L	Link No.	L
2	0.260	10	0.121
3	0.237	11	0.173
4	0.254	12	0.121
5	0.243	13	0.168
6	0.196	14	0.121
7	0.173	15	0.121
8	0.150	16	0.116
9	0.150		

(a) (b)

Figure 5. Parameters of the anatomical model (a) links of hand model (b) sizes of links in the model

The anatomical model calculates lengths of each joint by multiplying the ratio S related to L, the length from the wrist to the tip of middle finger. However, in most case is not so easy to find out the writ. Thus, we propose a hand model using modified parameters of the anatomical model. At first, we assume that user's fingers are fully shown. The assumption is essential to find out the convexes and concaves of user's hand. We already have the length of each finger, shown in Figure 4. By multiplying modified ratio to the calculated length of each finger, we can easily get the length of each joint. Modified anatomical concept and parameters are shown in Figure 6.



(a)

Link 1		Link 2		Link 3		Link 4		Link 5	
16	0.268	14	0.227	12	0.221	10	0.238	8	0.257
15	0.279	13	0.316	11	0.316	9	0.295	7	0.297
6	0.453	5	0.457	4	0.464	3	0.467	2	0.446

(b)

Figure 6. Modified hand model (a) modified concept of anatomical structure (b) parameters of the modified anatomical hand model

From the position and the direction of joints, a rough hand model is constructed with the combination of cylinders and hyper-rectangles. We assume that a user's hand consists of fingers and palm. The fingers are represented with a cylinder, and the palm is modeled with hyper-rectangle. Using the KLT tracking methods, we can continuously track and update 3D information of the joints. Thus, the deformation of user's hands can be reflected to the simplified hand model on the fly.

The generated hand model gives an opportunity to have more accurate interaction with augmented virtual objects. Especially, we can apply model-to-model collision which provides highly accurate collision detection algorithm with the consideration of user's hand with hand model [10].

3 Experimental Results

In this chapter, we show implementation of the proposed algorithm and applications. The proposed algorithm is implemented with OpenCV library version 3.0 [11]. Acquisition and tracking of binary image of markers is done with ARToolKit Directshow version 2.52 [12]. Collsion detection between the finger model and augmented virtual objects is done by Solid Library verson 3.5 [13]. Stereo camera, BumbleBee, of PointGrey Inc is adopted for stereo image acquisition [14].

Figure 7 shows the results of applying background subtraction.

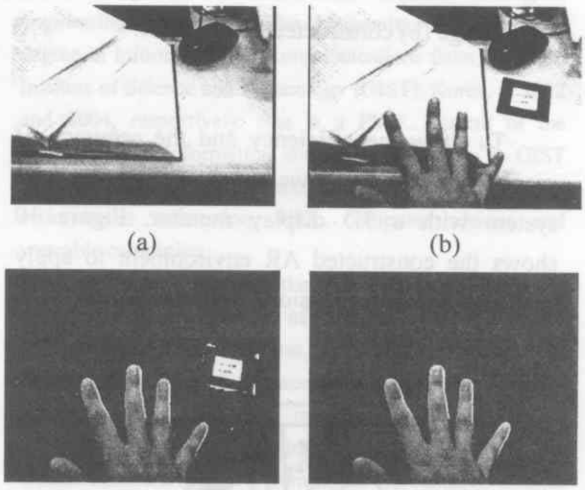
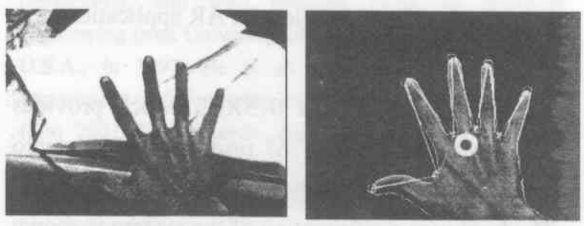


Figure 7. Foreground segmentation (a) background training (b) background subtraction (c) foreground (d) segmented hand

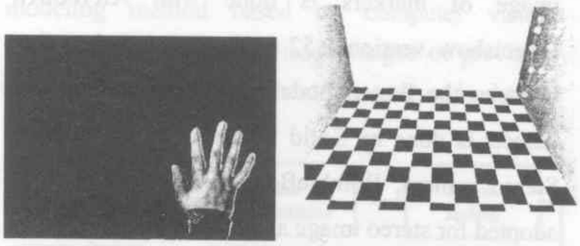


(a) (b)

Figure 8. Detected convexes and concaves (a)

original view (b) Detected convexes and concaves

With the segmented hand, we detected its convex points and concave points as well. Extracted points are shown in Figure 8. And, Figure 9 shows user's hand model using the subtracted hand image.



(a) (b)

Figure 9. Skeletal hand modeling (a) segmented hand image (b) constructed hand model.

To show the efficiency and the accuracy of proposed algorithm, we implemented an AR system with a 3D display monitor. Figure 10 shows the constructed AR environment to apply user's hand model to vision-based interaction.

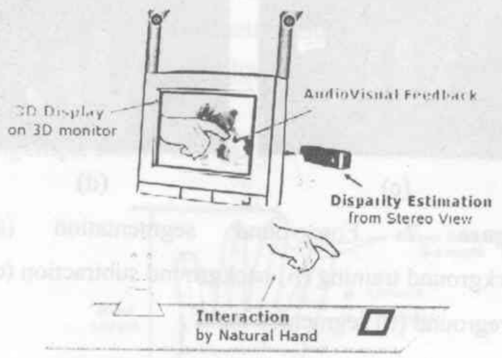
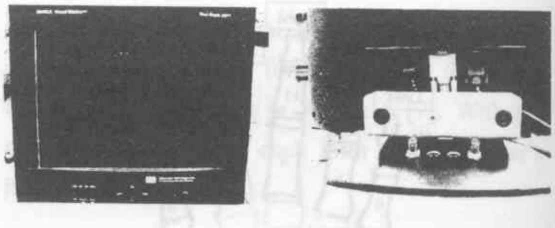


Figure 10. Configuration for AR applications

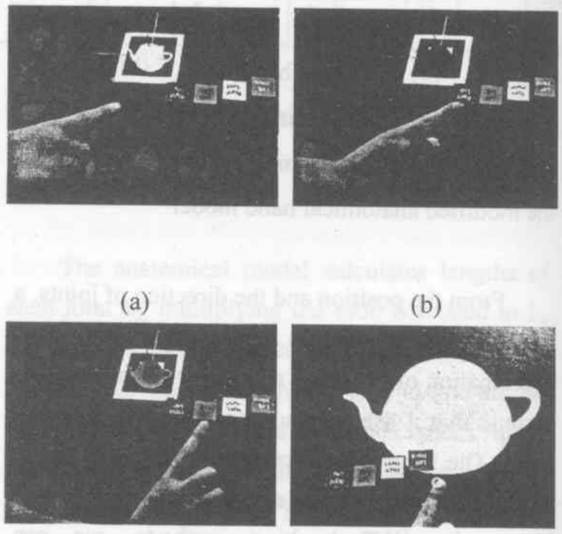
We adopted DTI's 015XLS which provides 3D display with 512 x 768 resolutions 0. A stereo camera is attached on the backside of the monitor. The hardware configuration of the system is shown in Figure 11.



(a) (b)

Figure 11. System hardware configuration (a) a stereo display monitor (b) a stereo camera on the backside of the monitor

Figure 12 shows a visual feedback as a result of collision between user's hand model and augmented virtual objects in implemented AR environment.



(c) (d)

Figure 12. Visual feedback of AR environment

4 Conclusions & Future Work

In this paper, we proposed a simplified hand model for AR environment. And we implemented an AR environment and experimented to show the usability of the proposed algorithm. We finally conclude the proposed algorithm provided user's hand model which can be fully applied to AR environment. As a future work, we will work on user's hand segmentation from a complex scene to be used for hand model-based interaction for wearable AR environments. Moreover, quantitative measurement of modeling accuracy will give a simple and more real-like hand model for more accurate interaction.

Acknowledgment The research is supported by Immersive Contents Research Center (ICRC) and Realistic Broadcasting Research Center (RBRC) in GIST.

References

[1] R. T. Azuma, "A Survey of Augmented Reality," Presence: Teleoperators and Virtual Environments, vol. 6, No. 4, pp. 355-385, 1997.

[2] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, No. 7, 1997.

[3] I.A. Kakadiaris, D. Metaxas, and R. Bajcsy, "Active Part-Decomposition, Shape and Motion Estimation of Articulated Objects: A Physics-Based Approach," Proc. IEEE C.S. Conf. Computer Vision and Pattern Recognition, pp. 980-984, 1994.

[4] A. C. Downton and H. Drouet, "Image Anaysis for Model-Based Sign language Coding," Progress in Image Anaysis and Processing II: Proc. 6th Int'l Conf. Image 'Analysis and Processing, pp. 637 – 644, 1991.

[5] M. Etoh. A. Tomono, and F. Kishino, "Stereo-Based Description by Generalized Cylinder Complexes From Occluding Contours," systems and Computers in Japan, vol. 22, no. 12, pp. 79-89, 1991.

[6] D. Hong and W. Woo, "Background Subtraction for a Vision-based Interface,"In the Proc. Of Parcific-Rim Conference on Multimedia, PCM03-CD Proceeding, pp. 1B3.3, 2003.

[7] O'Rourke and N.L. Badler, "Model-Based Image Analysis of Human Motion Using Constraint Propagation," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 2, pp. 522-536, 1980.

[8] Kato H. and Billinghurst, M., "Marker Tracking and HMD Calibration for a video-based Augemented Reality Conferencing System," Proc. International Workshop on Augmented Reality (IWAR99), 1999.

[9] F. Schuind et al., "Advances in the Biomechanics of the Hand and Wrist," Plenum Press, 1994.

[10] Gino van den Bergen, Collision Detection in Interactive 3D Environment, Morgan Kaufmann Publishers, 2004.

[11] Intel OpenCV library, http://www.intel.com/research/mrl/research/opencv/, 2004.

[12] ARToolkit library, ttp://www.hitl.washington.edu/research/shared_space/download/, 2004.

[13] Solid Collision Detection Library, ttp://www.dtecta.com/, 2004A.Field: Research on Subjects by Some methods, Trans. ISFA, vol.34, no.2, pp.83-120, 2001.

[14] Point Grey Inc., http://www.ptgrey.com/products/bumblebee/index.htm l

[15] Dimension Technologies Inc., http://www.dti3d.com/Products/dti_2015xls.htm, 2004.

LEE, Minkyung: received the B.S. degree in biomedical engineering from KyungHee University, and the M.S. degree in information and communications from Gwangju Institute of Science and Technology (GIST), Korea, in 2002 and 2004, respectively. She is a Ph.D. Student of the Department of Information and Communications in GIST from 2004. Her research interest includes augmented reality, vision-based interaction, HCI, computer vision, and wearable computing.

KIM, Sehwan: received the B.S. degree in electronic engineering from University of Seoul, and the M.S. degree in Department of information and communications from Gwangju Institute of Science and Technology (GIST), Korea, in 1998 and 2000, respectively. He is a Ph.D. candidate of the Department of Information and Communications in GIST from 2000. His research interest includes augmented reality, vision-based interaction, HCI, computer vision.

WOO, Woontack: received the B.S. degree in electrical engineering from Kyoungbook National University, and the M.S in electrical engineering from Pohang Institute of Technolgy (POSTECH), Korea, in 1987 and 1989, respectively, and then received Ph.D. in electrical engineering from University of Southern California (USC), U.S.A., in 1997. He is an assistant professor of the Department of Information and Communications in GIST from 2001. His research interest includes 3D computer vision, image processing and Mixed Reality, Ubiquitous/Wearable Computing, Human Computer Interaction.