A Stereo-Vision Approach for a Natural 3D Hand Interaction with an AR Object

Seonho Lee*, Junchul Chun*

*Department of Computer Science, Kyonggi University, San 94-6 Yiui-Dong, Yeongtong-Gu, Suwon, S. Korea

sunho36@naver.com, jcchun@kgu.ac.kr

Abstract—**Providing natural hand interaction between a virtual object and a user on Augmented Reality is a major issue to manipulate a rendered object in a convenient way. Conventional 2D image-based recognition and interaction technique in AR has a limitation to perform a natural interaction between the user and the virtual object. In this paper, we present a stereo-vision based natural 3D hand interaction with the augmented object. In the proposed 3D hand interaction approach, 3D hand location and finger direction can be easily obtained by using stereo-vision technique while user hand is approaching to the virtual object. Two types of hand manipulation for the augmented object such as the hand pointing and hand pinching are defined. The collision detection between user hand and the virtual object is determined by using a simple ray casting emitted from the user's finger-point against the virtual object in the case of hand pointing. In the hand pinching, the collision is occurred when the thumb and the index finger are approaching to the object and the degree of angle between two fingers becomes a predetermined value. From the experiments, the proposed 3D hand interaction method can control the virtual object in a natural way rather than using a vision-based 2D hand interaction since the stereo-vision technique can obtain the depth information from the AR environments.**

***Keyword*— Stereo-vision, Haptic interface, Augmented Reality, Collision detection, Hand detection**

I. INTRODUCTION

Recently one of the main issues in Augmented Reality (AR) is how to interact the overlaid virtual objects in a convenient fashion by users. Newly introduced tracking and interaction methods in AR allow users to work with and examine the real physical world, while controlling augmented objects in the system more feasible fashion. In general AR can be classified into two categories such as marker-based AR and marker-less AR. In marker-based AR a specific marker is used for overlaying an object in the scene. Meanwhile, marker-less AR does not require the forethought of adding markers to a scene in order to render a virtual object. It uses a detected feature from the scene as a marker instead of using a predetermined specific marker. Regardless of their types of AR system, most of AR systems need various interactions between users and augmented object in many AR applications. Therefore, the major issue in AR is how to interact with a virtual model in dynamic or convenient way. Conventional vision-based interaction techniques in AR are based on 2D image analysis and recognition methods and have a limitation to obtain three dimensional information of the virtual object and user who participates in the interaction.

In this paper, we present a vision-based 3D hand interaction with a virtual object in AR system with a natural way. To develop the 3D hand interaction we adopt stereo-vision technique which can expand the dimension of the AR interaction from 2D to 3D and provide more intuitive interactions between the virtual object and users. In the proposed 3D hand interaction, a user can directly contact with a virtual object by simply detecting the collision between user's hand and the virtual object while the interaction is being processed.

The rest of the paper includes related work, the description of the proposed vision-based 3D interaction method, the experimental results and the conclusion and future works.

II. RELATED WORKS

Both marker-based and marker-less AR systems require some indication of where exactly the virtual objects should be augmented. This has been conventionally accomplished by AR markers such as ARTag[1] or ARToolkit[2] in marker-based AR. In mobile AR system ARToolkitPlus[3] is well-adopted because it provides an efficient management of the memory and fixed point unit computing by optimizing the libraries of ARToolkit.

In order to register a virtual object to a detected marker on the scene, we can use the vision-based tracking technologies such as feature-based and model based approach. The main idea underlying feature-based methods is to find a correspondence between 2D image features and their 3D world coordinate system. The camera pose can be found from projecting the 3D coordinates of the features into the 2D image coordinate along with minimizing the difference between their corresponding 2D features. The ARToolKit library utilizes the four corners of a

Manuscript received July 29, 2013. This work was supported by Basic Science Research program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(No. 2012006018).

Seonho Lee is a graduate student in the department of computer science at Kyonggi University, South Korea (Tel: +82-31-249-8947; e-mail: sunho36@naver.com).

Junchul Chun is currently a professor in the department of computer science and a principal investigator of GIP(Graphics and Image Processing Lab) at Kyonggi University, South Korea. (Tel: +82-31-249-9668; Fax: +82-31-249-8949; e-mail: jcchun@kgu.ac.kr).

square marker to the positions of the 3D object rendered. Tracking algorithms for non-square visual markers such as ring shaped and circular shaped markers were also used[4]. Model-based methods tracking explicitly use a model of features of tracked objects such as 2D template object or CAD model.

Once a virtual object is registered on the marker, the users usually want to manipulate or interact with the augmented object. The most of introduced interaction techniques for AR applications allow end users to contact with virtual objects in an intuitive way. Tangible interface and tangible interaction metaphor have become one of the most frequently used AR interaction methods[5]. Tangible AR interactions leads to combining real object input with human gesture interaction and hand gesture interaction methods have been widely studied from the perspective of computer vision and AR[6].

In vision-based interaction, hand and fingertip tracking along with hand gesture recognition method are widely used to provide an easy way to interact with virtual object in AR. The approaches to use bare stretched human hand as a distinctive pattern instead of a marker for marker-less AR system are introduced[7]-[9]. In their work, 6-DOF camera pose was estimated by tracking fingertips and virtual objects are augmented on hand coordinate system. However, the limitation of these approaches is the inspection of the object is hindered when the fingertips are occluded by themselves when the hand is flipping or moving. Chun and Lee[10][11] presents a real-time vision-based approach to manipulate the overlaid virtual objects dynamically in a marker-less AR system using bare hand with a single camera. In their approach, the left bare hand is considered as a virtual marker in the marker-less AR and the right hand is used as a hand mouse. Thus the manipulation of the virtual objects on the marker-less AR system can be dynamically obtained and a vision-based hand control interface which exploits the fingertip tracking for the movement of the objects and pattern matching for the hand command initiation is developed. However, all of these interactions are based on vision-based 2D interaction with the augmented object in AR. Simple 3D interaction in AR[15] was presented however in this work we provide various scenarios of 3D interaction with more feasible fashion in actual AR applications.

III. Proposed Approach

A. *Manipulation of a Virtual Object in AR*

The interaction between a human and an augmented object in AR is by manipulating the virtual object by actual human hand illustrated in Fig 1. In this work, we introduce the two types of hand interaction by hand pointing and hand pinching.

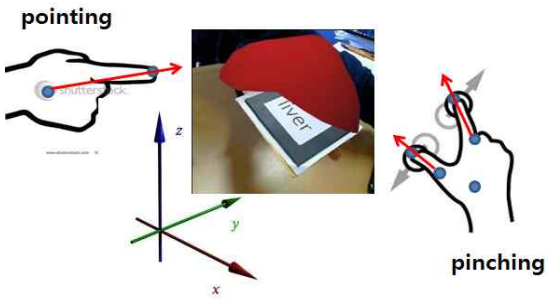


Fig. 1 Manipulating a virtual object in AR by hand pointing and pinching interaction

Fig 2 shows the overall steps of the proposed 3D hand interaction method, which consists of three major phase. In the first phase, the human hand is detected from input video image using the skin color model and image segmentation. In the second phase, the 3D locations of fingertip, the center of the palm and the center of the marker are evaluated using the disparity map of the stereo-vision. In the final stage, the user can interact with an augmented object by detecting collision between the human hand and the object.

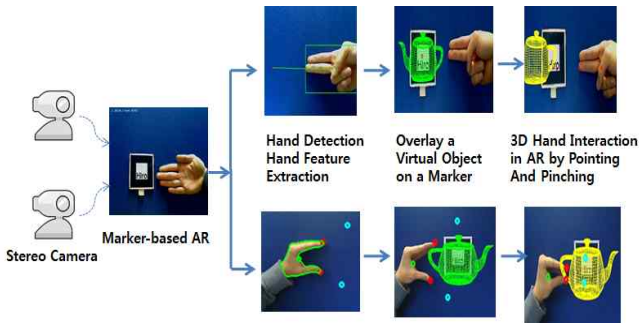


Fig. 2 Steps for stereo-vision based 3D hand interaction in AR

B. *Hand Detection Method*

For hand region detection like in our previous works, we use YC_bC_r skin color model which is proven to detect skin region effectively to segmenting hand region[10][11]. Skin color model has been widely used for hand and face detection because the use of color information can simplify the task of hand localization in complex environments. Mostly the primary components of RGB are used for skin segmentation. Other models such as HSI, normalized RGB, and YC_bC_r etc. are used for the segmentation of skin-like region[12]. Even though, color information is an efficient tool for identifying skin region if the skin tone color can be properly adapted for different lighting conditions, it has some limitation since the skin color model is sensitive to the light source varies significantly or complex background.

As a skin color model we adopt the YC_bC_r since it is perceptually uniform and it is similar to the TSL space in terms of the separation of luminance and chrominance. It is known that the chrominance components of the skin color are independent of the luminance component. Thus, the normalized

RGB color model is transformed to YC_bC_r model by using following equation (1).

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.16874 & -0.3313 & 0.500 \\ 0.500 & -0.4187 & -0.0813 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \quad (1)$$

By disregarding the luminance component (Y), robustness of skin detection can be obtained in the case of variations lighting conditions. Thus, in this work we only utilize C_b, C_r values from samples of skin color pixels. For the hand region detection, the threshold values of each chrominance components are derived from the set of hand images. The average values of C_b and C_r for hand region obtained through the experiments are as follow

$$(128 \le C_b \le 170) \cap (73 \le C_r \le 158) \quad (2)$$

With carefully selected threshold range values, a pixel value is assigned to the hand region if its value meets the range. Fig 2 shows the hand region with C_b, C_r values from the RGB based input hand image. In order to enhance the segmented hand region by removing noise a morphological operator are applied to the segmentation result. The center of the hand can be extracted from the medial axis of the segmented hand region $I(P)$. The distance transform which is used to compute the medial axis of the segmented hand[13] is used to obtain the single connected hand. The distance transform (DT) is the transformation that generates a map ***D*** whose value in each pixel p of the segmented region O is the smallest distance from this pixel to the background O^c. The distance map ***D*** can be defined as follow

$$D(p) = \min\{d(p,q) | q \in O^c\} = \min\{d(p,q) | I(q) = 0\} \quad (3)$$

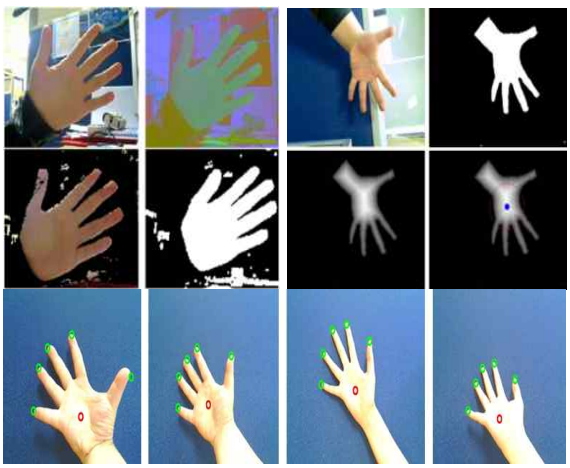


Fig 3. Steps for hand segmentation: (a) C_b C_r converted input image and segmentation of hand(left) (b) distance transform of the hand segmentation and the center of the hand(right) (c) Finger tip detection (bottom)

C. *Stereo Vision Technique*

In this work, we use stereo camera to get the 3D coordinates of the hand and the marker itself for detecting any possible collision between the user and the virtual object while manipulating the augmented object. Using two different views from the same scene then we can estimate the 3D coordinates of any point in the scene by finding the position of that point in the left image and in the right image and then apply some trigonometry. To get the accurate 3D coordinates of the points such as fingertip or the center of the marker, we calibrate the stereo camera. The calibration can make that the match of a point in the left image will appear in the exact same line on the right image and it can also fix the possible distortion caused by the lenses.

An offline camera calibration is performed by use of a simple planar grid pattern of known size in the field of view. Fig 4 illustrates two grid patterns for left and right un-calibrated images and their rectified images after calibrating those images. We can see each correspondent point of the left and right images are same epipolar line after calibrating the stereo images.

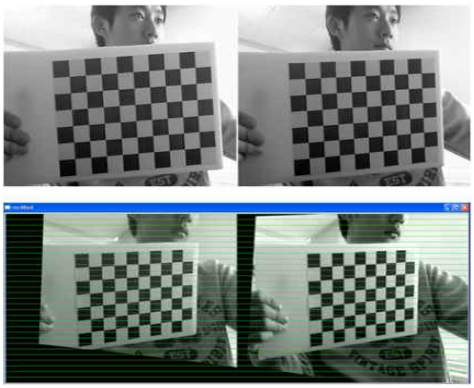
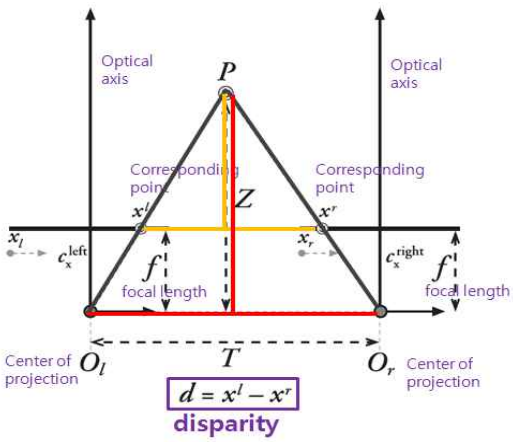


Fig 4. The rectified images from two images after calibration

Using the calibrated images, we can calculate the dense disparity map which provides the distance between the camera and the object. Fig 5 shows the geometry of the stereo vision and the relationship between disparity and camera distance.



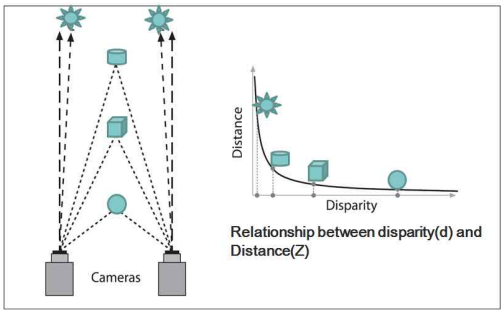


Fig 5. The geometry of stereo vision (top) and the relationship between disparity and camera distance (bottom)

Once the disparity (d) between x^l and x^r is calculated, the depth value (z) can be obtained by using T and the focal length f as follows:

$$z = f\frac{T}{d} \tag{4}$$

The calculated (z) value will be used for the (z) coordinate value of the objects such as the center of the hand or the marker in the marker-based AR system. Fig 6 shows a dense disparity map from two images taken from a left and right camera.

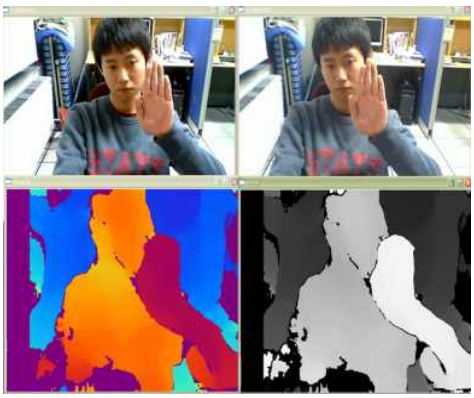


Fig 6. Dense disparity map from two input images

D. *Collision Detection Using Ray Casting*

It is natural that the collision between the human hand and the augmented object can be occurred during manipulating the virtual 3D object. In AR, however, the collision is occurred between a virtual and a real object thus the collision detection method may be different comparing with the ways in the real world. In this work, since we already have 3D information of the hand and the marker for the AR using stereo-vision technique, the 3D direction of the finger tip can be obtainable like Lee's work[14]. When both the calculated finger direction is aligned with the central position of the marker or the positions of augmented object and the hand is approaching to the object in the some distances, we can assume the collision between two objects.

Since the manipulation of an object can be done by two types of hand interaction i.e. pointing and pinching, the collision detection between the hand and a virtual object is also defined by two different approaches, respectively. Fig 7 illustrates the direction of the finger when we use two feature points (P_0, P_1) of the hand in the hand pointing.

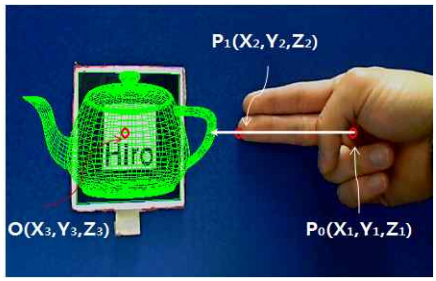


Fig 7. The direction of the finger pointing by simple ray casting

The direction of the finger can be easily calculated as a linear parametric function $f(t)$ between the centre of the hand $P_0(x_1, y_1, z_1)$ and the location of the fingertip $P_1(x_2, y_2, z_2)$ as follows:

$$\begin{aligned} f_x(t) &= x_1 + (x_2 - x_1)t \\ f_y(t) &= y_1 + (y_2 - y_1)t \\ f_z(t) &= z_1 + (z_2 - z_1)t \end{aligned} \tag{5}$$

The linear function $f(t)$ can be used for evaluating where the user is pointing at and the parameter t can control the magnitude of the function. For the collision detection, we can use finger ray casting which is known to be a fast for pointing at and selecting a distant object. The collision between the hand and the virtual object is detected when the ray touch the virtual object. In this work, we assume that two objects are collided in a certain range of distance (D) for example the parameter t of $f(t)$ is less than equal to 2 or the distance between the locations of the hand and marker is less than a predefined distance (ε). Three different reactions from the virtual object: pushing, pulling and changing the color of the virtual object are obtained when the collision between the user and the virtual object is detected.

Meanwhile, when the pinching is used the collision between the object and finger is determined by the angle between the thumb and the index finger as illustrated in Fig 8.

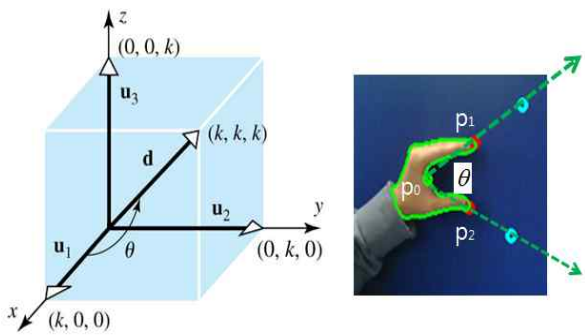
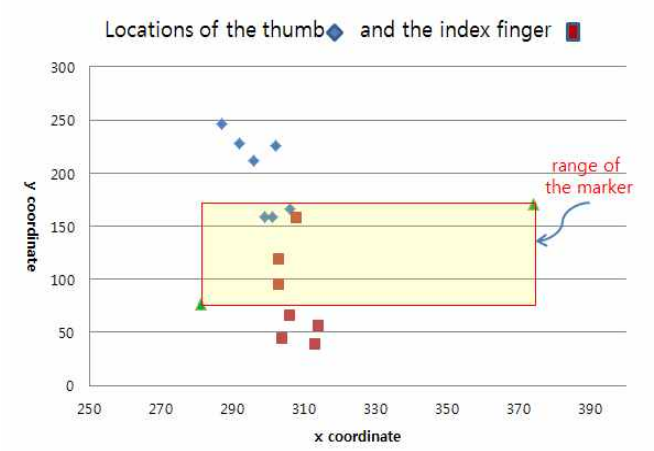


Fig 8. The angle of the thumb and a finger in pinching interaction

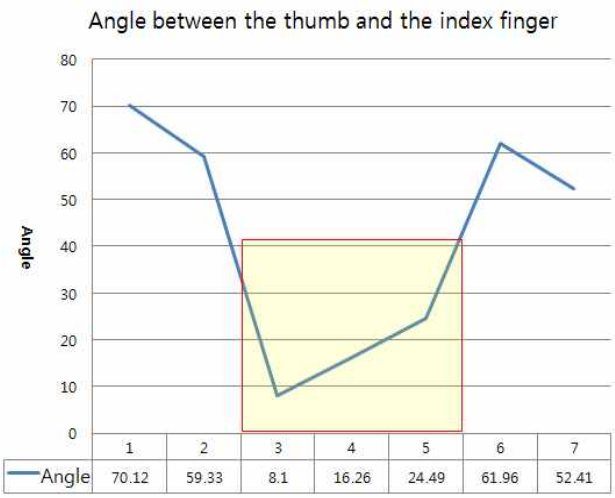
The degree of angle θ between $\vec{u} = (P_0, P_1)$ and $\vec{v} = (P_0, P_2)$ becomes as follow:

$$\cos\theta = \frac{\vec{u} \bullet \vec{v}}{\|\vec{u}\|\|\vec{v}\|} \quad (6)$$

Fig 9 (a) illustrates the locations of the thumb and the index fingers during pinching interaction in a sequences of image frames between the user and a virtual object when the z-coordinates of the fingers and the augmented object are same locations. The collisions between the fingers and the object are detected in the range of the marker. In this work, it is assumed that the collision between the finger and the virtual object is occurred when the thumb and the index finger are approaching to object and the degree of angle between two fingers is located in the range of 0° 0 to 42° during pinching the virtual 3D object as illustrated in Fig 9 (b).



(a) Locations of the thumb and the index finger in a sequence of image frames during pinching interaction



(b) Angle between the thumb and the index finger during pinching interaction

Fig 9. Collision detection in pinching interaction.

IV. Experimental Results

The proposed 3D hand interaction supports four different manipulations of the augmented object in AR. In this experiment, we use ARTookit to register the virtual object on a real scene. One of the difficulties in developing AR applications is the tracking the users viewpoint however ARToolkit can solve this problem using vision algorithm. In the experiment we use two Logitech Quick Cameras to obtain stereo images.

Fig 10 illustrates the sequence of hand images and detected 3D locations of the centre of hand, fingertips and finger ray using stereo-vision technique and our proposed method.

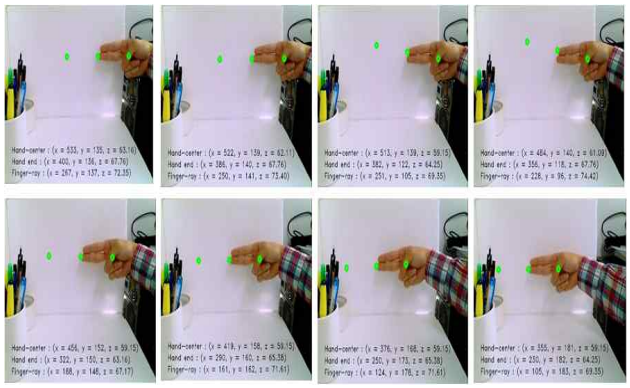
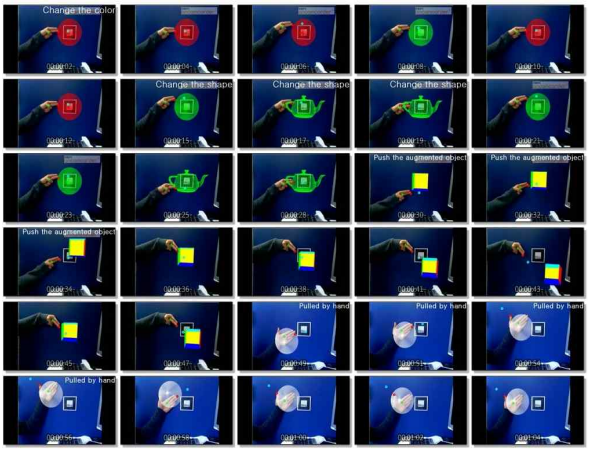


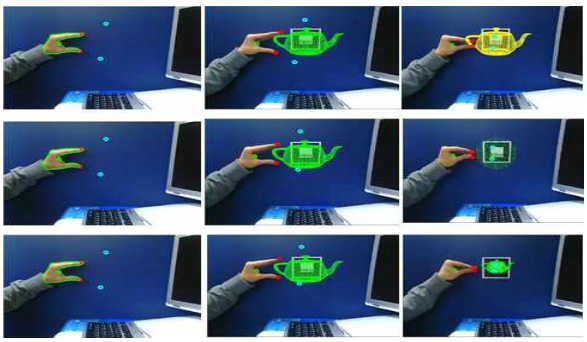
Fig 10. 3D locations of the center of hand, fingertip and its ray.

Based on the hand pointing and the hand pinching interaction with a virtual object in AR, we can produce various 3D interaction results. In the hand pointing interaction, 4 different reactions are generated when the hand is collided with the augmented object as illustrated in Fig 11(a). The first interaction is to change the colour feature of the virtual object when a user touches the augmented object. The second interaction is changing the shape of the object. The third interaction is pushing the augmented object from the currently registered position on the marker. While applying this interaction, the rendered object is translated from the current location of the marker to the opposite direction and got back to its original position when user's hand is moved away from the current marker. Final interaction is pulling the augmented object and locating the object on the hand. Like pushing, the object can be relocated into original position when the human hand is disappeared from the scene.

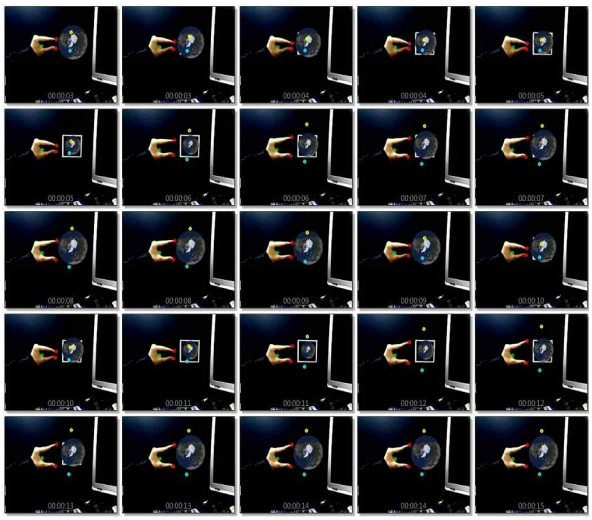
Meanwhile, in the hand pinching interaction, the color, shape and size of the virtual object is instantly changed when the object is touched by pinching as illustrated in Fig 11(b). The sequence of the image in Fig 11(c) is the size variation of the virtual 3D object when a user manipulates the object during pinching interaction.



(a) The hand pointing interaction



(b) Three different reaction with the hand pinching interaction



(c) The size variation of the object by pinching interaction

Fig 10. Manipulation of virtual object using the hand pointing and pinching interaction

CONCLUSION

In this paper we present a stereo vision-based natural 3D hand interaction with an augmented object in Augmented Reality System. In traditional AR interaction using 2D input image, it have been limitations to provide a natural interaction between the user and the augmented object because of the lack of *Z* information of the user and the virtual object. The proposed stereo-vision technique can overcome this disadvantage of 2D image and support more feasible manipulation of the virtual object in AR. In the proposed 3D hand interaction, the 3D information of the hand and the virtual object was utilized to detect the 3D collision between two models when the hand was approaching to the virtual object. The user interaction can change the features of the object such as color and shape. In addition, the user can change the geometric locations of the object by pushing or pulling the object from the current 3D locations of the object. This simple collision detection can make the 3D hand interaction come true in AR applications.

As for the future work, we will study to combine the proposed stereo-vision technique with actual haptic device for manipulating a virtual object in AR.

REFERENCES

[1] M Fiala. ARTag, a fiducial marker system using digital techniques, In CVPR 05, (2005), Vol 2., pp. 590-596.

[2] H. Kato and M. Billinghurst, Marker tracking and HMD calibration for a video-based augmented reality conferencing system, In IWAR ` 99(1999), pp. 85-94.

[3] D. Wagner and D. Schmalstieg ARToolKitPlus for Pose Tracking on Mobile Devices, In Procedings of 12th Computer Vision Winter Workshop'07 (2007), pp. 139-146.

[4] Cho, Y; Lee, J. & and Neumann, U. A multi-ring fiducial system and an intensity-invariant detection method for scalable augment reality, In *IWAR* `98, (1998). pp. 147-156..

[5] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto and K. Tachibana, Virtual object manipulation on a table-top AR environments, In ISAR `00, (2000), pp. 111-119.

[6] Lee, W. & Park, J. Augmented foam: A tangible augmented reality for product design. In ISMAR `05, (2005), pp. 106-109.

[7] T. Lee and T. Hollerer, Hand AR: Markerless Inspection of Augmented Reality Objects Using Fingertip Tracking, In Proc. IEEE ISWC `07,(2007), pp. 83-90.

[8] T. Lee and T. Hollerer, "Initializing Markerless Tracking Using a Simple Hand Gesture," In Proc. ACM/IEEE ISMAR `07, (2007), pp. 1-2.

[9] T. Lee and T. Hollerer, "Hybrid Feature Tracking and User Interaction for Markerless Augment Reality," In IEEE Int'l Conference on Virtual Reality`08 (2008), pp. 145-152.

[10] Lee, BS and Chun, JC, Interactive manipulation of augmented objects in marker-less AR using vision-based hand mouse, In Int'l Conference on Information Technology(ITNG), (2010), pp. 398-403.

[11] Chun, J.C and Lee, B.S, Dynamic Manipulation of a Virtual Object in Marker-less AR system Based on Both Human Hands, Transactions on Internet and Information Systems, (2010), Vol. 4, No.4, pp. 618-632.

[12] K. Min and J. Chun, A nonparametric skin color model for face detection from color images, PDCAT `04, pp. 115-119 , 2004

[13] G. Borgefors, Distance transformations in digital images, Computer Vision, Graphics and Image Processing,(1986), Vol. 34, pp. 344-371.

[14] M. Lee, R. Green, and M. Billinghurst, 3D Natural hand interaction for AR application, Image and Vision Computing New Zealand 2008,(2008), pp.1-6.

[15] JC Chun, SH Lee, A Vision-based 3D Hand Interaction for marker-based AR, International Journal of Multimedia and Ubiquitous Eng., Vol. 7, No. 3, (2012), pp. 51-57.

Seonho Lee is currently a graduate student in the department of computer science at Kyonggi University, South Korea. He received BS degree from Kyonggi University majoring computer science in 2012. His major research interests are augmented reality and haptic interaction in AR

Junchul Chun is currently a professor in the department of computer science and a principal investigator of GIP(Graphics and Image Processing Lab) at Kyonggi University, South Korea. He has been served as a chief research director at Contents Convergence Software Research Center supported by Geyonggi Regional Government in South Korea. He received BS from Chung-Ang University majoring computer science. He also received Ph.D and MS degrees of computer science and engineering from the University of Connecticut, U.S. respectively. His major research areas are vision-based interaction, augmented reality and computer graphics.