A Vision-based 3D Hand Interaction for Marker-based AR

Junchul Chun and Seonho Lee

Department of Computer Science, Kyonggi University
jcchun@kgu.ac.kr, sunho36@naver.com

Abstract

The hand interaction between a virtual object and human 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 natural 3D hand interaction with the augmented object by the use of stereo-vision technique. The proposed 3D hand interaction technique can obtain 3D hand position and finger direction while hand is approaching to the virtual object. 3D hand interaction can manipulate the virtual object in a feasible fashion. In addition, it can provide various interactions by detecting the collision between the 3D hand and the virtual 3D object on AR.

Keywords: *Augmented Reality, Hand Interaction, Stereo-Vision, Hand Detection, Collision Detection*

1. Introduction

Augmented Reality (AR) is a technology which allows virtual graphic models to exactly overlay physical objects in real time. Recently developed 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. Generally 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 systems, 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 feasible 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.

2. Related Work

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. Recently in developing mobile AR system, ARToolkitPlus [3] which provides an efficient management of the memory and fixed point unit computing by optimizing the libraries of ARToolkit is well-adopted.

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 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 introduced [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 investigated from the perspective of computer vision and Augmented Reality [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, 8, 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. To achieve more natural interaction with the virtual object, it is necessary to expand the dimension of the interaction.

3. Vision-based 3D Hand Interaction

3.1. Overview of the Proposed Method

Figure 1 shows the overall procedure of the proposed stereo-vision based 3D hand interaction, which consists of three major phase. In the first phase, the human hand is detected from input video image by use of the skin color model and image segmentation used in Chun's approach [11, 12]. In the second phase, the 3D positions of the hand such as fingertip, the center of the palm and the center of the marker are evaluated using disparity map of the stereo-vision. In the third phase, the user can interact with a virtual object by detecting collision between the human hand and the object. The user interaction can change the features of the object such as color and shape. In addition, it can change the geometric location of the object by pushing or pulling the object from the currently located position.

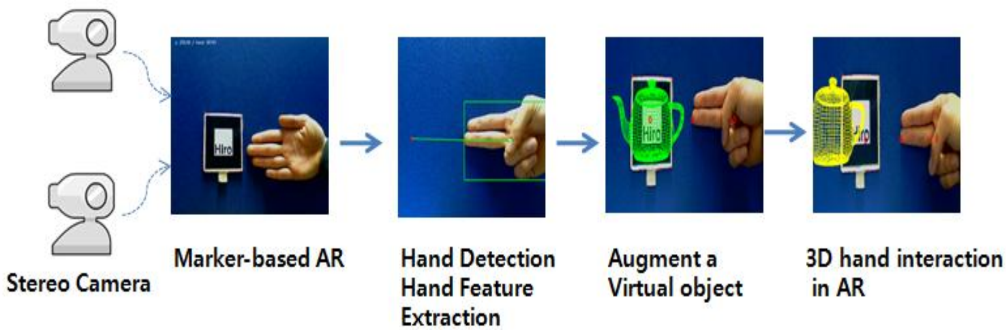

Figure 1. The Overall Procedure of the Proposed Method

In this paper, we utilize ARTookit to register the virtual object on a real scene. ARTookit is a software library for building AR applications such as overlaying a virtual imagery on a marker. One of the difficulties in developing AR applications is the tracking the users viewpoint however ARToolkit can solve this problem using vision algorithm. This library also provides a numbers of basic fiducial markers as illustrated in Figure 1 and uses a fast algorithm to track the location of the markers.

3.2. Stereo Camera Calibration

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. Using the calibrated images, we can calculate the dense disparity map which provides the distance between the camera and the object. Figure 2 shows the geometry of the stereo vision.

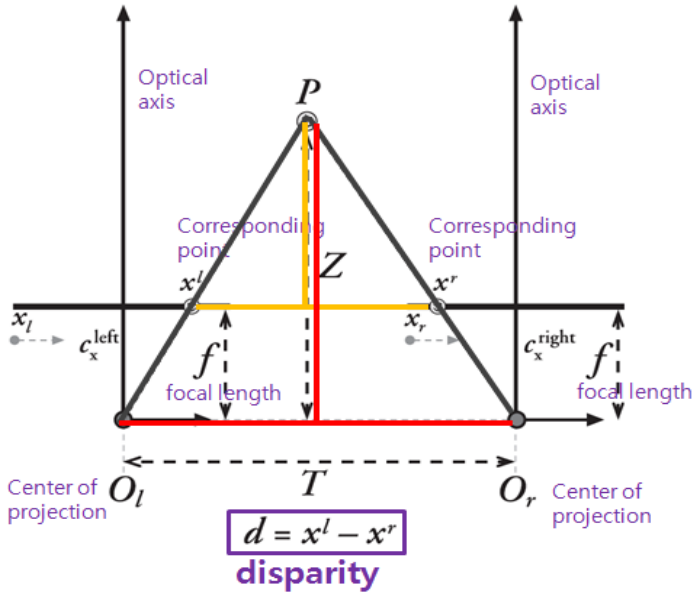

Figure 2. The Geometry of Stereo Vision

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:

$$z = f\frac{T}{d} \tag{1}$$

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.

3.3. 3D Interaction

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 happened between a virtual object and a real object thus the collision detection approach may be different comparing with the ways in the real world. In this proposed method, 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 pointing can be obtainable like Lee's work[13]. When both the calculated finger point 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. Figure 3 illustrates the direction of the finger when we use two feature points (P_0, P_1) of the hand.

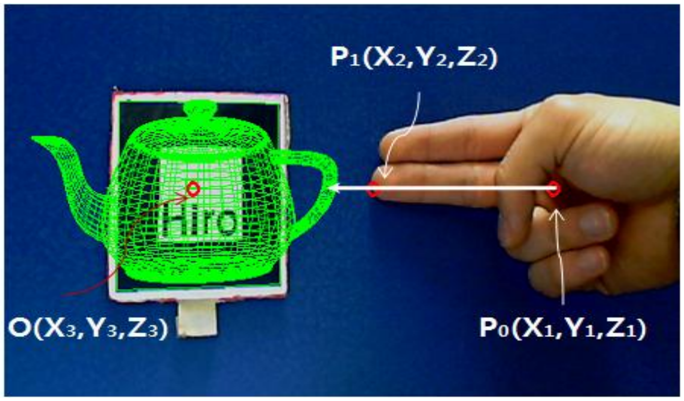

Figure 3. Direction of the Finger Pointing by a Simple Ray Casting

The direction of the finger can be easily calculated as a linear parametric function $f(t)$ between the center of the hand $P_0(X_1,Y_1,Z_1)$ and the position of the fingertip $P_1(X_2,Y_2,Z_2)$ as follows:

$$
\begin{aligned}
f_x(t) &= X_1 + (X_2 - X_1)t \\
f_y(t) &= Y_1 + (Y_2 - Y_1)t \\
f_z(t) &= Z_1 + (Z_2 - Z_2)t
\end{aligned}
\qquad (2)
$$

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. In this work we make the height of the virtual object when it is augmented on the marker $10cm$. Two different reactions from the virtual object such as changing color or shape of the virtual object are performed when the collision between the user and the virtual object is detected.

4. Experimental Results

The proposed 3D hand interaction supports four different manipulations of the augmented object in AR. The fist interaction is to change the color of the virtual object when user hand touches the virtual object and the second interaction is changing shape of the object as illustrated in Figure 4. In the experiment we use two Logitech Quick Cameras to obtain stereo images. The augmented objects in this work are designed to have $10cm$ height.

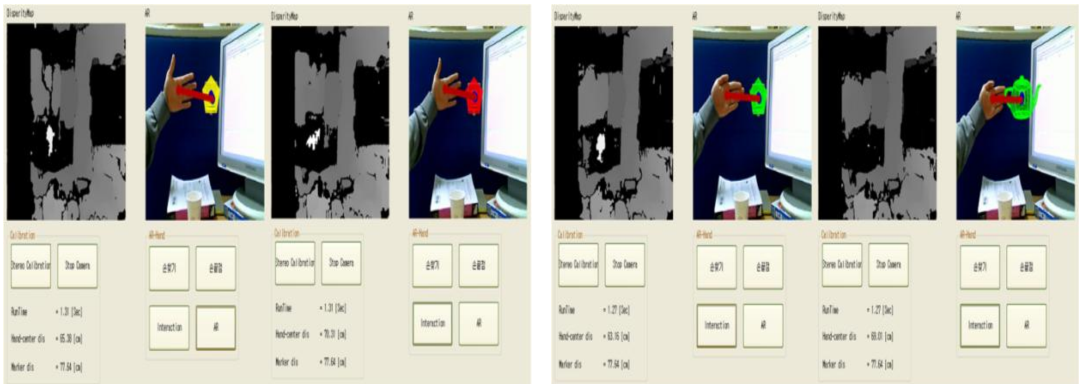

Figure 4. Direction of the Finger Pointing by a Simple Ray Casting

5. Conclusion

In this paper we present a stereo vision-based 3D hand interaction with an augmented reality on marker-based AR. AS described, there have been some limitations to make natural interaction between the user and the virtual object by simply using 2D images because the z-coordinate information of user and the virtual object are not available in the sequential input video images. The proposed stereo-vision technique can overcome the disadvantage of 2D image to make dynamic manipulations of the virtual object in AR. In the proposed 3D hand interaction method, the 3D information of the hand and the virtual object was used to detect the collision between two models when the user's hand was approaching to the virtual object. The collision detection can make the 3D hand interaction come true in AR applications. From the experiments we show the proposed 3D interaction can control the virtual object efficiently when the collision is detected.

Acknowledgements

This work was supported by Kyonggi University Research Grant 2011.

References

[1] M. Fiala, "ARTag, a fiducial marker system using digital techniques", In CVPR 05, vol. 2, **(2005)**, 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] Y. Cho, J. Lee and U. Neumann, "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] W. Lee and J. Park, "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] B. S. Lee and J. C. Chun, “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] J. C. Chun and B. S. Lee, “Dynamic Manipulation of a Virtual Object in Marker-less AR system Based on Both Human Hands”, Transactions on Internet and Information Systems, vol. 4, no.4, **(2010)**, pp. 618-632.

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

[13] 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.

Authors

Junchul Chun

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 CCSRC(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.

Seonho Lee

Seonho Lee is currently a graduate students 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 fields are augmented reality and haptic interaction in AR.