

Depth Map based Real Time 3D Virtual Image Composition

Hye-Mi Lee*, Nam-Hoon Ryu*, Eung-Kon Kim*

* Department of Computer Engineering, Sunchon National University, Korea

lhrooh@sunchon.ac.kr, nhryu@sunchon.ac.kr, kek@sunchon.ac.kr

Abstract— To complete an image, it is needed to go through the process to capture the actual actor's motion and compose it with virtual environment. Due to the excessive cost for production or lack of post-processing technology, however, it is mostly conducted by manual labor. The actor plays his role depending on his own imagination at the virtual chromakey studio, and at that time, he has to move considering the possible collision with or reaction to an object that does not exist. And in the process of composition applying CG, when the actor's motion does not go with the virtual environment, the original image may have to be discarded and it is necessary to remake the film. The current study suggested and realized depth-based real-time 3D virtual image composition system to reduce the ratio of remaking the film, shorten the production time, and lower the production cost. As it is possible to figure out the mutual collision or reaction by composing the virtual background, 3D model, and the actual actor in real time at the site of filming, the actor's wrong position or acting can be corrected right there instantly.

Keywords—Image Composition, Chromakey, Kinect, Depthmap

I. INTRODUCTION

In the broadcasting media including movie, using the digital technology to produce the video is a general phenomenon. The computer graphic technology, which can realize the creative idea, has remove the limitation of the production space and allowed the production of diverse contents by composing the actual image and virtual environment. In case of video hard for live action, the background image is composed with the actual actor using the computer graphic technology and it must be composed without join and the sense of difference in the images must be minimized. The realistic and high quality video invested with hugh capital raised the demand level of the audience and to overcome that, the development of low cost and high efficient video production technology is desperately needed.

In this article, the system to monitor the expected result screen simultaneously with the filming by composing the integrated virtual environment applied with 2D background and 3D virtual model with the actor was intended to be designed and realized. Since the conflict between the actor and 3D virtual objects and the reaction can be judged during the filming, the filming can be completed correcting the wrong action and the location of the actor. In Chapter II, all the technologies required to composite images are explained, in Chapter III, depth information-based real-time 3D virtual

image composite system is explained, in Chapter IV, the realization results are examined and in Chapter V, the conclusion and the expected effects are presented.

II. RELATED WORKS

A. Virtual Studio Composite Technique

To complete the video, the action of actual actor is captured and passed through the process of compositing with virtual environment and at this moment, it is general that the compositing process is made manually due to the cost problem and the lack of post-processing technology. Although the prior preparation is made in order to match the actual actor and the virtual object through the preproduction process, the director cannot verify the result image during the actual filming. For that, the case that the movement of the acting object extracted in the post production process and the 3D virtual environment are not matched correctly occurs frequently.

In the post production process, the object of actual actor extracted by chromakey or Rotoscoping and 3D virtual environment are passed through the task composing them in 1 frame unit. The action taken by the actor in the chromakey environment is determined depending on pure imagination and moved considering collision and the interaction with the not existing objects. However, in case that the location, eye line, feeling, etc. of the actor is mismatched delicately, the task, which the filming must be made again in the chroma studio [1] and refilmed video is composed by frame must be repeated. To film again, the problems occur such as the environment should be same as previous filming environment (distance between cameras, angle, lighting environment, structures, etc), and all the manpower such as actor, extra, staffs, etc. should be mobilized again. In addition, since while the task was made manually, the time is passed and it is hard to bring back the previous emotions, the probability that the completeness of action would be dropped is increased, too.

B. Foreground Object Separation Technique

For the visual effect compositing method, which extracts the certain area from the original video, there are the chromakey technique and the rotoscoping technique. Chromakey is the technology of overlaying the new background plate and the interested area after filming the subject in the studio built with single color so that the computer can separate the interested area from the background

easily. Since the blue is generally used for single color, it is called as blue screen compositing technique but since specific color can be assigned in the video entered through the camera at one's discretion, there is no limitation for color. Figure 1 shows the picture applied with virtual environment by the chromakey.



Figure 1. Virtual Image Synthesis by Chromakey

Rotoscoping technique refers to separate the required interested area from the video manually when the subject should be isolated from the background with no single color. Since the separation and compositing is made manually by each frame, it is very slow and labor-intensive but depending on the situation, it becomes unique solution [2].

These foreground object separation techniques separate the video with 2-layer only; foreground and background. Since the chromakey background has the structure that the entire video is replaced with the virtual video, the partial replacement is impossible. 3-layer double chromakey can composite the 3D model additionally with the subject, which is the foreground and the virtual environment, which is the background, but since the real-time z-buffering is not applied, the front forward and backward arrangement order of background-subject-3D model is not reversed. The multiplane compositing can connect 2D- and 3D- works but 2D images are arranged along the z-axis in the 3D space, the scene is shown through the movement of virtual camera [2]. Since such technology is the process made manually in the post production later, it is not suitable to visualize the compositing results in real-time when the video, which the action is made together with 3D object.

III. REAL-TIME VIRTUAL VIDEO COMPOSITION SYSTEM

In this chapter, the video compositing system which replaces the 2D background by converging the depth information and the real object and the 3D object can be responded mutually in real-time is designed excluding the fragmentary compositing of chromakey. Since the purpose of this system is to produce the video according to the story or scenario, all the works like the virtual environment and the animation of 3D object should be made in advance. This system is divided into 2-layer composit module, 3D virtual space generation module and 2D-3D virtual video compositing engine.

A. Text Font of Entire Document

In case of the foreground and background separation and the background superimposing by the existing chromakeying, the depth buffer is not generated. Since to link with the 3D model, 3D space data is generate, the foreground and background are separated through the depth-keying method.

As the hardware to obtain precise depth data is expensive equipment, to obtain the depth image and color image with low cost at the same time, obtain RGB image and the depth stream using Kinect. Kinect is the device that generates the 3D space data, and provides diverse user experiences by measuring the depth of the video entered through the infrared pattern recognition [3], [4]. Figure 2 shows the 2-layer compositing module pipeline.

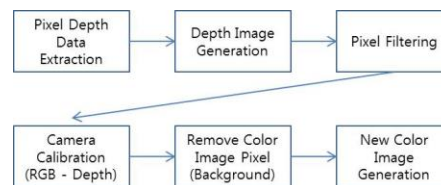


Figure 2. Pipeline of the 2-Layer Composition Module

First, to extract the foreground to separate object, each pixel depth value should be obtained from the raw depth video entered. The depth value of each pixel is mapped responded to the physical distance of the video's field of vision. the bit field stored the depth information of each pixel is total 16 bits, among which the depth bits use 13 bits and the player index uses 3 bits. Out of 13 bits of depth bits, 12 bits are used for the depth information of each pixel and 1 bit is used to express if the depth measurement is possible or not. Player Index is the information to detect the figure of human and is corresponded to the value to extract the foreground. Figure 3 shows the layout of the depth bit [5].

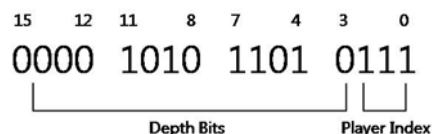


Figure 3. Layout of Depth Bits

To separate the foreground from the depth image generated, the filtering should be performed. Kinect returns the depth value from 0 to 4095 and each value responded to length unit of mm. In 2-layer composite module, the threshold value for the foreground position is set to 800 ~ 4000, and the pixel, of which depth threshold is exceeded 4000, is processed to background through filtering. Figure 4(a) shows the color video entered through RGB camera, and (b) shows the depth image visualizing the depth of space.



Figure 4. The process of Separating the Foreground from the Depth Image

There is a problem that the pixel of depth image entered through the depth camera is not converted in 1 vs 1 with the color stream entered through the RGB camera. The reason is because the positions of two cameras are different and convert the position of the depth image pixel position to the pixel position of color image through the camera calibration [6], [7]. And then remove the pixel, which is not the foreground and generate the new color image and draw it to the area corresponded to background. At this moment, since the noise is made around the boundary of the foreground, eliminate the noise by compensating the edge. Figure 5(a) shows the RGB image of the foreground after filtering the background and Figure 5(b) shows the image that the background of actual filming image entered through the device was converted and composed with the foreground.

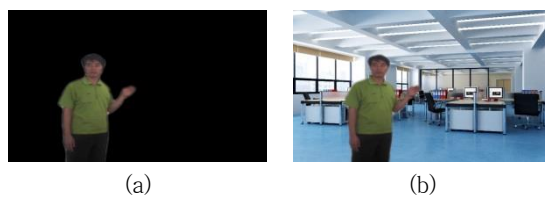


Figure 5. Synthesis of the Foreground and Background Image based on the Depth Information

B. 3D Virtual Space Generation Module

This is the module to build the 3D virtual environment needed in order for the actor to act according to the scenario. 3D object is loaded to agent tool in order for the user to control such as zoom-in/zoom-out, rotate, move, etc. at his/her discretion before being composed with the final outcome. In the agent tool, the user input data for the object is stored and sent to virtual image composite engine. Figure 6 shows the picture that the coordinates is being adjusted at discretion in order to select the composite location after loading the virtual model of Bumblebee of Transformer to the agent tool.

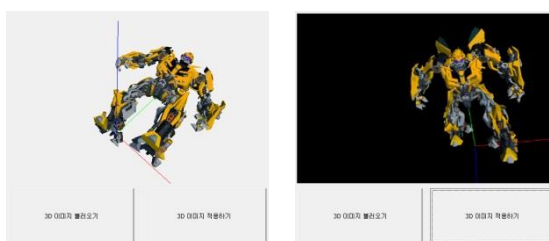


Figure 6. Agent tool for 3D object

C. 2D-3D Virtual Video Composition Engine

In the engine needed to verify the final video outcome, 2D background and 3D virtual environment are composed. The background completed the processing in the compositing module is replaced and composed with the space for 3D virtual model rendering. Perspective is given by applying the perspective projection and the forward and backward arrangement sequence can be verified visually by processing the area, which the foreground and 3D virtual model are

overlapped, with hidden surface. 2D background assigned as background are drawn to entire screen at the last from the Z-axis. The depth test for the extract image of the foreground actor and the 3D virtual model is calculated in real-time and the Z-buffering is determined according to the movement of two objects.

IV. SYSTEM REALIZATION RESULTS

Since this system is the technology to reduce refilming rate when filming virtual images, the purpose of this system is that the position and eye line of the actual actor is not mismatched according to the virtual objects appeared in video. The actor acts in the virtual studio where the 3D objects to be arranged in final outcome are not existed but in the video, the virtual environment and 3D objects appear together. Therefore, the movement and the flow of eye line of the actual actor matched with the virtual environment can be observed.

Figure 7 shows the situation that virtual background, actual actor, 3D virtual model are continuously moving. (a) is the results realized as if the 3D virtual model is closer to the camera than the actor and (b) is the results realized as if the actor is far away from the camera than 3D virtual model. Each object is moving continuously and are rendering in the screen in real-time.

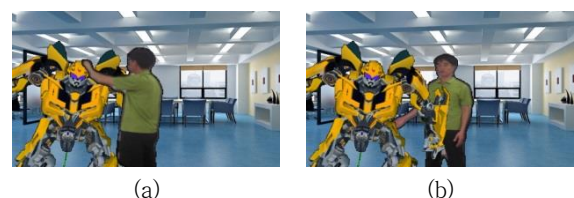


Figure 7. Screen of the Final Implementation Results

V. CONCLUSIONS

In the dangerous scene or SF movie that the live action is hard, the video production using the computer graphics technology accompanies many problems such as technical problem, working time, cost, etc. To complete the high quality video, the work should be done manually but if the action of the actor is not matched with the 3D virtual environment delicately, the problem occurs that the live action should be made from the beginning discarding the original video. The proportion that the refilming should be made for that reason is reached to approximately 10% out of filmed video and lots of cost and time are wasted.

Since the depth-based real-time 3D virtual image compositing system can verify the position to be inserted the virtual environment and 3D model in the video in advance and monitor the outcome composed with the action of actor and virtual environment through the screen in the filming site in real-time, it can reduce the error to be occurred in the post production process. If the refilming proportion is reduced through this system, since the production period for high quality video can be reduced and significant amount of production cost can be saved, the vitalization of diverse and experimental video production can be expected.

ACKNOWLEDGMENT

This research was financially supported by the Ministry of Education (MOE) and National Research Foundation of Korea(NRF) through the Human Resource Training Project for Regional Innovation (No. 2013H1B8A2032217).

REFERENCES

- [1] C.-K. Lee, and G. Park, "A Study on Comparison of background chroma studio for Virtual Studio," J. KOSST, vol. 7, no. 2, pp. 36~41, 2012.
- [2] S. Wright, *Compositing Visual Effects*, Focal Press, 2011.
- [3] H.-M. Lee, N.-H. Ryu, and E.-K. Kim, "Expression Technique of the Plurality IR LED Marker using OOK Method for Augmented Reality," J. of The Korean Institute of Electronic Communication Sciences, vol. 7, no. 2, pp. 433-438, 2012.
- [4] J. Park, and J. Park, "Upper Body Exercise Game using a Depth Camera," J. of Korean Society For Computer Game, vol. 25, no. 1, pp. 61-66, 2012.
- [5] J. Webb, and J. Ashley, *Beginning Kinect Programing with the Microsoft Kinect SDK*, O'Reilly, 2012.
- [6] K.-I. Kim, S.-H. Han, and J.-S. Park, "A Distortion Correction Method of Wide-Angle Camera Images through the Estimation and Validation of a Camera Model," vol. 8, no. 12, pp. 1923-1932, 2013.
- [7] S.-C. Kwon, W.-Y. Kang, and Y.-H. Jeong, "Stereoscopic Video Composition with a DSLR and Depth Information by Kinect," J. Korea Inst. Commun. Inform. Sci. (KICS), vol. 38C, no. 10, pp. 920-927, 2013.



Hye-Mi Lee received the B.E., and M.S. degrees in computer engineering from Sunchon National University, Korea in 2004 and 2010, respectively. Since 2010, she has been a Principal Researcher in Withusbiz Co., Ltd., Korea. Her research interests are image processing, computer graphics, and HCI.



Nam-Hoon Ryu received the M.S., and Ph.D. degrees in computer science from Sunchon National University, Korea in 2009 and 2012, respectively. Since 2012, he has been a CEO in Withusbiz Co., Ltd., Korea. His research interests are image processing, computer graphics, and algorithm.



Eung-Kon Kim received B.E. degrees in electronic engineering from Chosun University, Korea in 1980. He received the M.E. degrees in computer engineering from Hanyang University, Korea in 1986. He received the Ph.D. degrees in computer engineering from Chosun University, Korea in 1992. Since 1993, he has been a professor of computer engineering in Sunchon National University, Korea. His research interests are image processing, computer graphics, multimedia, and HCI.