2.5D Video Avatar for Networked VRPhoto System*

Youngjung Suh, Dongpyo Hong and Woontack Woo

KJIST U-VR Lab.
Gwangju 500-712, S. Korea
+82-62-970-3157
{ysuh, dhong, wwoo}@kjist.ac.kr

Abstract

In this paper, we propose a novel yet simple way to generate a photo-realistic 2.5D video avatar on the fly for a networked VRPhoto system, which allows users at a distance taking a photo/video int eractively in a shared 3D virtual environment through networks. The avatar is an important medium for users at a distance to feel they seem to be in a same space. The proposed algorithm consis ts of three key steps: (1) 2.5D video avatar generation from a natural background, (2) mesh simplification for efficient transmission of the generated 2.5D video avatar through a limited network bandwidth, (3) real-time augmentation of video avatar into 3D virtual space over network. The pro posed algorithm can be applied to various types of VR applications that require real-time interactions between users at a distance.

1 Introduction

Over last few years, various research activities on avatar generation have been reported. At first, a CG-based avatar has been proposed but it is used in limited applications due to its weakness, i.e., the lack of reality. To overcome the weakness, an image (video)-based avatar has been developed, where the texture of the avatar is segmented from background image and then augmented into virtual world. However, augmenting 2D video avatars onto virtual world is unnatural, since 2D video avatars do not have 3D position information, in general [1]. As a result, it is difficult to allow 2D video avatars to interact with real or virtual objects. To relieve these problems, a 3D video avatar has been developed. However, it usually takes time in generating 3D video avatar and requires extensive computational power in modelling and rendering the avatars [2]. Though 2.5D video avatar has been proposed as a compromise between 2D and 3D video avatars, it also has various limitations since the avatar is generated by exploiting chroma-keying in segmenting a user from the captured image [3].

In this paper, we propose a novel yet simple way to generate a photo-realistic 2.5D video avatar from the natural scene on the fly for its efficient delivery through a network and augmentation into a virtual environment [4]. Figure 1 shows the proposed system block diagram. The proposed algorithm, at first, generates 2.5D video avatar from a natural scene. Then, it performs mesh simplification, which is to represent the mesh model of 2.5D video avatar efficiently for network delivery. At last, the simplified mesh model of 2.5D video avatar is naturally augmented into 3D virtual environment over a network.

* This work is supported by KIST and ICU DML.

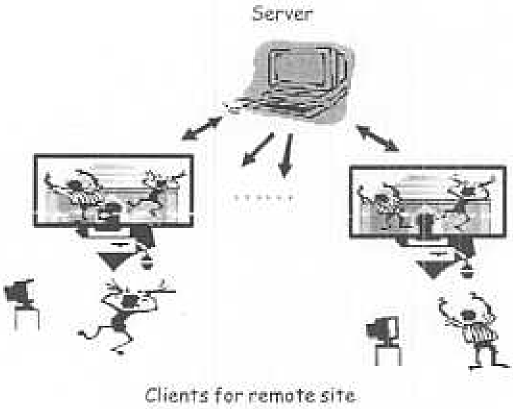


Figure 1: Conceptual diagram of Networked VRPhoto System

The resulting 2.5D video avatar can be a medium for users from different places to provide immersive feeling that they seem to be in a same space. Thus, the avatars can observe existence of the others as well as themselves in a same virtual space, navigating VR space by exploiting their 3D depth information. The proposed algorithm can be widely exploited for VR applications requiring the real-time interactions as well as various collaborations between users at a distance such as teleconference, tele-education, game, and broadcasting.

This paper is organized as follows: in section 2, user segmentation from natural background and real-time mesh model generation are described. In section 3, we explain network delivery and augmentation of the generated 2.5D video avatar. Brief description on the Networked VRPhoto System exploiting the proposed 2.5D video avatar and conclusion are in Section 4 and 5, respectively.

2 2.5D Video Avatar generation

In the proposed algorithm, we adopt a statistical algorithm for detecting moving objects from a natural background scene [5][6][7]. First, we calibrate distorted color values due to the physical limitations of digital cameras. Then we model static background of the compensated image sequences over time. The segmentation algorithm exploits the differences between a trained background image and a current image to segment moving objects from a natural scene. We, first, segment objects in RGB color space by comparing a current image and the modelled reference image. However, RGB color model has weakness in segmenting object without shadow. To segment shadow from the segmented moving objects, we introduce a normalized RGB color model, rgb, with pixelwise dynamic threshold. The rgb color model can detect shadow from the segmented object due to its characteristics. The proposed algorithm overcomes the weakness of the previous background segmentation algorithms by introducing pixel-wised dynamic threshold values considering the characteristics of color channels [8].

For mesh model generation, we first find correspondence between pairs of stereo sequences to estimate 3D image. Then, as shown in Figure 2, we assume that 3D point of each pixel with disparity value might be adjacent to one another also in 3D OpenGL space. Under this assumption, mesh model is generated by triangulation algorithm that connects points according to a specific rule. Also, each vertex that makes up triangular mesh has the texture coordinate with color information. By setting the material property of polygon in model as RGBA array, texture of color image is mapped onto triangular mesh, resulting 2.5D video avatar.

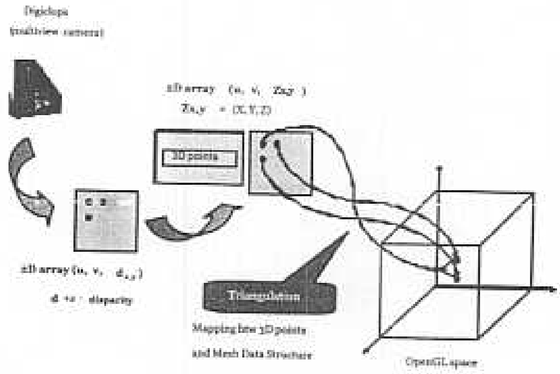


Figure 2: Mesh model generation

3 Augmentation of 2.5D video avatar over networks

Simplifying the generated mesh model of 2.5D video avatar is necessary for efficient transmission and augmentation of 2.5D video avatar through the limited network bandwidth. We can simplify the three components, or vertex, edge, and face of the mesh model. In the proposed algorithm, we simplify the vertices of mesh model using a regular hexahedron. As shown in Figure 3, we form the bounding box which encloses the entire model by exploiting the smallest and largest value of the coordinates of mesh model. After that, the generated bounding box is split into several small cells. Then, if a point of mesh model is included within a specific cell, it is mapped onto the representative point of the cell. Consequently, the number of vertices of mesh model is reduced.

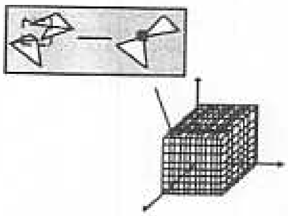


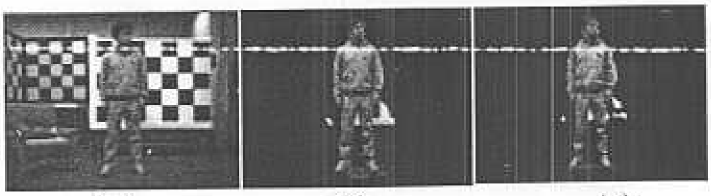
Figure 3: Mesh simplification

The simplified 2.5D video avatar is augmented with virtual cultural heritage considering the camera parameters of real and virtual cameras. The process of the avatar's augmentation over a network is as follows: the environment in which Quanta library is available is set to the clients and a server. A server plays a role in sending data from one client to the others (more than one) simultaneously. Each client sends 3D points and corresponding color values of the avatars, and receives same types of data for one frame from the others to render those.

4 Implementation of VRPhoto System

To prove the usefulness of the proposed algorithm, we implemented "Networked VRPhoto System." In the implemented VRPhoto system, the users from different places can take a photo/video interactively in a shared 3D virtual environment through a network. The system is implemented at Pentium III Dual Xeon 1.0GHz CPU. To obtain images, Digiclops of IEEE 1394 camera is used. Sony UP-DP10 is used as a digital photo printer.

Figure 4 shows the result of the proposed background subtraction. As shown in Figure 4(b), user and its shadows are subtracted from background in RGB color space. However, Figure 4(c) shows only user without shadows, where it is subtracted in the normalized RGB color space. The proposed algorithm can be applied to estimate the light source for VR augmentation applications such as VR studio.



(a) (b) (c)

Figure 4: Background subtraction result. (a) input image, (b) subtraction in RGB color space, (c) subtraction in the normalized RGB color space.

Figure 5 shows the process of constructing the mesh model of 2.5D video avatar exploiting 3D depth information from a 3D camera.

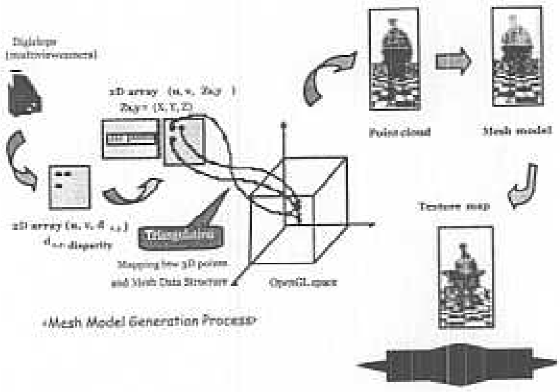
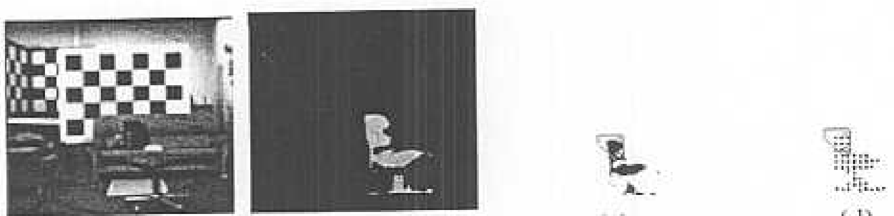


Figure 5: Mesh model generation of 2.5D video avatar

Then, we simplify the polygonal mesh model of generated 2.5D video avatar. Since the delivering cost, as well as rendering cost, of using a mesh model is directly related to the number of polygons, it is useful to efficiently represent the mesh models. Figure 6 shows the results of mesh simplification.



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

Figure 6: Mesh simplification. (a) Input image, (b) depth of segmented moving object, (c) original model, (d) simplified model

After all, the simplified mesh models with textures are augmented into a shared virtual space in a server. As shown in Figure 7, the simplified 2.5D video avatar is augmented into 3D virtual environments considering the camera parameters of real and virtual cameras over a network.

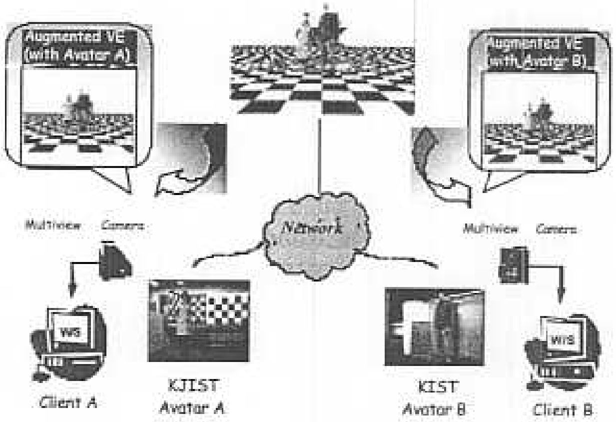


Figure 7: Networked VRPhoto system

5 Conclusion

In this paper, we introduced a novel approach to segment moving objects from natural scene to generate 2.5D video avatar and to augment the avatar naturally into 3D virtual environment through a network using the calibrated camera parameters and depth information of the 2.5D avatar. The proposed algorithm can be applied for VR applications requiring the real-time interactions for various collaborations between users at a distance such as teleconference, tele-education, game, and broadcasting. In addition, without loss of generality the video avatar can be transmitted over the Network by efficiently representing the mesh model.

6 References

[1] "The ALIVE System: Wireless, Full-body Interaction with Autonomous Agents", M.I.T. Media Laboratory Perceptual Computing Technical Report No. 257 http://alive.www.media.mit.edu/projects/alive/

[2] "Virtualized Reality: Constructing Virtual Worlds from Real Scenes", IEEE MultiMedia, vol.4, no.1, Jan.-Mar. (1997),pp.34-47. http://www.cs.cmu.edu/afs/cs/project/VirtualizedR/www/VirtualizedR.html

[3] Hirose, M., Ogi, T., Yamada, T., Tanaka, K., Kuzuoka H., "Communication in Networked Immersive Virtual Environments", 2nd International Immersive Projection Technology Workshop (1998)

[4] Y.Suh, D.Hong, and W.Woo, "2.5D Video Avatar Augmentation for VRPhoto," ICAT'02, pp. 182-183, Dec. 3-6, (2002).

[5] T. Horprasert, D. Harwood, and L.S. Davis, "A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection," Proc. IEEE ICCV'99 FRAME-RATE Workshop, Kerkyra, Greece, September (1999)

[6] W. Woo, N. Kim, K. Wong and M. Tadenuma, " Sketch on Dynamic Gesture Tracking and Analysis Exploiting Vision-based 3D Interface," in Proc. SPIE PW-EI-VCIP'01, vol. 4310, pp. 656-666, Jan.(2001).

[7] N. Kim, W. Woo and M. Tadenuma, " Photo-realistic Interactive 3D Virtual Environment Generation Using Multiview Video," in Proc. SPIE PW-EI-VCIP'01, , vol. 4310, pp. 245-254, Jan. (2001).

[8] D.Hong, W.Woo, "Background subtraction for a vision-based interface", ICIP'03 (under review)