ICAT 2002
December 4-6, Tokyo, JAPAN

2.5D Video Avatar Augmentation for VRPhoto

Youngjung Suh, Dongpyo Hong, Woontack Woo

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

Abstract

In this paper, we introduce a novel and natural way to augment photo-realistic 2.5D video avatars on the fly for the VRPhoto system, which allows users taking a photo/video interactively in a shared 3D virtual environment. The proposed algorithm generates polygonal mesh models of 2.5D video avatars from natural scenes in real-time and augments them into 3D virtual environment. Due to its simplicity and robustness of the proposed algorithm, it can be applied to various applications exploiting real-time interaction such as media art, art education, entertainment, broadcasting etc.

Key words: video avatar, augmentation, background segmentation

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 (or 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, 3D video avatars have been developed. However, it usually takes time in generating 3D video avatars and requires extensive computational power in modeling and rendering the avatars [2]. Though 2.5D video avatars have been proposed as a compromise between 2D and 3D video avatars, it also has various limitations since the avatars are generated by exploiting chroma-keying in segmenting users [3].

Meanwhile, the proposed algorithm introduces a novel approach to segment moving objects from natural scenes to generate 2.5D video avatars and to augment the avatars naturally into 3D virtual environment using the calibrated camera parameters and depth information of the 2.5D avatars. As a result, the proposed algorithm generates photo-realistic video avatars from natural scenes on the fly. In addition, the resulting 2.5D video avatars can navigate VR space by exploiting the 3D depth information. The proposed algorithm consists of three key steps: (1) user segmentation from natural background, (2) real-time 3D depth map acquisition and mesh model generation, (3) augmentation of video avatar into 3D virtual space.

2. Background subtraction and real-time 2.5D mesh model generation

In the proposed algorithm, we adopt a statistical algorithm for detecting moving objects from a natural background scene [4][5]. The segmentation algorithm exploits the differences between a trained background image and a current image to segment moving objects from a natural scene. Since 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 can be applied to estimate the light source for VR augmentation applications such as VR studio. Figure 1 shows the results of segmented moving object which satisfies threshold values considering channel characteristics of each color.

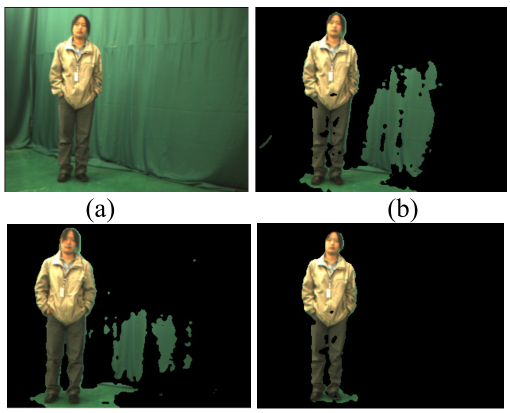

(a) (b) (c) (d)

Fig 1. Moving object segmentation
(a) current image (b) RGB-based segmentation (c) rgb-based segmentation (d) proposed rgb-based segmentation

As shown in Figure 1 (c) and (d), it is more robust to segment moving object using pixelwise dynamic threshold values considering each color channel than conventional approaches using threshold without considering the characteristics of each color channel.

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 neighboring disparities might be adjacent to each other in 3D OpenGL space. Given the assumption, mesh model is generated by triangulation algorithm. 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 in 2.5D video avatars.

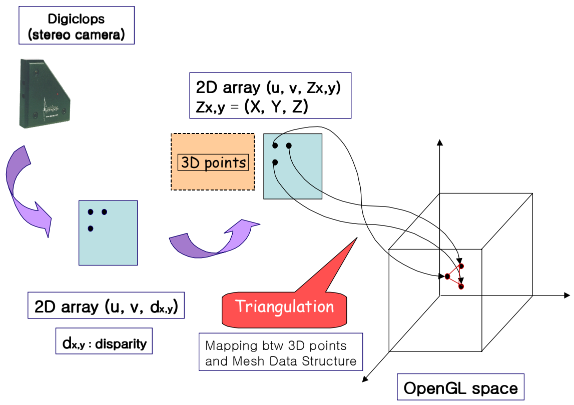
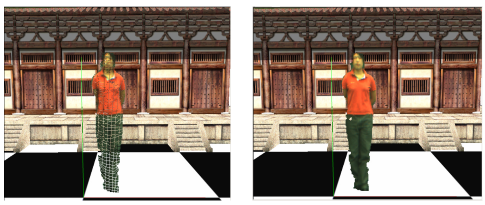


Fig 2. mesh model generation process

3. Video avatar augmentation

As shown in Figure 3, the 2.5D video avatar is augmented with virtual cultural heritage considering the camera parameters which are obtained through the camera calibration [6].

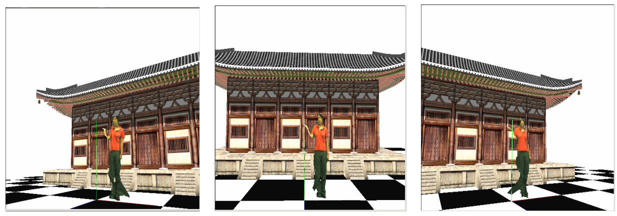


(a) (b)

(b)

Fig 3. Enlarged 2.5D Video Avatar
(a) point cloud (b) mesh model

According to experimental results, we observed that the remarkable difference between point cloud-based and mesh-based 2.5D video avatars when it comes to navigation. The 2.5D video avatar with mesh model does not have the empty space among points when the scene is zoomed in, which is a main weakness of the point cloud-based avatar. In addition, the mesh-based avatar can navigate 3D virtual space and interact with virtual object in the shared space effectively. Even though the 2.5D video avatar does not have a complete 3-D shape, the 3D appearance can be maintained for a limited area. In general, if the optical center of the virtual camera moves within 10 degrees along the video avatar, then the avatar may be perceived as having a 3D model, as shown in Figure 4.



(a) (b) (c)

Fig 4. Rotated 2.5D Video Avatar (a) Rotated (-10˚) (b) Original Scene (0˚) (c) Rotated (+10˚)

4. VRPhoto System

Current PC-based VRPhoto system can process 8~9 frame per second which includes disparity estimation, mesh model generation, texture mapping, augmentation, etc. By optimizing each component of the system and distributing overload into several high-performance computers, the processing speed of the VRPhoto can be improved to be used in real-time applications. Due to its simplicity and robustness of the proposed algorithm, it can be applied to various types of real-time interactive systems such as media art, art education, entertainment, broadcasting etc. In addition, without loss of generality the video avatar can be transmitted over the Network by efficiently representing the mesh model.

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. 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

5. 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.

6. 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.