Poster: A Virtual Body for Augmented Virtuality by Chroma-Keying of Egocentric Videos

Frank Steinicke* Gerd Bruder† Kai Rothaus‡ Klaus Hinrichs§

Department of Computer Science
University of Münster
Einsteinstr. 62, 48149 Münster, Germany

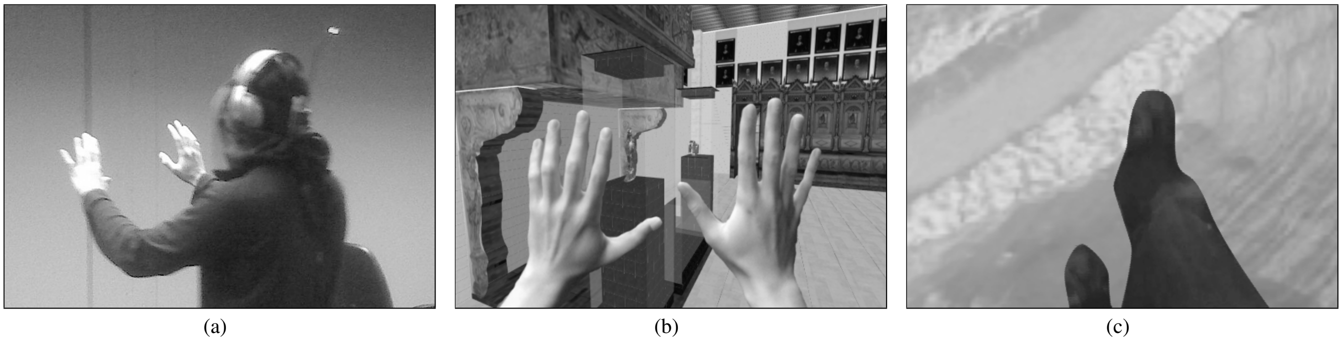

Figure 1: Virtual body in an augmented virtuality scenario: (a) A user in an immersive virtual environment with a video-see-through HMD mockup. (b) The user's view in the virtual world with visualization of his virtual hands in an indoor museum environment and (c) with virtual lower part of the body on the glass bridge of a virtual model of the Grand Canyon Skywalk.

Abstract

A fully-articulated visual representation of oneself in an immersive virtual environment has considerable impact on the subjective sense of presence in the virtual world. Therefore, many approaches address this challenge and incorporate a virtual model of the user's body in the VE. Such a "virtual body" (VB) is manipulated according to user motions which are defined by feature points detected by a tracking system. The required tracking devices are unsuitable in scenarios which involve multiple persons simultaneously or in which participants frequently change. Furthermore, individual characteristics such as skin pigmentation, hairiness or clothes are not considered by this procedure.

In this paper we present a software-based approach that allows to incorporate a realistic visual representation of oneself in the VE. The idea is to make use of images captured by cameras that are attached to video-see-through head-mounted displays. These egocentric frames can be segmented into foreground showing parts of the human body and background. Then the extremities can be overlayed with the user's current view of the virtual world, and thus a high-fidelity virtual body can be visualized.

Keywords: augmented virtuality, virtual body

1 Introduction & Motivation

Digital representations of the user, so-called avatars, are in common use in video and multi-player on-line games, but are usually controlled only by keyboard or mouse. Only few current-state VR setups incorporate fully-articulated virtual bodies. This lack of digital body representations in VEs may be due to the fact that today's tracking systems require a considerable instrumentation of the user in order to provide a fully articulated VB. It is nevertheless highly recommended to provide a realistic and naturally articulated virtual body in an HMD environment that can be controlled in real-time by the viewer's own movements and viewed from a first-person perspective. Virtual human models for VR applications have been presented and analyzed for their impact on social interaction [6]. Furthermore, as mentioned above, the existence of a VB has been shown to increase a participant's sense of presence measurably. The reasoning is as follows: if a body is in a certain location and if a person has a certain association with that body, it is likely that this person will believe that she is in that location [4].

Camera images from real users in order to incorporate avatars in VR environments has been used in video conferencing systems and 3D model reconstruction. In conferencing systems videos of real users are added to virtual surroundings and allow users to interact face-to-face [2]. The blue-c system [1] uses visual hull based approaches to reconstruct 3D models from video streams. Steinicke et al. have presented the concept of virtual reflection where users were able to see their own reflection captured by an external web camera in a semi-immersive environment [5]. However, for most of these approaches several cameras are required in order to diminish reconstruction errors. Furthermore, these approaches are focussed on 3D reconstruction by means of several static perspectives rather than using a dynamic egocentric camera perspectives to present a virtual body.

In contrast to augmented reality, our *augmented virtuality* environment refers to predominantly virtual spaces, where physical elements are dynamically integrated to support interaction with the virtual world in real-time [3].

*e-mail: fsteini@math.uni-muenster.de
†e-mail:g_brud01@math.uni-muenster.de
‡e-mail:rothaus@math.uni-muenster.de
§e-mail:khh@math.uni-muenster.de

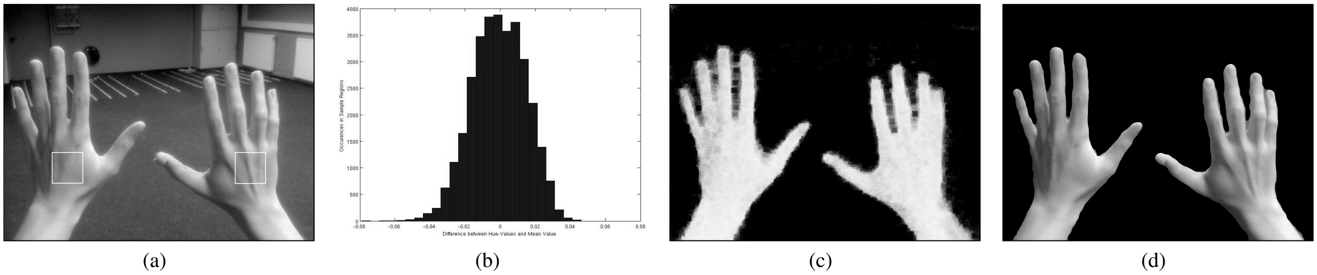

Figure 2: Process to obtain a virtual body: (a) back of the right and left hand covered by a white, square training regions, (b) centralized empirical observation set, (c) region confidence map representing plausibility of pixels to be part of the skin, and (d) segmented skin pixels.

2 Obtaining a Virtual Body

2.1 Video-see-through Augmented Virtuality

We use a customized video-see-through HMD version based on a 3DVisor Z800 (800x600@60 Hz, 40° diagonal FoV) for the visual presentation, and attached a camera setup consisting of one USB camera with a resolution of 640×480 pixels and update rate of 30 frames per second (see Figure 1(a)). On top of the HMD an infrared LED is fixed. We track the position of this LED within the room with an active optical tracking system (Precise Position Tracking of World Viz), which provides sub-millimeter precision and sub-centimeter accuracy. The update rate is 60 Hz providing real-time positional data of the active markers. For three degrees of freedom (DoF) orientation tracking we use an InertiaCube 2 (InterSense) with an update rate of 180 Hz. The InertiaCube is also fixed on top of the HMD.

2.2 Classification and Segmentation

As mentioned above, this camera shows the real world from the position and orientation of the user in the VR laboratory space, while head movements are used to render the VE according to tracked motions. In the following we describe how to realize an efficient algorithm for skin detection. HSV color-space is best suitable for skin detection, we transform the captured images into hue H, saturation S and intensity value V. As usual in supervised pattern recognition, the task is divided in two phases, a training phase and a classification task. To account for different skin colors, the first phase of our approach is to train the skin classifier. The user is asked to move the hands, so that the backs of the right and left hand cover the white, square training regions (see Figure 2 (a)). The hue values of the pixels within these regions are taken as observation set to compute the mean skin color μ and the standard deviation σ. Figure 2(b) shows the centralized empirical observation set. The classification phase is realized in three main steps. First the hue values are computed and centralized by $H' = H - \mu$. A slight smoothing using a Gaussian filter kernel completes the preprocessing step. Secondly, for each pixel we estimate a value which represents the plausibility of the pixel to be part of a skin region (see confidence map in Figure 2(c)). The plausibility value $P(p)$ of a pixel p depends on the centralized hue value $H'(p)$ and the hue contrast $C(P)$, which is the difference of the minimal and maximal H' in a 21×21 neighborhood of p:

$$P(p) = \left(1 - 2\left|H'(p)\right|\right) \cdot (1 - C(p))^2$$

Hence, a white-colored pixel corresponds to a high plausibility that the pixel is part of the skin, a black-colored pixel corresponds to a low plausibility. The third step is the segmentation of the skin pixel. Therefore, we first smooth the plausibility map P. All pixels with a plausibility value higher than $1 - 25\sigma$ are taken as skin pixel candidates. The resulting binary image contains some holes in skin regions and some wrong classified skin regions in the background. Both failures are reduced by using an extension of the median filter technique, i. e., by counting the number of skin pixel candidates in a 7×7 neighborhood of p. A pixel is redefined as skin pixel candidate if and only if there are at least 13 skin pixel candidates in its neighborhood. We repeat this procedure by increasing the lower bound to 17, 21, and 25.

2.3 Chroma-Keying

We define the background in the foreground image to be a particular color, which ideally does not appear in the displayed VE. During the composition step these background pixels from the foreground image are neglected, and only those pixels with a different color, i. e., regions showing the virtual body, replace the corresponding pixels from the image showing the virtual world. This procedure is implemented in a fragment shader and requires only one comparison, and therefore it can be realized in real-time.

3 Conclusion

We have shown how the human's extremities can be segmented from these egocentric videos, and we have presented how we merge such foreground images with the user's current view of the virtual environment.

We plan to extend our approach by a stereo-based setup in order to derive a three-dimensional representation of the virtual hand. This can be used for global effects, but also interaction, for instance grabbing can be supported with such information.

References

[1] M. Gross, S. Würmlin, M. Naef, E. Lamboray, C. Spagno, A. Kunz, E. Koller-Meier, T. Svoboda, L. V. Gool, S. Lang, K. Strehlke, A. V. Moere, and O. Staadt. Blue-c: A spatially immersive display and 3d video portal for telepresence. In 819-827, editor, *ACM Transactions on Graphics (TOG)*, volume 22, 2003.

[2] H. Maeda, T. Tanikawa, J. Yamashita, K. Hirota, and M. Hirose. Real world video avatar: Transmission and presentation of human figure. In *Proceedings of the IEEE Virtual Reality 2004*, page 237, Washington, DC, USA, 2004. IEEE Computer Society.

[3] P. Milgram and F. Kishino. A Taxonomy of Mixed Reality Visual Displays. In *IEICE Transactions on Information and Systems, Special issue on Networked Reality*, 1994.

[4] M. Slater, A. Steed, J. McCarthy, and F. Maringelli. The influence of body movement on subjective presence in virtual environments. *Human Factors*, 40(3):469–477, 1998.

[5] F. Steinicke, K. Hinrichs, and T. Ropinski. Virtual Reflections and Virtual Shadows in Mixed Reality Environments. In *10th International Conference on Human-Computer Interaction (INTERACT2005)*, pages 1018–1021, 2005.

[6] N. M. Thalmann and D. Thalmann. The artificial life of synthetic actors. *Transactions of the Institute of Electronics, Information and Communication Engineers D-II*, J76D-II(8):1506–14, 1993. MIRALab, Geneva Univ., Switzerland.