Medicine Meets Virtual Reality 19
J.D. Westwood et al. (Eds.)
IOS Press, 2012

doi:10.3233/978-1-61499-022-2-273

Design and Evaluation of a Medical Teamwork Training Simulator using Consumer-Level Equipment

Stefan MARKS [a,1], John A. WINDSOR [b,2] and Burkhard WÜNSCHE [a,3]
[a] *Department of Computer Science, Faculty of Science*
[b] *Department of Surgery, Faculty of Medicine and Health Sciences, The University of Auckland, New Zealand*

Abstract. Virtual environments (VE) are increasingly used for teamwork training purposes, e.g., for medical teams. One shortcoming is lack of support for nonverbal communication channels, essential for teamwork. We address this issue by using an inexpensive webcam to track the user's head and using that data for controlling avatar head movement, thereby conveying head gestures and adding a nonverbal communication channel. In addition, navigation and orientation within the virtual environment is simplified. We present the design and evaluation of a simulation framework based on a game engine and consumer-level hardware and the results of two user studies showing, among other results, an improvement in the usability of the VE and in the perceived quality of realism and communication within the VE by using head tracking avatar and view control.

Keywords. medicine, training, teamwork, virtual environment

Introduction

In recent years, virtual environments (VEs) have become increasingly popular due to technological advances in graphics and user interfaces [9]. One of the many valuable uses of VEs is teamwork training. The members of a team can be located wherever it is most convenient for them (e.g., at home) and solve a simulated task in the VE collaboratively, without physically having to travel to a common simulation facility. Medical schools have realized this advantage and, for example, created medical simulations within Second Life [3,4].

Communication is a vital aspect of teamwork, and failure of communication can lead to critical or even lethal events, not only in medicine [5]. An ideal VE would therefore facilitate all communication channels that exist in reality – verbal as well as non-verbal. Due to technical limitations, this is not possible, and therefore, existing communication in VEs is currently mostly limited to voice. Other channels like text chat, avatar body

[1]Corresponding Author: Stefan Marks; E-mail: stefan.marks.ac@gmail.com.
[2]E-mail: john.windsor@auckland.ac.nz
[3]E-mail: b.wuensche@auckland.ac.nz

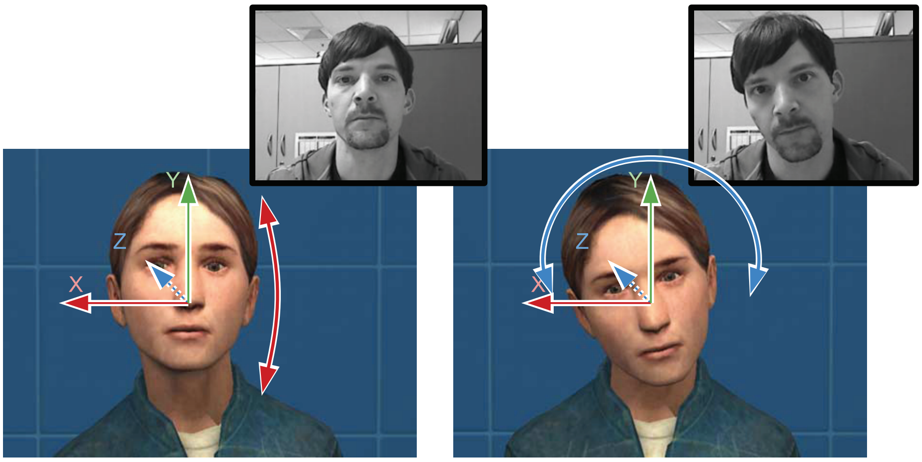

Figure 1. Examples of the user controlling the head movement of the avatar by physical head movement.

gestures, and facial expressions have to be controlled manually and thus do not reflect the real-time communicative behavior of the user, especially unconscious behavior.

Analysis of communication in medical teamwork has shown that nonverbal communication cues like gesture, touch, body position, and gaze are equally important to verbal communication in the analysis of the team interactions [1]. VEs that do not consider those nonverbal channels are likely to render the communication among the team members less efficient than it would be in reality.

Additional disadvantages of existing VE-based solutions are complex user interfaces, e.g., in Second Life, that require the user to learn first how to use the simulator, before actually being able to participate in simulations [2]. Complex menus or a list of keystrokes make it difficult to operate the VE and the user constantly has to translate a desired action or gesture into a corresponding mouse movement or the push of a button. Specifically camera control can pose a problem, especially for users that are not familiar with computer games.

In this paper, we describe the design and implementation of an inexpensive VE for training collaborative tasks, specifically medical tasks. We then propose an extension of the VE by camera-based head tracking to increase the "communication bandwidth" and the ease of use. This extension does not require a complex and expensive special camera – any customer-level webcam is sufficient.

Head tracking measures the position and the orientation of the user's head relative to the camera and the screen. The rotational tracking information can be used to control the head rotation of the user's avatar. That way, other users in the VE can see rotational head movement identical to the movement actually performed physically by the user, like nodding, shaking, or rolling of the head (see Figure 1).

The translational tracking information can be used to control the view "into" the VE. This so called Head Coupled Perspective (HCP) enables intuitive control, like peeking around corners by moving the head sideways, or zooming in by simply moving the head closer to the monitor (see Figure 2). The use of head tracking information has therefore the potential to simplify the usage of a VE by replacing non-intuitive manual view control by intuitive motion-based view control. This makes it unnecessary for the user to learn

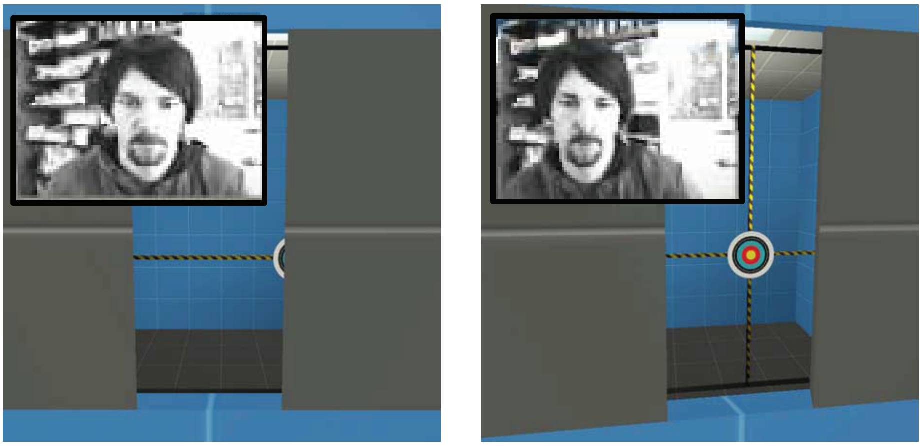

Figure 2. Screenshots demonstrating how the view of the VE can be controlled by simple head movement. The participant intuitively moves the head sideways to see the hidden disc.

additional key strokes, or joystick/mouse movements. Especially in medical applications, this method has the advantage of freeing the hands of the user, enabling the use of other simulated instruments or tools, e.g., an endoscope.

At the end of this paper, we also present a summary of the user studies we have conducted, demonstrating the benefits of camera-based avatar control for VEs.

1. Methods and Materials

We used a modified game engine for the implementation of our VE after conducting an extensive evaluation of possible platforms that facilitate multi-user simulation scenarios with a reasonable realism for interaction with simulated physical objects [6]. For this, we performed a requirement analysis and identified the most important features of suitable engines to be stability, availability, flexibility, proper documentation, networking ability, and advanced physical simulation capabilities. We surveyed 280 game engines based on these requirements and developed some VE prototypes. As a result, we chose the Source Engine [10], which fulfils all of the above, provides a good level of realism, a physics engine that allows collaborative manipulation of simulated objects (e.g., surgical instruments), and readily provides mechanisms for creating a simulation server that several clients can connect to.

This game engine was extended by software modules that receive the input of a webcam and measure the user's head position and orientation relative to the monitor (see Figure 3). A design goal was to keep these modules completely separate from the game engine so that they can easily be utilized by other VEs and simulation engines. Also, the data model produced by the head tracking module is flexible and requires only little bandwidth when being transmitted to the simulation server. The tracking data sent to the game engine is processed in two ways:

First, the translational information is used to control the view of the virtual environment on the screen in real-time. When the user moves sideways, the view also moves sideways, giving a more realistic 3D impression of the scene on the screen. Furthermore,

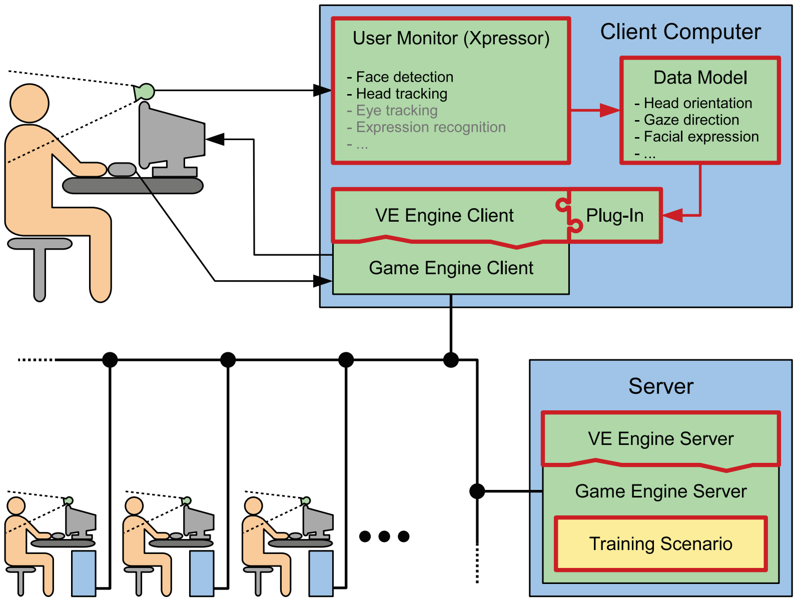

Figure 3. The functional blocks of the simulation framework. Blocks marked with a thick border have been developed specifically for this project.

moving closer to the screen allows for "zooming in" on objects, e.g., to inspect them in more detail (see Figure 2). Using a combination of these movements, the user can look at or around objects in an intuitive way by using physical head movement as in real life. Furthermore, this control mechanism, called head-coupled perspective (HCP), frees up the hands which can then be used to, e.g., control simulated surgical instruments like endoscopes.

We conducted user studies to measure the effect and usability of HCP for controlling the view of a VE, comparing manual control by keyboard and mouse with HCP [7]. The results are presented in the next section.

Second, the rotational information of the tracking data is used by the game engine to control the head rotation of the user's avatar (see Figure 1). This rotation data is transmitted over the network to all clients that are connected to the simulation server, allowing all other users of the training scenario to see each others' head movement in real-time.

This rotational information replicated on all clients opens up an additional channel for nonverbal communication cues. Verbal messages like confirmation or rejection can now be emphasized by nodding or shaking of the head. Furthermore, this data, together with all interactions, movements, and verbal communication, can be recorded on the simulation server and analyzed during the teamwork debriefing phase.

An additional advantage of camera-based avatar control is that the user does not need to be aware of those nonverbal communication cues in order to, e.g., trigger any avatar animations manually. VEs which use manual command input for avatar control, e.g., by menus or keyboard shortcuts, force the user to constantly be aware of his or her nonverbal expressions. This puts additional workload on the user, distracting from the task at hand. Alternatively, if the user does not use this functionality because of the increased cognitive load, the avatars in the training scenario remain sterile and emotionless.

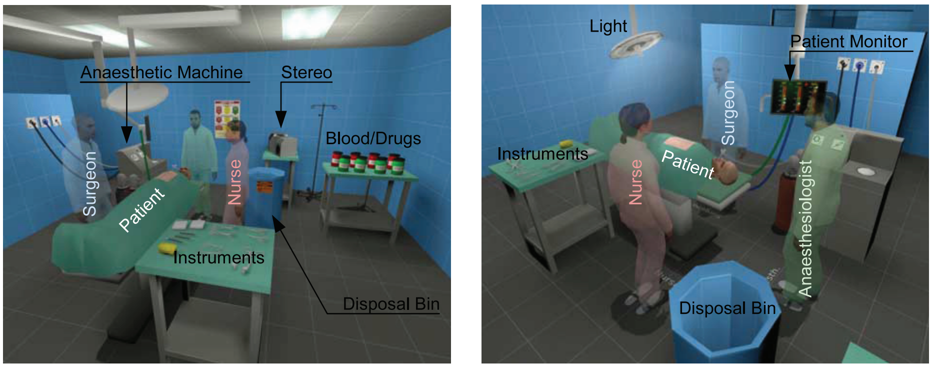

Figure 4. View of the Virtual Environment running our simplified surgical teamwork training scenario.

Our camera-tracking framework also allows for the inclusion of eye gaze and facial expressions, adding more communication channels to the simulation. Even with relatively inexpensive webcams, it is possible to track the point that the user is looking at on the screen. This can be projected into the VE, calculating the point and the object that the user is looking at and in turn modifying the avatar's eye gaze to look at the same point. This enables other participants to get a better idea of what the other team colleagues are looking at. This also includes eye contact in face-to-face conversations, where a slight shift in the gaze might signalize full attention to the speaker or distraction by something going on in the background.

Facial expressions can also be mapped onto the avatars, communicating smiling, frowning, raised eyebrows, and other emotions and expressions.

To analyze the effect of adding head rotation to the VE on users and their teamwork, we conducted a study with 27 participants working on a teamwork task with and without the use of camera-based head tracking [8] (See Figure 4). The results of this study are also summarized in the next section.

2. Results

The experimental data and user feedback from our first study, focusing on HCP, lead to the following results, documented in full detail in [7]:

- Users were able to control their view of the VE with the same precision using HCP compared to manual control by mouse and keyboard.
- Using HCP, it took the participants slightly longer on average to perform viewing tasks compared to manual control. However, participants with little or no experience with computer games were significantly faster than with manual control.
- The participants stated that HCP was more intuitive to use than manual control and improved their 3D perception of the VE and their feeling of immersion.

The results of our study about the effect of camera-controlled avatars and HPC on a teamwork training scenario can be summarized as follows (see paper [9] for more details):

- Avatars with tracking-based head movement are perceived as more natural than avatars without tracking-based head movement.
- Avatars with tracking-based head movement facilitate a communication that is perceived as more natural than with avatars without tracking-based head movement.

We could not measure a significant influence of the camera-based head tracking on the teamwork performance. This was most probably due to learning effects masking the effect of the tracking as discussed in the paper.

3. Conclusions and Discussion

In this paper, we have addressed a shortcoming of current VE-based teamwork simulators, the lack of nonverbal communication channels, by utilizing inexpensive consumer-level hardware and an extension to the simulation framework to create an intuitive and non-obtrusive way of adding nonverbal communication. In addition, the extension allows the user to easily navigate and control their view of the VE without using any mouse or keyboard control. The participants of both of our studies perceived an improvement in the usability and naturalness of the VE and an improvement of the quality of communication, but did not experience any major negative limitations imposed upon them by the tracking technology.

Not all possibilities of this new technology have been explored yet. For our second user study, the head rotation was simply mapped onto the avatars without any further semantic analysis. One future extension will be to calculate the actual target of the user's head direction on the screen and to use that information to transform the tracking data so that the user's avatar looks at the same target in the VE. An additional level of realism would be introduced by eye tracking and a similar transformation of the avatar's gaze. Body gestures can be included by using advanced consumer-level hardware like the Microsoft Kinect. An extension of our framework by automatic recognition of facial expressions is also possible. We expect that the inclusion of one or more of the above communication channels can further increase at least the perceived naturalness and ease of use – if not more.

References

[1] John Cartmill, Alison Moore, David Butt, and Lyn Squire. Surgical Teamwork: Systemic Functional Linguistics and the Analysis of Verbal and Nonverbal Meaning in Surgery. *ANZ Journal of Surgery*, 77(Suppl 1):A79–A79, May 2007.

[2] Emily Conradi, Sheetal Kavia, David Burden, Alan Rice, Luke Woodham, Chris Beaumont, Maggi Savin-Baden, and Terry Poulton. Virtual patients in a virtual world: Training paramedic students for practice. *Medical Teacher*, 31(8):713–720, 2009.

[3] Douglas Danforth, Mike Procter, Robert Heller, Richard Chen, and Mary Johnson. Development of Virtual Patient Simulations for Medical Education. *Journal of Virtual Worlds Research*, 2(2):3–11, August 2009.

[4] Scott Diener, John Windsor, and David Bodily. Design and Development of Medical Simulations in Second Life and OpenSim. In *EDUCAUSE Australasia Conference*, Perth, Australia, May 2009.

[5] Lorelei Lingard, Sherry Espin, S Whyte, Glenn Regehr, G R Baker, Richard Reznick, J Bohnen, B Orser, D Doran, and E Grober. Communication failures in the operating room: An observational classification of recurrent types and effects. *Quality & Safety in Health Care*, 13(5):330–334, March 2004.

[6] Stefan Marks, John Windsor, and Burkhard Wünsche. Evaluation of Game Engines for Simulated Surgical Training. In *GRAPHITE '07: Proceedings of the 5th international conference on Computer graphics and interactive techniques in Australia and Southeast Asia*, pages 273–280, New York, NY, USA, December 2007. ACM.

[7] Stefan Marks, John Windsor, and Burkhard Wünsche. Evaluation of the Effectiveness of Head Tracking for View and Avatar Control in Virtual Environments. *25th International Conference Image and Vision Computing New Zealand (IVCNZ) 2010*, November 2010.

[8] Stefan Marks, John Windsor, and Burkhard Wünsche. Head Tracking Based Avatar Control for Virtual Environment Teamwork Training. In *Proceedings of GRAPP 2011*, 2011.

[9] Paul R. Messinger, Eleni Stroulia, Kelly Lyons, Michael Bone, Run H. Niu, Kristen Smirnov, and Stephen Perelgut. Virtual Worlds – Past, Present, and Future: New Directions in Social Computing. *Decision Support Systems*, 47(3):204–228, June 2009.

[10] Valve Corporation. Source Engine, 2007. `http://source.valvesoftware.com`.