Virtual Reality System for Treatment of the Fear of Public Speaking Using Image-Based Rendering and Moving Pictures

JAE M. LEE,[1] JEONG H. KU,[1] DONG P. JANG,[1] DONG H. KIM,[2]
YOUNG H. CHOI,[3] IN Y. KIM,[1] and SUN I. KIM[1]

ABSTRACT

The fear of speaking is often cited as the world's most common social phobia. The rapid growth of computer technology enabled us to use virtual reality (VR) for the treatment of the fear of public speaking. There have been two techniques used to construct a virtual environment for the treatment of the fear of public speaking: model-based and movie-based. Virtual audiences and virtual environments made by model-based technique are unrealistic and unnatural. The movie-based technique has a disadvantage in that each virtual audience cannot be controlled respectively, because all virtual audiences are included in one moving picture file. To address this disadvantage, this paper presents a virtual environment made by using image-based rendering (IBR) and chroma keying simultaneously. IBR enables us to make the virtual environment realistic because the images are stitched panoramically with the photos taken from a digital camera. And the use of chroma keying allows a virtual audience to be controlled individually. In addition, a real-time capture technique was applied in constructing the virtual environment to give the subjects more interaction, in that they can talk with a therapist or another subject.

INTRODUCTION

THE FEAR OF SPEAKING is often cited as the world's most common social phobia. Public speaking anxiety is the fear of social situations and interpersonal interaction that can automatically bring on self-consciousness, judgment, evaluation, and criticism when a person speaks in front of the general public. People having such fear often show symptoms such as shame and timidity in daily personal relationships. In addition, their heart rate and blood pressure are elevated when they speak in front of the public.[1] People with such fear are often afraid that other people will be puz-zled or feel insulted by their speech, and they also fear that they would be underestimated for their mistakes.[2,3] This fear does not appear to be limited by age, gender, economic, or educational variables.

Though there are many treatment methods, virtual reality therapy (VRT) has been used in treating psychological disorders, including the fear of public speaking. Unlike traditional treatment methods, VRT can provide an exact replica of the anxiety-producing situation for the subject, which is impossible to do in the real world. Many researchers have shown that VRT is successful in reducing the fear of public speaking.[3,4]

[1]Department of Biomedical Engineering, Hanyang University, Seoul, Korea.
[2]School of Engineering and Applied Science, George Washington University, Washington, D.C.
[3]Department of Psychiatry, Seoul Paik Hospital, Inje University, Seoul, Korea.

There have been two techniques used to develop a VR system for the treatment of the fear of public speaking: model-based and movie-based.

The model-based technique has a disadvantage in that the virtual environment and audiences designed with it are unreal because the virtual audiences are made with limited polygon data and unnatural motions.[3] In the case of the movie-based technique, which uses a movie file shot with real audiences, there is also a disadvantage in that that virtual audiences cannot be controlled individually because all virtual audiences are included in one movie file.[5,6]

To sort out the abovementioned problems, this paper presents a virtual environment made by using image-based rendering (IBR) and chroma keying simultaneously. In addition, a real-time capture technique is presented, which gives subjects more interaction possibilities because they can talk with a therapist or another subject.

MATERIALS AND METHODS

Image-based rendering

IBR is a popular way to simulate a visual VR experience. Instead of building and rendering a complete three-dimensional (3D) model of the environment, a collection of images is used to render a scene, which supports virtual camera motion. For example, a single cylindrical image surrounding the viewer enables the user to pan and zoom inside an environment created from real images.[7,8] A number of techniques have been developed for capturing panoramic images of real world scenes. The first way is to record an image onto a long film strip using a panoramic camera, which directly captures a cylindrical panoramic image.[9] A second way is to use a lens with a very large field of view such as a fisheye lens.[10] A third way, a less hardware-intensive method for constructing full-view panoramas, is to take many regular photographic or video images in order to cover the whole viewing space. These images must then be aligned and composed to make panoramic images using an image mosaics or "stitching" algorithm.[11] In this paper, the last method was used; only a regular still camera is needed.

To capture a full 360° (horizontal) × 180° (vertical) view of the scene as in Figure 1e, many images were captured in rows. That is, in addition to horizontally capturing images in a 360° circle, we captured rows of images with the camera vertically tilted up and down. A panorama might consist of five rows of images, as in Figure 1a. Figure 1b shows that each row of images consists of 12 shots. The images in each row are captured at 30° increments:
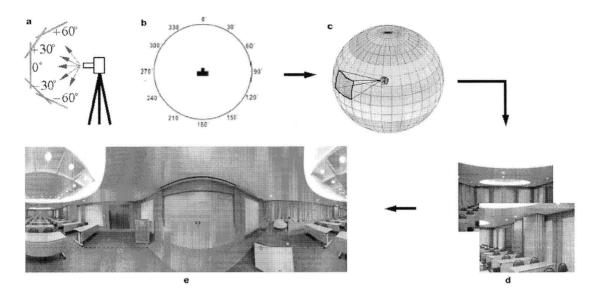
- The first row is captured with the camera tilted up at a 60° pitch.
- The second row is captured with the camera tilted up at a 30°pitch.
- The third row is captured with the camera leveled at a 0° pitch.
- The fourth row is captured with the camera tilted down at a −30° pitch.
- The fifth row is captured with the camera tilted down at a −60° pitch.

After stitching the images by each row, as in Figure 1d, we stitched all the rows together to produce the final image. But the final panorama image cannot reach to the topmost and bottommost. For covering those parts, two images, taken straight up and down, are added to the top and bottom of the panorama image (Fig. 2A, B).

Movie-making and chroma keying

The movie-based VR therapy system, developed to treat an anthrophobia in previous studies, has a disadvantage in that a therapist cannot control every virtual audience individually because all virtual audiences belong to one movie scene.[5,6] But if each audience is shot individually, it will be possible to flexibly control the virtual audiences and compose many more situations.

Figure 3a shows a virtual audience made of film (with a blue screen for its background), shot through a digital camcorder. And then, the original scene (720 × 480 pixels), shot through a camcorder, was clipped of its outer frames and scaled down to 128 × 128 pixels. Because the movie scene was blurred after this

FIG. 1. Image stitching. (a) Five rows are captured with the camera vertically. (b) Each row of images consists of 12 shots horizontally. (c) Three-dimensional diagram. (d) Stitching the images by each row. (e) Stitching all the rows together to produce the final image.

procedure, a sharpening algorithm was applied to the movie scene.

To make all regions except the actor's body be transparent, as in Figure 3b, if the intensity of pixels in a movie scene is over a specific threshold value, the pixel becomes transparent by application of the alpha channel. This is the point of chroma keying technique. A total of 12 virtual audiences were captured, and each audience has nine kinds of actions, which are performed for about 10 sec. Those actions consist of a neutral action (handclapping), four positive actions (laughing, taking an interest, understanding, acclaiming), and four negative actions (indifference, chatting, dozing, yawn-
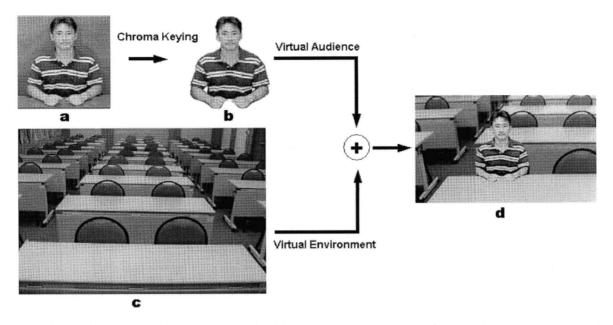
ing). An operator can select a specific action of each virtual audience during a subject's speech. By these methods, a subject can see the audiences responding positively or negatively according to his or her speech.

Real-time capture

So far, most studies have included an animated avatar as a virtual audience. The avatar acted only as previously arranged. These methods cannot, however, show suitable and various actions, except actions made previously. For solving that problem, the real-time capture system (Fig. 4) (shot through a PC camera) was applied to a virtual environment. This system will be especially helpful on giving subjects more interaction with other people or a therapist in real time. One can talk with a subject and express various responses to the subject's speech. Being shown another person responding in real time will enable a subject to feel more presence and interaction than when he or she sees only the animated avatar. (Fig. 5).

Intel Computer Vision Library was used to capture high-speed color video frames into memory (e.g., 30 fps at 160×120). Captured frames were saved to AVI sequences, and then Captured AVI sequences were mapped as a texture into the surface designated previously.

FIG. 2. Capturing a shot straight up and down. (A) Capturing a shot straight up. (B) Capturing a shot straight down.

FIG. 3. Chroma keying. **(a)** Making a movie with a blue screen. **(b)** Application of chroma keying. **(c)** The virtual environment. **(d)** Taking a virtual audience's position in the virtual environment.

Implementation of the virtual reality system

The hardware used for creation of the virtual world was a personal computer system having a Pentium III 866-MHz CPU, 256 MB RAM, and a 3D acceleration graphic card. To enhance the immersion of the user in the virtual environment, a head-mounted display (HMD) and head tracker having three degrees of freedom were used. 3D StudioMax (Kinetix) was used as the 3D modeling tool. A SONY MVC-CD1000 digital camera and a SONY DCR-VX2000 digital camcorder were used to capture the real images for image stitching and to get the movie files. We built the real-time virtual environment with Visual C++ 6.0, DirectX 7.0a SDK (Software Development Kit).

Immediately on starting this system, a subject can see 12 virtual audiences sitting at their desks, and can also see another person staring at him or her directly. That person could be another subject or a therapist being captured by PC camera in real time. A therapist can give a

subject directions during treatment and help him or her cope with the events.

A therapist operates virtual audiences to act positively and negatively according to the success of the speech.

CONCLUSION

Compared with previous work, our VR public speaking system has several outstanding attributes. First, our virtual environment system is more realistic than the existing system using many graphical textures because real photos taken from a digital camera were used by IBR technology. Second, an operator can select a specific action of each virtual audience during a subject's speech. It is possible to flexibly control the virtual audiences and compose many more situations. Third, being shown another person responding in real time through a PC camera enables the subject to feel more
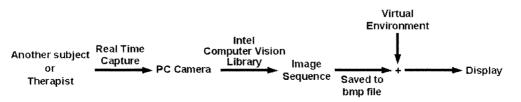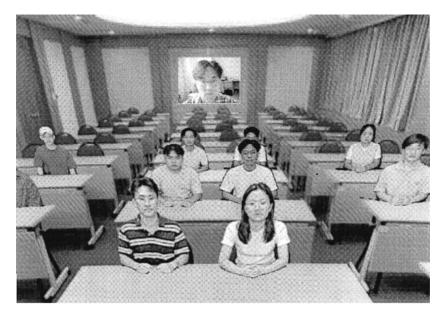

FIG. 4. Real-time capture.

FIG. 5. The implementation scene of a virtual reality system.

presence and interaction than when he or she sees the animated avatars. Therefore, this system should help people suffering from the fear of public speaking.

In the future, additional study is needed to conduct a clinical test on subjects with anthrophobia to verify whether this system has the effect of treating those subjects successfully or not. In addition, virtual audiences of various ages and ethnicities must be studied.

ACKNOWLEDGMENTS

REFERENCES

1. Davidson, R.J., Marshall, J.R., Tomarken, A.J., et al. (2000). While a phobic waits: regional brain electrical and autonomic activity in social phobics during anticipation of public speaking. *Biological Psychiatry* 47:85–95.
2. Antony, M.M. (1997). Assessment and treatment of social phobia. *Canadian Journal of Psychiatry* 42: 826–834.
3. Slater, M., Pertaub, D.P., & Steed, A. (1999). Public speaking in virtual reality: facing an audience of avatars. *IEEE Computer Graphics & Application* 19:6–9.
4. North, M.M., North, S.M., & Coble, J.R. (1998). Virtual reality therapy: an effective treatment for the fear of public speaking. *International Journal of Virtual Reality* 3:2–6.
5. Anderson, P., Rothbaum, B.O., & Hodges, L.F. (2000). Social phobia: virtual reality exposure therapy for fear of public speaking. Presented at the Annual Meeting of the American Psychological Association, Washington, D.C.
6. Jo, H.J., Ku, J.H., Jang, D.P., et al. (2001). Movie-based VR therapy system for treatment of anthropophobia. Presented at the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society.
7. Chen, S.E. (1995). QuickTime VR—an image-based approach to virtual environment navigation. *Computer Graphics* Aug:29–38.
8. McMillan, L., & Bishop, G. (1995). Plenoptic modelling: an image-based rendering system. *Computer Graphics* Aug:39–46.
9. Meehan J. (1990). *Panoramic photography.* Watson-Guptill.
10. On-line data. Available: www.behere.com.
11. Szeliski, R., & Shum, H.-Y. (1997). Creating full-view panoramic image mosaics and environment maps. In: *Proceedings of the 24th Annual Conference on Computer Graphics & Interactive Techniques.* pp. 251–258.

Address reprint requests to:
Dr. Sun I. Kim
Sungdong P.O. Box 55
Seoul, 133-605, Korea

E-mail: sunkim@hanyang.ac.kr