ADVANCES IN SHADER LAMPS AVATARS FOR TELEPRESENCE

Ryan Schubert *, Greg Welch †*, Peter Lincoln *, Arjun Nagendran †, Remo Pillat †, and Henry Fuchs *

* The University of North Carolina at Chapel Hill, Department of CS
Chapel Hill, NC 27599-3175 USA Email: {res, welch, plincoln, fuchs}@cs.unc.edu

† The University of Central Florida, IST and Department of EECS
Orlando, FL 32826-3281 USA Email: welch@ucf.edu, {arjun, rpillat}@cs.ucf.edu

ABSTRACT

Since the introduction of the *animatronic Shader Lamps Avatar* in 2009, several aspects of the system's design have been improved to create a more effective telepresence experience. This includes an analysis of the tradeoffs between front and rear projection and between rigidly aligning the projector with the display surface or needing to track it. A new control mode is implemented for the pan-tilt unit that enables the avatar head pose to better emulate real human motion and a new hardware device allows the motorized wheelchair that carries the avatar to be controlled by the remote inhabiter. With the recognition of the importance of high quality audio, the audio components for both the inhabiter and the avatar have also been updated. Finally, some preliminary live tests are presented and areas of continued research are discussed.

Index Terms — Teleconferencing, Avatars, Animatronics, Mobile robots

1. INTRODUCTION

There are many display paradigms which can be effective in conveying the 3D appearance of a remote user in a telepresence scenario. Most such systems provide a 3D virtual image of a user, as with stereoscopic [1], auto-stereoscopic [2], or holographic displays [3]. However, the *animatronic Shader Lamps Avatar* (SLA) was introduced as a telepresence paradigm that provided a physical, three-dimensional representation of a user, spatially augmented with virtual imagery to provide a variable appearance [4]. This allowed for many of the benefits of physical presence, such as accurately conveying head pose and gaze direction [5] for multiple viewers simultaneously, combined with the flexibility of being able to display the appearance of any one of many possible users. This was achieved by using a 3D face-shaped display surface onto which live video of any given user's face could be projected, giving the physical avatar's face the 3D appearance of the user.

Since the original proof-of-concept animatronic SLA system was built, several changes and improvements have been incorporated into two new mobile animatronic SLAs and a new setup for the inhabiter has been created. This paper will discuss the more significant recent design tradeoffs, improvements, and ongoing issues with animatronic SLA 3D telepresence systems, including a discussion of front versus rear projection, rigid alignment between the projector and display surface, a new pan-tilt-unit control method for more realistic and accurate avatar head movement, and new developments in remote mobility.

2. RELATED WORK

A more detailed discussion about related work prior to 2011 can be found in [4] and [6] but we provide a summary here, as well as some newer work from 2011 and 2012. The basic idea of *Shader Lamps* is to use projected imagery to alter the appearance of physical 3D objects [7]. *Hypermask* is described as a system in which animated, expressive synthesized faces were projected onto a face-shaped mask, worn by an actor [8]. Lincoln et al. also present a fair amount of previous work looking at emulating human motion, such as some of Hiroshi Ishiguro's work at Osaka University's Intelligent Robotics Laboratory developing complex animatronic humanoid robots such as the Repilee [9] and teleoperated Geminoid. While the robots developed by Ishiguro could be made to look like a single synthetic person, the WD-2 robot from the Takanishi Laboratory has a total of 56 degrees of freedom for a flexible face-shaped mask, allowing the entire face surface to closely match any specific typical user [10]. Susumu Tachi's TELESAR project uses humanoid animatronic robot surrogates with retro-reflective surfaces showing the remote inhabiter through autostereoscopic projection [11]. Al Moubayed et al. projected animated human faces onto a 3D physical head, and investigated the perceived gaze direction [12]. In particular, they showed that the 3D model resulted in an accurate perception of gaze direction, unlike the Mona Lisa gaze effect typically seen when using a 2D surface to display 3D face imagery. Sirkin et al. developed a videoconferencing system that used a 2D display for the user's face, but the display could swivel to indicate gaze direction [13].

3. PROTOTYPE SYSTEM

The full telepresence system can be broken into two parts, the *inhabiter station* and the *avatar*, with full-duplex video and audio channels between them.

3.1. Inhabiter station

Seen in Figure 1, the *inhabiter station* is comprised of three 70-inch LCD TVs, one or more color cameras to capture live imagery of the inhabiter, a tracking system, a microphone, and earbuds. The three 70-inch TVs are each individually rotated sideways and arranged in a half-circle to provide roughly a 180-degree horizontal field of view for the inhabiter. The large size of the displays was selected for its immersive field of view and life-size representation of remote objects. The prototype system uses NaturalPoint's OptiTrack system to acquire head position and pose of the inhabitor.

Figure 1. Inhabiter station setup

Figure 2. Mobile avatar setup

3.2. Avatar

The *avatar* side, shown in Figure 2, has a face-shaped display surface mounted on a pan-tilt-unit, stereo microphones, a speaker, and three wide-angle HD cameras to capture the environment in front of the avatar (each camera maps directly to one of the three TVs in the inhabiter station). The three HD cameras provide significantly better image quality and field of view than the two non-HD cameras used in the original prototype system. The entire avatar-side system is mounted on a motorized cart, and capable of being powered by an on-board battery. Video from the three cameras as well as the inhabiter's face imagery can be streamed over a wireless network, allowing for free mobility in the remote environment. The pan-tilt-unit is used to match the current pose of the tracked inhabiter's head while live imagery from the inhabiter is projected on the display surface. This gives the inhabiter the ability to interact with multiple people through a physical 3D presence at the remote location.

3.3. Front versus rear projection

The original proof-of-concept animatronic SLA used front projection onto a solid styrofoam head-shaped display surface [4]. This was a natural extension of previous Shader Lamps works, and allowed for a simple display surface material. In practice, however, this paradigm has two potential drawbacks. For a more compact SLA design (necessary for a mobile system), the projector requires a relatively short focal distance such that it can be

placed near the avatar face, and this can cause the projector itself to occlude the avatar for certain viewer positions. Additionally, any close proximity interaction with the avatar can block the projector with respect to the display surface.

One solution to this problem is to use a thin, semi-transparent, face-shaped shell of a display surface and have the projector mounted behind the surface for rear projection. While this complicates the properties of the display surface, it eliminates both forms of occlusion that can occur in a front projection setup. In both front and rear-projected cases, the material properties of the display surface itself can have a significant effect on the quality of the end result. Inter-reflection and specular reflections can cause distracting visual cues that do not match the reflectance or surface characteristics of the target appearance.

3.4. Tracked display surface versus rigid alignment

The original animatronic SLA also tracked the display surface in order to properly align imagery from the projector (calibrated in tracker space) onto the surface, which could change position and orientation due to the pan-tilt-unit [4]. However, rather than allowing the projector and display surface to move independently, they can be mechanically joined such that both are driven by the pan-tilt-unit and the relative alignment remains rigid. This allows for the computation of a one-time calibration between the projector and display surface, which is unaffected by pan or tilt changes. Such a rigid alignment requires a more powerful pan-tilt unit to handle the additional weight of the projector and any necessary optics.

A rear-projection setup lends itself particularly well to a rigid alignment, and we now have two mobile animatronic SLA systems, one of which is a tracked front projection system while the other is a rigidly aligned rear projection system that uses a wide angle lens and a mirror to fold the optical path of the projector. Both animatronic SLA display systems can be seen in Figure 3.
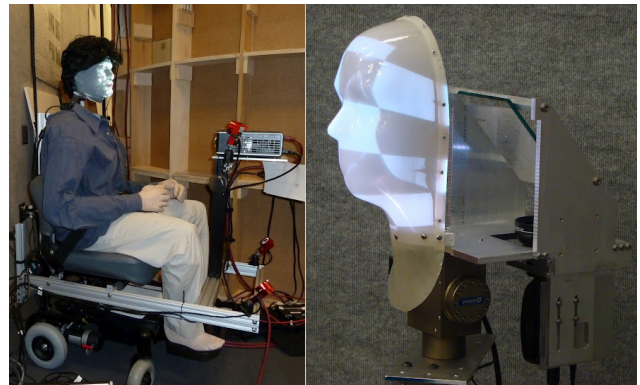

Figure 3. Front projected, tracked display surface compared to rear projected, rigidly aligned display surface

3.5. Pan-tilt unit control

In order to handle the additional weight of the projector/head assembly, the rigid-aligned rear-projection animatronic SLA is equipped with a Schunk PW 70 pan-tilt unit. Head pose data from the inhabiter is received by the Schunk which responds by minimizing the error between its current orientation and the received orientation in minimum time. In [4, 6] it was operated in position control mode, i.e. new target orientation angles were commanded at 100 Hz.

Due to high-frequency noise in the head tracking data, movements were fairly jerky and discontinuous. Delay was increased because the unit internally uses the orientation targets to plan independent position/velocity/acceleration trajectories, and the Schunk's internal brake was regularly and automatically triggered to hold the head in a stable position.

To achieve smooth and responsive control, velocity control mode was used. In this mode, the Schunk unit still receives target orientations at 100 Hz, but the achievement of orientation targets is validated through an external control system. The time-critical execution of the control loop is handled in a separate thread, thereby providing fine-grained management of the motion trajectories.

Target orientations are received by the tracking system at the inhabiter and the controller uses the orientation error (difference between desired and current position) to set appropriate target velocities for each axis of the Schunk. The control system's parameters were tuned by observing the system's characteristics in response to a step position command. Each movement axis required a slightly different control scheme owing to varying apparent inertias (an effect of gravity); the tilt axis uses a proportional (P) controller, while the pan axis employs a proportional-derivative (PD) controller. Variations of standard tuning rules were used to provide the desired response without overshoot or oscillations. It should be noted that, as with any long distance system operation, errors due to latency in the data transfer are a potential issue, but have not yet been addressed in this work.

3.6. Remote mobility

A major aspect of physical presence is our ability to freely move in a given space. Human mobility provides the underpinning for many social as well as utilitarian interactions. To this end, we present a system that allows mobile control of the avatar by a remote inhabiter while also making it easy to revert to local wired operation when desired.

The mobile avatar platform is based on a commercial electric scooter, the DriveMedical Sunfire Plus EC, which provides local control through a PG Drives Technology (PGDT) VR2 joystick. Built-in safety features and a proprietary protocol for the communication between joystick and motor controller make it difficult to directly control the chair via a computer. The 2-axis potentiometers in the joystick knob were therefore emulated through three analog outputs: one for each axis and one for the center tap voltage. This solution provides enough flexibility to allow fully autonomous computer control, yet retains all proprietary features like velocity control and brake applications. In addition, this emulation principle can be transferred to any mobile platform utilizing potentiometer-based joystick control.

3.7. Full-duplex audio

High-quality audio plays an important role in allowing for natural interaction between the inhabiter and people around the avatar. Stereo audio provides an important 3D directional cue when interacting with multiple people and, especially in combination with high fidelity sound, can greatly increase the inhabiter's sense of immersion in the remote environment [14]. Likewise, high fidelity capture of the inhabiter's voice from the avatar itself creates a more faithful and human-like reproduction of the inhabiter.

Although the system described in [6] used stereo microphones to capture the audio of the environment around the avatar, the quality of the audio system was not a high priority. A substantial im-

provement to the overall effectiveness of the system has now been achieved using higher quality audio components. Non-networked (tethered) tests have been done capturing stereo sound with two MXL FR355K lavalier mics which are then fed through a four-channel Mixer and directly to a pair of earbuds for the inhabiter. The mics are positioned near the sides of the SLA head to mimic the virtual ear positions of the inhabiter. Inhabiter voice capture is done with a Countryman EMW Omni Classic lavalier mic that feeds through a PYLE PT710 amplifier and then directly to a Klipsch Gallery G-12 speaker mounted behind the avatar head. Although USB mixers can be used to stream the audio over traditional teleconferencing channels, these are not typically designed to maximize the fidelity of the sound.

3.8. Rendering method

The rendering method for mapping the inhabiter imagery onto the avatar display surface is unchanged from what is described in [4]. It still requires the creation of topologically identical geometric models of both the 3D display surface and each inhabiter who wishes to use the system. The capture cameras in the inhabiter station setup are calibrated in the coordinate system of the tracker that is used to track the head. In this way, the 3D point on the model of the inhabiter can be calculated given any point in the camera image. Similarly, the projector on the avatar is calibrated in the space of the tracked display surface, or, in the case of a rigid alignment, the static calibration between the projector and the display surface is manually determined as a one-time process. With the above information, the standard graphics method of texture projection can be employed to map the 2D camera image onto the 2D projector image such that the salient features from the inhabiter fall in the correct locations on the display surface.

4. EXPERIMENTAL RESULTS

In May 2012, a rigidly-aligned, rear-projection animatronic SLA located in Singapore was inhabited by a user from an inhabiter station in the United States in an early intercontinental test of an animatronic SLA system for long distance telepresence. The experiment included streaming video from all three HD cameras mounted on the avatar along with streaming video of the inhabiter being mapped onto the avatar's face. Stereo audio was not used in this experiment, and instead two-way audio was achieved via mono microphones over Skype. The user was able to successfully converse with a group of people, standing around the avatar, conveying attention by head and gaze direction.

5. AREAS OF ONGOING RESEARCH

For systems that use a tracked display surface without it being rigidly attached to the projector, "sloshing" type misregistration continues to be a problem. Even on the inhabiter station side, video capture of the user's face suffers from similar artifacts due to tracker latency combined with camera image acquisition latency. Both types of latency-based error can manifest themselves as a temporary misregistration between features in the live imagery and geometric landmarks on the display surface. Such misregistration can lead to particularly distracting artifacts, especially in cases where high contrast imagery is incorrectly registered with areas of high geometric detail on the display surface. This causes noticeable distortion of the imagery, which changes over time based

on the inhabiter's head movement. We are currently looking at tradeoffs with using a display surface with less geometric detail to mitigate the error from this type of geometric distortion.

One downside to the approach presented here, is that a single, static display surface shape for any given avatar system is used. Due to the process of warping the imagery of the inhabiter to fit the 3D geometry of the display surface, there tends to be a "homogenization" of the appearance of users. Everyone looks, to varying degrees, like they have the same facial structure as the display surface itself. This appearance homogenization can also be further affected by the amount of geometric detail present in the display surface geometry. Techniques to more accurately represent different facial appearances on the same display shape, thereby effectively diminishing the apparent homogenization of facial geometry, form a part of the ongoing work. There are additional challenges with the audio system presented, such as echo cancellation with respect to the speaker and microphones mounted on the avatar, as well as latency issues. Currently no synchronization between the audio and the video, in either direction is ensured. Additionally, current and ongoing work includes working on more complex remote animatronics, to allow for the use of pointing or gestures and to better convey the body pose of the inhabiter, beyond just the head.

6. CONCLUSION AND ACKNOWLEDGEMENTS

A number of new advances in the use of animatronic SLAs as an effective telepresence system have been presented. This includes wide field-of-view, life-size imagery and high quality stereo audio for the inhabiter, and two new mobile animatronic SLA prototypes. The tradeoffs between front and rear projection and whether or not the display surface can move independently of the projector have been outlined. A new pan-tilt unit control method allowed the avatar to more accurately mimic the head motions of the inhabiter and the mobile avatar platform is now remotely controllable, further increasing the sense of physical presence and utility. Finally, several ongoing issues are being explored and addressed including research dealing with latency-based error, geometric distortion due to misalignment, differences between the display shape and the inhabiter's face as well as more complex remote animatronics.

7. REFERENCES

[1] Kibum Kim, John Bolton, Audrey Girouard, Jeremy Cooperstock, and Roel Vertegaal, "Telehuman: effects of 3d perspective on gaze and pose estimation with a life-size cylindrical telepresence pod," in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, New York, NY, USA, 2012, CHI '12, pp. 2531–2540, ACM.

[2] Wojciech Matusik and Hanspeter Pfister, "3d tv: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes," in *ACM SIGGRAPH 2004 Papers*, New York, NY, USA, 2004, SIGGRAPH '04, pp. 814–824, ACM.

[3] P. A. Blanche, A. Bablumian, R. Voorakaranam, C. Christenson, W. Lin, T. Gu, D. Flores, P. Wang, W. Y. Hsieh, M. Kathaperumal, B. Rachwal, O. Siddiqui, J. Thomas, R. A. Norwood, M. Yamamoto, and N. Peyghambarian, "Holographic three-dimensional telepresence using large-area photorefractive polymer," *Nature*, vol. 468, no. 7320, pp. 80–83, Nov. 2010.

[4] Peter Lincoln, Greg Welch, Andrew Nashel, Adrian Ilie, Andrei State, and Henry Fuchs, "Animatronic shader lamps avatars," in *Proceedings of the 2009 8th IEEE Intl Symposium on Mixed and Augmented Reality*, Washington, DC, USA, 2009, ISMAR '09, pp. 27–33, IEEE Computer Society.

[5] Samer Al Moubayed, Jens Edlund, and Jonas Beskow, "Taming mona lisa: Communicating gaze faithfully in 2d and 3d facial projections," *ACM Trans. Interact. Intell. Syst.*, vol. 1, no. 2, pp. 11:1–11:25, Jan. 2012.

[6] Peter Lincoln, Greg Welch, Andrew Nashel, Andrei State, Adrian Ilie, and Henry Fuchs, "Animatronic shader lamps avatars," *Virtual Real.*, vol. 15, pp. 225–238, June 2011.

[7] Ramesh Raskar, Greg Welch, and Wei chao Chen, "Tabletop spatially-augmented reality: Bringing physical models to life with projected imagery," in *In: Proceedings of the 2nd IEEE and ACM international workshop on augmented reality (IWAR99*, 1999, pp. 64–73.

[8] Tatsuo Yotsukura, Frank Nielsen, Kim Binsted, Shigeo Morishima, and Claudio S. Pinhanez, "Hypermask: Talking head projected onto real object," *The Visual Computer*, vol. 18, no. 2, pp. 111–120, April 2002.

[9] Robert Epstein, "My date with a robot," *Scientific American Mind*, vol. June/July, pp. 68–73, 2006.

[10] Takanishi Laboratory, "Various face shape expression robot," http://www.takanishi.mech.waseda.ac.jp/top/research/docomo/index.htm, August 2009.

[11] S. Tachi, K. Watanabe, K. Takeshita, K. Minamizawa, T. Yoshida, and K. Sato, "Mutual telexistence surrogate system: Telesar4 - telexistence in real environments using autostereoscopic immersive display -," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, sept. 2011, pp. 157–162.

[12] Samer Al Moubayed, Jonas Beskow, Jens Edlund, Björn Granström, and David House, "Animated faces for robotic heads: gaze and beyond," in *Proceedings of the 2010 international conference on Analysis of Verbal and Nonverbal Communication and Enactment*, Berlin, Heidelberg, 2011, COST'10, pp. 19–35, Springer-Verlag.

[13] D. Sirkin, G. Venolia, J. Tang, G. Robertson, T. Kim, K. Inkpen, M. Sedlins, B. Lee, and M. Sinclair, "Motion and attention in a kinetic videoconferencing proxy," in *Proceedings of the 13th IFIP TC 13 international conf. on Human-computer interaction - Vol Part I*, Berlin, Heidelberg, 2011, INTERACT'11, pp. 162–180, Springer-Verlag.

[14] Barbara Olasov Rothbaum, Larry Hodges, Renato Alarcon, David Ready, Fran Shahar, Ken Graap, Jarrel Pair, Philip Hebert, Dave Gotz, Brian Wills, and David Baltzell, "Virtual reality exposure therapy for ptsd vietnam veterans: a case study.," *Journal of Traumatic Stress*, vol. 12, no. 2, pp. 263–71, 1999.