Animatronic Shader Lamps Avatars

Peter Lincoln   Greg Welch   Andrew Nashel   Adrian Ilie   Andrei State   Henry Fuchs

The University *of* North Carolina *at* Chapel Hill
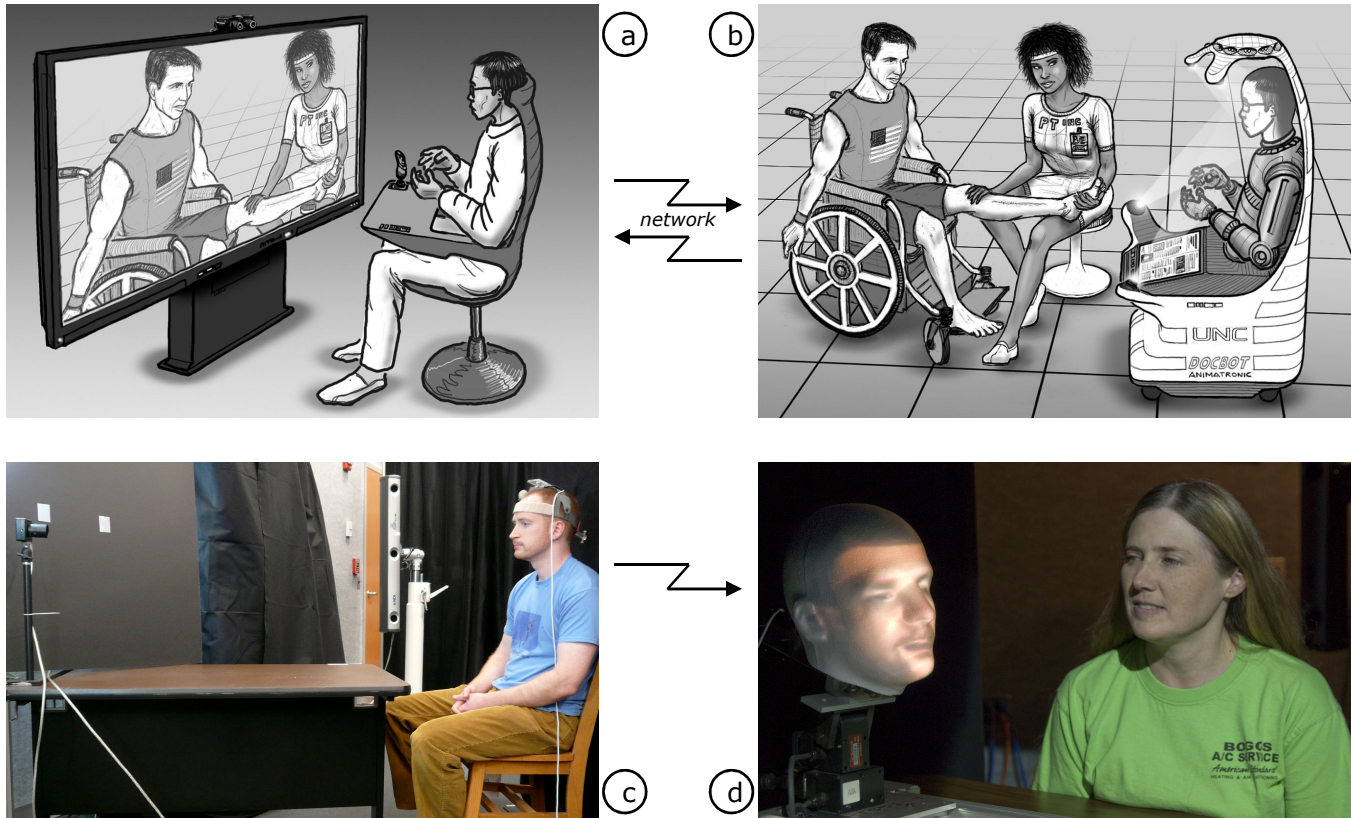Department of Computer Science*

Figure 1: The upper images conceptually illustrate one possible use of animatronic Shader Lamps Avatars (SLA): full-duplex telepresence for medical consultation. The physician in (a) interacts with a remote patient and therapist in (b) by means of a camera-equipped SLA. The SLA allows the physician to both *see* and be *seen* by the patient and therapist. The lower two figures show our current uni-directional proof-of-concept prototype. The user in (c) wears a tracking system and is imaged by a video camera. In (d) we show the avatar of the user, consisting of a styrofoam head mounted on a pan-tilt unit and illuminated by a projector.

ABSTRACT

Applications such as telepresence and training involve the display of real or synthetic humans to multiple viewers. When attempting to render the humans with conventional displays, non-verbal cues such as head pose, gaze direction, body posture, and facial expression are difficult to convey correctly to all viewers. In addition, a framed image of a human conveys only a limited physical sense of presence—primarily through the display's location. While progress continues on articulated robots that mimic humans, the focus has been on the motion and behavior of the robots.

We introduce a new approach for robotic avatars of real people: the use of cameras and projectors to capture and map the dynamic motion and appearance of a real person onto a humanoid animatronic model. We call these devices *animatronic Shader Lamps Avatars* (SLA). We present a proof-of-concept prototype comprised of a camera, a tracking system, a digital projector, and a life-sized styrofoam head mounted on a pan-tilt unit. The system captures imagery of a moving, talking user and maps the appearance and motion onto the animatronic SLA, delivering a dynamic, real-time representation of the user to multiple viewers.

Index Terms: H.4.3 [Information Systems Applications]: Communications Applications—Computer conferencing, teleconferencing, and videoconferencing H.5.1 [Multimedia Information Systems]: Animations—Artificial, augmented, and virtual realities I.3.7 [Computer Graphics]: Three Dimensional Graphics and Realism—Virtual Reality; I.3.8 [Computer Graphics]: Applications;

*e-mail: {plincoln, welch, nashel, adyilie, andrei, fuchs}@cs.unc.edu

1 INTRODUCTION

The term "telepresence" describes technologies that enable activities as diverse as remote manipulation, communication, and collaboration. Today it is a moniker embraced by companies building commercial video teleconferencing systems and by researchers exploring immersive collaboration between one or more participants at multiple sites. In a collaborative telepresence system, each user needs some way to perceive remote sites, and in turn be perceived by participants at those sites. In this paper we focus on the latter challenge—how a user is *seen* by remote participants, as opposed to how he or she *sees* the remote participants.

There are numerous approaches to visually simulating the presence of a remote person. The most common is to use 2D video imagery; however, such imagery lacks a number of spatial and perceptual cues. Even with 3D captured or rendered imagery and 3D or view-dependent displays, it is difficult to convey information such as body posture and gaze direction to multiple viewers. Such information can indicate the intended recipient of a statement, convey interest or attention (or lack thereof), and direct facial expressions and other non-verbal communication. To convey that information to specific individuals, each participant must see the remote person from his or her own viewpoint.

Providing distinct, view-dependent imagery of a person to multiple observers poses several challenges. One approach to is to provide distinct tracked and multiplexed views to each observer, such that the remote person appears in one common location. However, approaches involving head-worn displays or stereo glasses are usually unacceptable, given the importance of eye contact between all (local and remote) participants.

Another approach is to use *multi-view* displays. These displays can be realized with various technologies and approaches, however each has limitations that restrict its utility:

- "Personal" (per-individual) projectors and a retroreflective surface at the location corresponding to the remote user [15, 16]. Limitations: no stereo; each projector needs to remain physically very close to its observer.
- Wide-angle lenticular sheets placed over conventional displays to assign a subset of the display pixels to each observer [13, 20]. Limitations: difficult to separate distinct images; noticable blurring between views; no stereo—approach trades limited range of stereo for a wide range of individual views.
- High-speed projectors combined with spinning mirrors used to create 360-degree light field displays [11]. Advantages: lateral multiview with stereo. Limitations: small physical size due to spinning mechanism; binary/few colors due to dividing the imagery over 360 degrees; no appropriate image change as viewer moves head vertically or radially.

Our alternative approach is to use a human-shaped display surface that intrinsically provides depth cues. This one-to-many approach also scales to any number of observers, who do not need to be head-tracked. To convey appearance, we capture live video imagery of a person, warp the imagery and use Shader Lamps techniques [3, 18, 19] to project it onto the human-shaped display surface. As a result, all observers view the remote user from their own perspectives. To convey motion and orientation we track the user and use animatronics to vary the pose of the display surface accordingly, while continually projecting the appropriate imagery.

A fundamental limitation of this approach is that it does not result in a general-purpose display—it is a *person* display. More general multi-view displays [13, 10] can—and often are—used to display artifacts like coffee cups and pieces of paper along with the remote person. However, to use such displays for multi-viewer teleconferencing, one needs either many cameras (one per view) or real-time 3D reconstruction.

Figure 1 shows conceptual sketches and real results from our current proof-of-concept prototype. Our method and prototype are described in detail in Section 3. In Section 4 we present results, and in Section 5 we conclude with thoughts on the current state of our work and discuss future possibilities.

2 RELATED WORK

The related technical and scientific work in this area is vast. Some of the most visible results have been in theme park entertainment, which has been making use of projectively illuminated puppets for many years. The early concepts consisted of rigid statue-like devices with external film-based projection. Recent systems include animatronic devices with internal (rear) projection such as the animatronic Buzz Lightyear that greets guests as they enter the *Buzz Lightyear Space Ranger Spin* attraction in the Walt Disney World Magic Kingdom. While the presented method currently uses front projection, using internal projection would reduce the overall footprint of the robot, which would make it less intrusive and potentially more useful.

In the academic realm, *Shader lamps*, introduced by Raskar et al. [19], use projected imagery to illuminate physical objects, dynamically changing their appearance. The authors demonstrated changing surface characteristics such as texture and specular reflectance, as well as dynamic lighting conditions, simulating cast shadows that change with the time of day. The concept was extended to *dynamic shader lamps* [3], whose projected imagery can be interactively modified, allowing users to paint synthetic surface characteristics on physical objects.

Hypermask [25] is a system that dynamically synthesizes views of a talking, expressive character, based on voice and keypad input from an actor wearing a mask onto which the synthesized views are projected. While aimed at storytelling and theatrical performances, it deals with many of the issues we discuss here as well, such as the construction of 3D models of human heads and projecting dynamic face imagery onto a moving object (in this case, the mask).

Future versions of the technology we introduce here will require complex humanoid animatronics (robots) as "display carriers," which can be passive (projective, as shown here) or active (covered with flexible self-illuminated display surfaces such as the ones currently under development in research labs at Philips, Sony and others). Significant work in the area of humanoid robots is being conducted in research labs in Japan. In addition to the well-known Honda ASIMO robot [6], which looks like a fully suited and helmeted astronaut with child-like proportions, more recent work led by Shuuji Kajita at Japan's National Institute of Advanced Industrial Science and Technology [2] has demonstrated a robot with the proportions and weight of an adult female, capable of human-like gait and equipped with an expressive human-like face. Other researchers have focused on the subtle, continuous body movements that help portray lifelike appearance, on facial movement, on convincing speech delivery, and on response to touch. The work led by Hiroshi Ishiguro [9] at Osaka University's Intelligent Robotics Laboratory stands out, in particular the lifelike Repliee android series [5] and the Geminoid device. They are highly-detailed animatronic units equipped with numerous actuators and designed to appear as human-like as possible, also thanks to skin-embedded sensors that induce a realistic response to touch. The Geminoid is a replica of principal investigator Hiroshi Ishiguro himself, complete with facial skin folds, moving eyes, and implanted hair—yet still not at the level of detail of the "hyper-realistic" sculptures and life castings of (sculptor) John De Andrea [4], which induce a tremendous sense of presence despite their rigidity; Geminoid is teleoperated, and can thus take the PI's place in interactions with remote participants, much like the technology we advocate here. While each of the aforementioned robots take on the appearance of a single synthetic person, the Takanishi Laboratory's WD-2 [12] robot
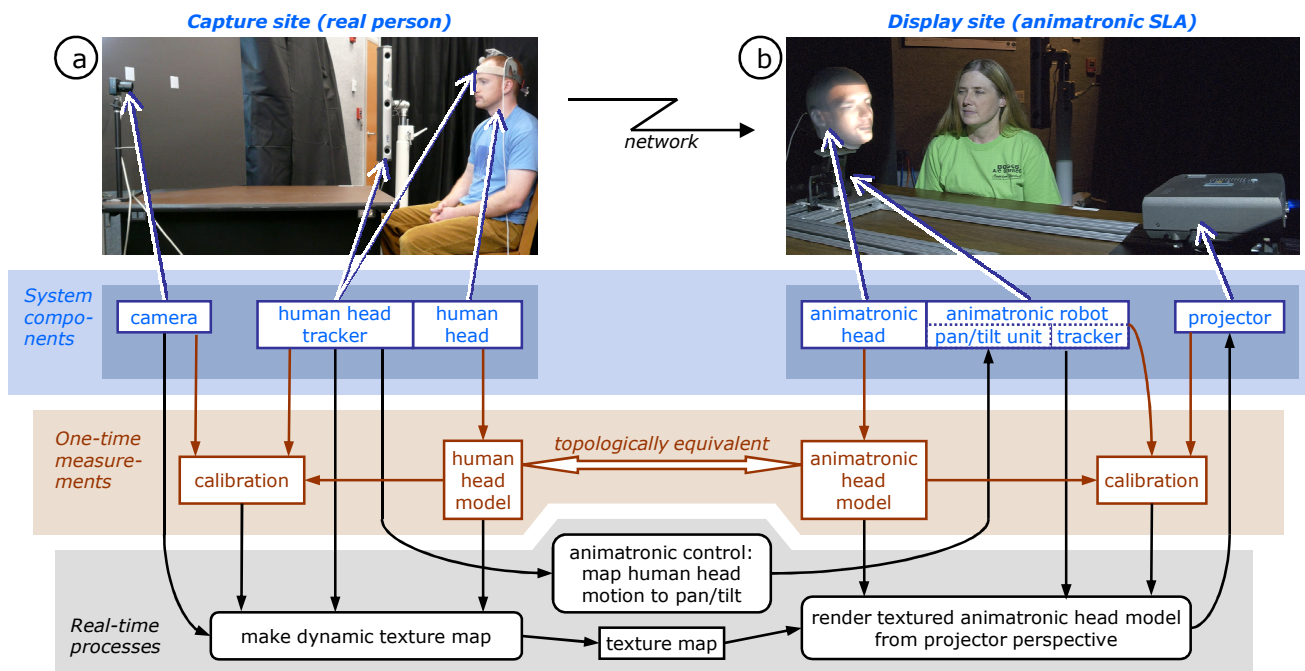
Figure 2: Proof-of-concept implementation and diagram. At the capture site shown in (a), a camera captures a person, also tracked using a headband. At the display site shown in (b), a projector displays images onto an avatar consisting of a styrofoam head placed on an animatronic robot. The diagram in the lower part of the figure highlights the system components and the processes involved.

is capable of changing shape in order to produce multiple expressions and identities. The WD-2 also uses rear-projection in order to texture a real user's face onto the robot's display surface. The researchers are also interested in behavioral issues and plan to investigate topics in human-geminoid interaction and sense of presence.

When building animatronic avatars, one is inevitably faced with the challenge of mapping human motion to the animatronic avatar's motion. The avatar's range of motion, as well as its acceleration and speed characteristics, will generally differ from a human's; with current state-of-the art in animatronics, they are a subset of human capabilities. Hence one has to "squeeze" the human motion into the avatar's available capabilities envelope, while striving to maintain the appearance and meaning of gestures and body language, as well as the overall perception of resemblance to the imaged person. In the case of our current prototype, we are only concerned with the mapping of head movements; previous work has addressed the issue of motion mapping ("retargeting") as applied to synthetic puppets. Shin et al. [22] describe on-line determination of the importance of measured motion, with the goal of deciding to what extent it should be mapped to the puppet. The authors use an inverse kinematics solver to calculate the retargeted motion. They also introduce filtering techniques for noisy input data (not an issue with our current tracker, but may become one with alternative, tetherless vision-based methods). Their work is geared towards complete figures, not just a single joint element as in our case, but their methods could be applied to our system as well.

The TELESAR 2 project led by Susumu Tachi [24, 23] integrates animatronic avatars with a display of a person. The researchers created a roughly humanoid robot equipped with remote manipulators as arms, and retro-reflective surfaces on face and torso, onto which projected imagery of the person "inhabiting" the robot can be shown. In contrast to the work we present here, the robot-mounted display surfaces do not mimic human face or body shapes; the three-dimensional appearance of the human is recreated through stereoscopic projection. The robot also contains cameras; it is con-

trolled by a human from a remote station equipped with multi-degree of freedom controls and monitors displaying imagery acquired by the robot's cameras. The work is part of an extensive project that aims to enable users to experience "telexistence" in any environment, including those that are not accessible to humans.

3 METHOD

In this section we describe our proof-of-concept system and present some details about the methods we employ. We begin by listing the system components and the relationships between them. Next, we describe one-time operations such as calibration and model construction. We continue with the adjustments performed before each run and finish by describing the real-time processes that take place during the use of the system.

3.1 System Components

The components of our proof-of-concept system, as shown in Figure 2, are grouped at two sites: the capture site and the display site. The capture site is where images and motions of a human subject are captured. In addition to a designated place for the human subject, it includes a camera and a tracker, with a tracker target (a headband) placed onto the human's head as shown in Figure 3 (a). We currently use a single 640x480 1/3" CCD color camera running at 15 FPS for capturing imagery. The focus, depth of field, and field of view of the camera has been optimized for being able to comfortably move around in a fixed position chair. The NDI Optotrak system is currently being used for tracking. Future systems may choose to employ computer-vision-based tracking, obviating the need for a separate tracker and allowing human motion to be captured without cumbersome tracker targets. The display site includes a projector, an avatar, and a tracker with a tracker target (a probe) mounted onto the avatar as shown in Figure 3 (b). The *avatar* consists of an animatronic head made of styrofoam that serves as the projection surface, mounted on a pan-tilt unit that allows moving the head to mimic the movements of the human at
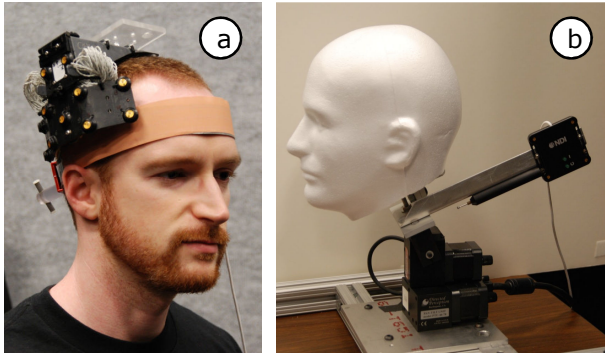
Figure 3: Tracker targets. (a) Headband tracker placed on a human head. (b) Tracker probe attached to the avatar, with pan-tilt unit in reference pose (zero pan and tilt).

the capture site. The 1024x768 60Hz DLP projector is mounted a few feet away from the avatar and is configured to only project upon the maximal range of the mounted avatar; the projector's focus and depth of field is sufficient to cover the illuminated half of the avatar. Instead of a tracker, future systems may choose to use position-reporting features of more sophisticated pan-tilt units to determine the pose of the styrofoam head.

3.2 One-time Operations

One-time operations are performed when the system components are installed. They include camera and projector calibration, as well as head model construction and calibration.

3.2.1 Camera and Projector Calibration

To calibrate the intrinsic and extrinsic parameters of the camera at the capture site, we use a custom application [8] built on top of the OpenCV [17] library. We capture multiple images of a physical checkerboard pattern placed at various positions and orientations inside the camera's field of view, and save them to disk. We automatically detect the 2D coordinates of the corners in each image using the OpenCV *cvFindChessboardCorners* function. Using the ordered lists of checkerboard corners for each image, we compute the intrinsic parameters via the OpenCV *cvCalibrateCamera*2 function. We then compute the extrinsic parameters in the tracker coordinate frame as follows. We first place the pattern in a single fixed position, capture an image of it and detect the 2D corners in the image as before. Next we use the tracker probe to capture the 3D locations corresponding to the pattern corners in the tracker's coordinate frame. Finally, we call the *cvFindExtrinsicCameraParams*2 OpenCV function using the captured 3D points, the corresponding 2D corner locations, and the previously computed intrinsic matrix; this produces the camera's extrinsic matrix in the coordinate frame of the capture-side's tracker. In the case of this system, these techniques are capable of projection a reprojection error on the order of a pixel or less.

We calibrate the projector at the display site using a similar process. Instead of capturing images of the checkerboard pattern, we place the physical checkerboard pattern at various positions and orientations inside the projector's field of view, and use our custom application to render and manually adjust the size and location of a virtual pattern until it matches the physical pattern. We save the rendered checkerboard images to disk and proceed using our custom OpenCV-based application and the tracker probe as outlined above in the camera case. Similarly, this produces the projector's intrinsic and extrinsic matrices in the coordinate frame of the display-side's tracker.

3.2.2 Head Model Construction

We built our 3D head models (human and animatronic) using Face-Worx [14], an application that takes in two images of a person's head (front and side view), allows manual identification of distinctive features such as eyes, nose and mouth, and produces a textured 3D model. The process consists of importing a front and a side picture of the head to be modeled and adjusting the position of a number of given control points overlaid on top of each image—see Figure 4 (a,e). The program provides real-time feedback by showing the resulting 3D model as shown in Figure 4 (b,f). A key property of all FaceWorx models is that they have the same topology, only the vertex positions differ. This allows a straightforward mapping from one head model to another. In particular, we can render the texture of a model onto the shape of another. In Figure 4, the projection-ready model (i) is obtained using the shape from the avatar head (h) and the texture from the human head (c).

3.2.3 Head Model Calibration

Capturing the human head model and rendering the animatronic head model "on top of" the styrofoam projection surface requires finding their poses in the coordinate frames of the trackers at each site. Both the human's and avatar's heads are assumed to have static shape, which simplifies the calibration process. The first step in this calibration is to find the relative pose of each head model with respect to a *reference coordinate frame* which corresponds to a physical tracker target rigidly attached to each head being modeled. We use a tracker probe to capture a number of 3D points corresponding to salient face features on each head, and compute the offsets between each captured 3D point and the 3D position of the reference coordinate frame. Next, we use a custom GUI to manually associate each computed offset to a corresponding 3D vertex in the FaceWorx model. We then run an optimization process to compute the 4×4 homogeneous transformation matrix that best characterizes (in terms of minimum error) the mapping between the 3D point offsets and the corresponding 3D vertices in the FaceWorx model. This transformation represents the relative pose and scale of the model with respect to the reference coordinate frame. We multiply it by the matrix that characterizes the pose of the reference coordinate frame in the tracker's coordinate frame to obtain the final transformation.

The calibration transformation matrices obtained through the optimization process are not constrained to be orthonormal. This can result from instabilities in the optimization computation, which can arbitrarily prefer undesirable homogenous transformations such as skew. As a final step on the calibration process, we perform manual adjustments of each degree of freedom in the matrices by moving the animatronic head or asking the human to move their head and using the movements of the corresponding rendered models as real-time feedback. The same error metric can be used in the manual adjustment phase in order to both reduce error while optimizing desirable transformations.

3.3 Per-run Calibrations

The headband used to track the human head is assumed to be rigidly mounted onto it. Unfortunately, each time the user puts the headband on his or her head, the position and orientation is slightly different. Although a complete calibration prior to each run would ensure the best results, in practice small manual adjustments are sufficient to satisfy this assumption.

Two adjustments are required for each run of the system. We first align the poses of the pan-tilt unit and of the human head as follows. We ask the human to rotate his or her head and look straight at the camera, and capture a *reference pose*. We set this pose to correspond to the zero pan and zero tilt pose of the pan-tilt unit—see Figure 3 (b), which positions the styrofoam head as if it were directly facing the projector. Finally, we perform additional manual
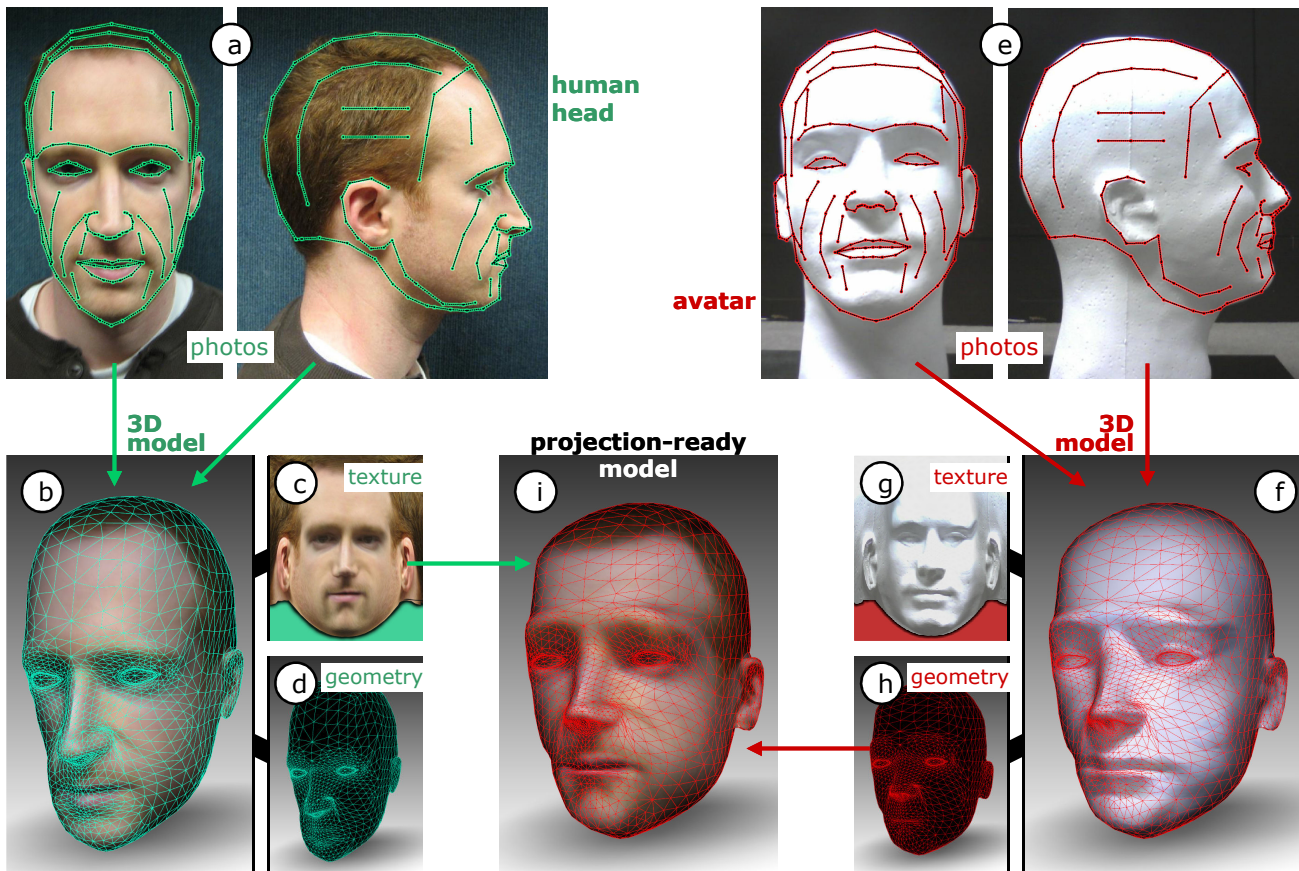
Figure 4: Head model construction and mapping. FaceWorx [14] is used to move control points in photographs showing the fronts and sides of heads (a,e), resulting in 3D models (b,f), which are comprised of texture (c,g) and geometry (d,h). The final model (i) is built using the texture of the human head (c) and the geometry of the avatar head (h).

adjustments to the headband by asking the user to rotate and shift the headband to ensure the projection of salient face features in the projected image are aligned with the corresponding features on the animatronic head; these features include the positions of the eyes, tip of the nose, and edges of the mouth. Essentially the shifting operations return the head band to it originally calibrated position on the human's head. When these features continue to correspond as the user move his or her head, the manual adjustments are complete.

3.4 Real-time Processes

Once the system is calibrated, it becomes possible for the avatar on the display side to mimic the appearance and motion of the person on the capture side. In this section we describe the real-time processes that implement this system operation.

3.4.1 Animatronic Control

Given a pose for a human head tracked in real time and a reference pose captured as described in Section 3.3, it is possible to compute a relative orientation. This orientation constitutes the basis for the animatronic control signals for the avatar. The pose gathered from the tracker is a 4×4 orthonormal matrix consisting of rotations and translations from the tracker's origin. We use the rotation component of the matrix to compute the roll, pitch, and yaw of the human head. The relative pitch and yaw of the tracked human are mapped to the pan and tilt capabilities of the pan-tilt unit and transformed into commands issued to the pan-tilt unit. Using this process, the avatar emulates the *motions* of its human "master."

3.4.2 Dynamic Texturing

Given a calibrated input camera, a tracked human, and a calibrated 3D model of the human's head, we compute a texture map for the model. This is achieved through *texture projection*; essentially, the imagery of the camera is projected upon the surface of the head model as though the camera were a digital projector and the human head the projection surface. In our system, we use OpenGL vertex and pixel shaders to achieve this process, which allows us to view a live textured model of the human head from any point of view.

In the case of the avatar, however, it is desirable to compute a texture map using the calibrated model of the human head and project the resulting live imagery onto the calibrated model of the avatar head. Since both heads are modeled in FaceWorx, they have the same topology. It is then possible to perform the warping operation shown in Figure 4 to transform the texture projection to target the avatar's head. We use an OpenGL vertex shader that takes as input the avatar's tracker, calibration, and model vertex positions to compute the output vertices. We use an OpenGL pixel shader that takes the human's tracker, the calibration model and the vertices computed by the vertex shader as input to compute the output texture coordinates. Through these shaders, it is possible to render a textured model of the avatar from any perspective, using a live texture from camera imagery of the human head. By selecting the perspective of the calibrated projector, the live texture is projected upon the tracked animatronic head, and the model shape is morphed to that of the animatronic head model. Using this process, the animatronic head emulates the *appearance* of its human counterpart.

Figure 5: Humans and avatars as seen from different viewpoints. Column 1 shows the live camera images; column 2 shows the warped head models; column 3 shows photos of the models projected onto the avatar; and column 4 shows the un-illuminated styrofoam head in poses matching the column 3 images. In row 1, the photos in columns 3 and 4 are taken from the left side of the projector; in row 2, these photos are taken from behind the projector.

4 RESULTS

The general result of the system is the presentation of a physical proxy for a live human. As implemented, the avatar can present elements of a user's facial appearance and head motion.

Visual appearance is generated through the use of (currently) a single camera and single projector and thus is limited to certain perspectives. In particular, high-quality imagery is limited to the front of the face. As in-person communication is generally face-to-face, it is reasonable to focus visual attention onto this component. Since the human's facial features are mapped to the avatar's corresponding features taking advantage of the identical topology of their 3D models, the avatar can present the human's eyes, nose, mouth, and ears in structurally appropriate positions. The quality of this matching is demonstrated in Figure 5. As both relationships (camera/human and projector/avatar) are approximately the same in terms of direction, the imagery is generally appropriate, and the features well matched. There is about a 0.3 second discrepancy between the camera and tracking system; a small amount of buffering allows us to synchronize the two data source. Thus, as the user moves, the tracker and camera imagery update correspondingly to project the proper texture on the virtual model of the head.

Using the pan-tilt unit, the avatar is also capable of movement that matches the yaw and pitch components of the human's head motion. As long as the human's orientation stays within the limits of the pan-tilt unit and tracker, the avatar can rotate to match the latest reported human head orientation. Because the human's features are texture-mapped to the corresponding locations of the, an observer can stand anywhere on the projected side of the head and see both a representation of the avatar's user and accurately gadge in which direction the user is looking. However, humans are capable of moving faster than the available pan-tilt unit; a human can move much faster than the pan-tilt unit's maximum rotational speed of 40 degrees/second. As a result, the avatar's head motion may lag behind the most recently reported camera imagery and corresponding tracker position. This issue could be mitigated with a faster and more responsive pan-tilt unit.

Fortunately, the capture and playback sides of the system can be decoupled; the motion of the avatar need not match that of the human user in order to show relevant imagery. Because the texture produced by the input camera is displayed on the avatar via projective texturing of an intermediate 3D model, the position and orientation of the avatar is independent of the human's position and orientation. The image directly projected on the avatar is dependent on the avatar's model and the current tracker position for the pan-tilt unit. Through this decoupling, the motion of the avatar can be disabled or overridden and the facial characteristics of the avatar will still match to the best degree possible. However, if the relative orientations between the human and camera and the avatar and projector are significantly different, the quality of the projective texture may be degraded due to missing information. For example, if the human looks to his or her right side, but the avatar is still looking straight at the projector, only the left side of the person's head can be captured by the camera and projected onto the avatar. This issue could resolved by using additional cameras and/or projectors to collect and provide better coverage.

5 CONCLUSIONS & FUTURE WORK

We introduced animatronic Shader Lamps Avatars (SLAs), described a proof-of-concept prototype system, and presented preliminary results. We are currently exploring passive vision-based methods for tracking the real person's head [1, 7, 21] so that we can eliminate the separate tracking system. We also hope to add, very soon, additional cameras and projectors. Both will involve the dynamic blending of imagery: as the real person moves, textures from multiple cameras will have to be dynamically blended and mapped onto the graphics model, and as the physical avatar moves, the projector imagery will have to be dynamically blended (intensity and perhaps color) as it is projected. We are also considering methods for internal projection. In terms of the robotics, we will be exploring possibilities for more sophisticated animation, and more rigorous motion retargeting methods [22] to address the limitations of the animatronic components (range and speed of motion, degrees
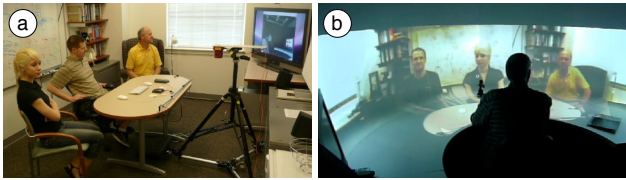
Figure 6: Remote panoramic video for Avatar control. A tripod-mounted PointGrey Ladybug camera is used to capture panoramic imagery of a remote scene in (a). The real-time video is mapped to a projector-based 270° surround display as shown in (b). The camera would eventually be mounted above the SLA.

of freedom) while still attempting human-like performance. Some of the filtering techniques in [22] could be useful if we use vision-based face tracking as above. Finally, together with collaborators at the Naval Postgraduate School we plan to undertake a series of human subject evaluations using our next generation prototype.

While our current prototype supports only half-duplex (one-way) communications, we envision full-duplex capability via the use of cameras associated with the SLA and a display associated with the user. For example, outward-looking cameras could be mounted in a canopy over the SLA to provide remote imagery for the user as depicted in Figure 1 (b) and (a) respectively. Figure 6 shows a preliminary demonstration of a panoramic camera and a surround display that could be used for viewing the Avatar's surroundings. Figure 6 also illustrates the one-to-many nature of the paradigm.

In the longer term, we have a vision for SLAs mounted on mobile platforms with outward-looking cameras that enable users to explore remote facilities such as hospitals, factories and shopping centers, while interacting with multiple remote individuals—both *seeing* and being *seen*. For some disabled individuals, this could provide a "prosthetic presence" that is otherwise unattainable. SLAs may also be useful as role players in immersive training environments for medicine and defense, robotic teachers that visually transform between historians and historic individuals, or personal robotic companions that take on different real or synthetic appearances during live interactions. In fact SLAs could some day support the limited integration of a virtual "second life" into our "first lives"—allowing people to visit remote real places, using a real or alternate persona, seeing and being seen as if they (or their persona) were really there.

ACKNOWLEDGEMENTS

REFERENCES

[1] J. Ahlberg and R. Forchheimer. Face tracking for model-based coding and face animation. *International Journal of Imaging Systems and Technology*, 13(1):8–22, 2003.

[2] AIST. Successful development of a robot with appearance and performance similar to humans. http://www.aist.go.jp/aist_e/latest_research/2009/20090513/20090513.html, May 2009.

[3] D. Bandyopadhyay, R. Raskar, and H. Fuchs. Dynamic shader lamps: Painting on real objects. In *Proc. IEEE and ACM international Symposium on Augmented Reality (ISAR '01)*, pages 207–216, New York, NY, USA, October 2001. IEEE Computer Society.

[4] J. L. DeAndrea. AskART. http://www.askart.com/askart/d/john_louis_de_andrea/john_louis_de_andrea.aspx, May 2009.

[5] R. Epstein. My date with a robot. *Scientific American Mind*, June/July:68–73, 2006.

[6] Honda Motor Co., Ltd. Honda Worldwide - ASIMO. http://world.honda.com/ASIMO/, May 2009.

[7] T. S. Huang and H. Tao. Visual face tracking and its application to 3d model-based video coding. In *Picture Coding Symposium*, pages 57–60, 2001.

[8] A. Ilie. Camera and projector calibrator. http://www.cs.unc.edu/~adyilie/Research/CameraCalibrator/, May 2009.

[9] H. Ishiguro. Intelligent Robotics Laboratory, Osaka University. http://www.is.sys.es.osaka-u.ac.jp/research/index.en.html, May 2009.

[10] A. Jones, M. Lang, G. Fyffe, X. Yu, J. Busch, I. McDowall, M. Bolas, and P. Debevec. Achieving eye contact in a one-to-many 3d video teleconferencing system. In *SIGGRAPH '09: ACM SIGGRAPH 2009 papers*, pages 1–8, New York, NY, USA, 2009. ACM.

[11] A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec. Rendering for an interactive 360° light field display. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, volume 26, pages 40–1 – 40–10, New York, NY, USA, 2007. ACM.

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

[13] P. Lincoln, A. Nashel, A. Ilie, H. Towles, G. Welch, and H. Fuchs. Multi-view lenticular display for group teleconferencing. *Immerscom*, 2009.

[14] LOOXIS GmbH. FaceWorx. http://www.looxis.com/en/k75.Downloads_Bits-and-Bytes-to-download.htm, February 2009.

[15] D. Nguyen and J. Canny. Multiview: spatially faithful group video conferencing. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 799–808, New York, NY, USA, 2005. ACM.

[16] D. T. Nguyen and J. Canny. Multiview: improving trust in group video conferencing through spatial faithfulness. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1465–1474, New York, NY, USA, 2007. ACM.

[17] OpenCV. The OpenCV library. http://sourceforge.net/projects/opencvlibrary/, May 2009.

[18] R. Raskar, G. Welch, and W.-C. Chen. Table-top spatially-augmented reality: Bringing physical models to life with projected imagery. In *IWAR '99: Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality*, page 64, Washington, DC, USA, 1999. IEEE Computer Society.

[19] R. Raskar, G. Welch, K.-L. Low, and D. Bandyopadhyay. Shader lamps: Animating real objects with image-based illumination. In *Eurographics Workshop on Rendering*, June 2001.

[20] O. Schreer, I. Feldmann, N. Atzpadin, P. Eisert, P. Kauff, and H. Belt. 3DPresence-A System Concept for Multi-User and Multi-Party Immersive 3D Videoconferencing. pages 1–8. CVMP 2008, Nov. 2008.

[21] Seeing Machines. faceAPI. http://www.seeingmachines.com/product/faceapi/, May 2009.

[22] H. J. Shin, J. Lee, S. Y. Shin, and M. Gleicher. Computer puppetry: An importance-based approach. *ACM Trans. Graph.*, 20(2):67–94, 2001.

[23] S. Tachi. http://projects.tachilab.org/telesar2/, May 2009.

[24] S. Tachi, N. Kawakami, M. Inami, and Y. Zaitsu. Mutual telexistence system using retro-reflective projection technology. *International Journal of Humanoid Robotics*, 1(1):45–64, 2004.

[25] T. Yotsukura, F. Nielsen, K. Binsted, S. Morishima, and C. S. Pinhanez. Hypermask: Talking head projected onto real object. *The Visual Computer*, 18(2):111–120, April 2002.