HeadSPIN: A One-to-many 3D Video Teleconferencing System

Andrew Jones[⋆] Magnus Lang[⋆] Graham Fyffe[⋆] Xueming Yu[⋆] Jay Busch[⋆] Ian McDowall[†] Mark Bolas[⋆‡] Paul Debevec[⋆]

[⋆] University of Southern California
Institute for Creative Technologies

[†] Fakespace Labs

[‡] University of Southern California
School of Cinematic Arts

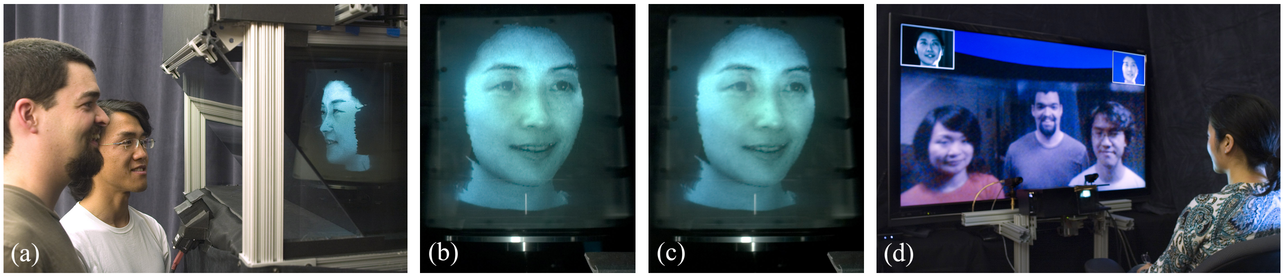

Figure 1: *(a) An audience interacts with a remote participant (RP) rendered in 3D on an autostereoscopic display. (b,c) A cross-fusable stereo pair where the RP appears life-size in correct perspective, able to make eye contact with the members of the audience. (d) The RP looks back at the audience via geometrically calibrated wide-angle 2D video while being scanned, transmitted, and rendered at 30Hz.*

When people communicate in person, numerous cues of attention, eye contact, and gaze direction provide important additional channels of information, making in-person meetings more efficient and effective than telephone conversations and 2D teleconferences. Two-dimensional video teleconferencing precludes the impression of accurate eye contact: when a participant looks into the camera, everyone seeing their video stream sees the participant looking toward them; when the participant looks away from the camera (for example, toward other participants in the meeting), no one sees the participant looking at them. In this work, we develop a one-to-many teleconferencing system which uses 3D acquisition, transmission, and display technologies to achieve accurate reproduction of gaze and eye contact. In this system, the face of a single remote participant is scanned at interactive rates using structured light while the participant watches a large 2D screen showing an angularly correct view of the audience. The scanned participant's geometry is then shown on the 3D display to the audience.

Real-time 3D Scanning The face of the remote participant is scanned at 30Hz using a structured light scanning system based on the phase-unwrapping technique of Zhang and Huang [2006]. The system uses a monochrome Point Grey Research *Grasshopper* camera and greyscale video projector running at a frame rate of 120Hz. Generally, we found 120Hz capture to be relatively robust to artifacts resulting from temporal misalignment, though fast facial motion can produce waviness in the recovered geometry.

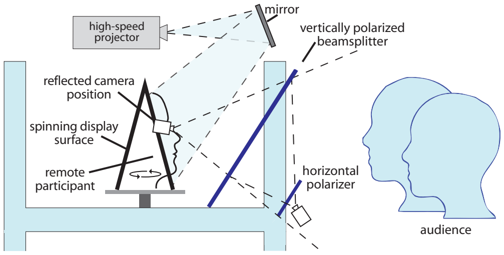

Figure 2: *Layout of the 3D Display apparatus*

Autostereoscopic 3D Display Our 3D display is similar to [Jones et al. 2007] with several key differences. First, the size, geometry, and material of the spinning display surface have been optimized for the display of a life-sized human face. The display surface has been replaced by a two-sided tent shape with symmetrical sides made from $20cm \times 25cm$ thin brushed aluminum sheet metal. We found that brushed aluminum had high reflectivity and anisotropy making it an inexpensive substitute for the holographic diffuser material used by [Jones et al. 2007]. The two-sided shape provides two passes of a display surface to each viewer per full rotation, achieving 30Hz visual update for 900 rpm rotation compared to the 15Hz update rate achieved by [Jones et al. 2007]. A monochrome high speed projector from Polaris Road, Inc. projects frames at 4,320 1-bit (black or white) frames per second using a specially coded DVI video signal. Effectively, the display projects seventy-three unique views of the scene across the 180-degree field of view, for an angular view separation of 2.5 degrees. For a typical inter-pupillary distance of 65mm, this provides binocular stereo for viewing positions up to 1.5m away.

2D Video Feed A 90° field of view 2D video feed allows the remote participant to view the central audience interacting with their three-dimensional image on the 3D display. A polarized beam-splitter as seen in Fig. 2 is used to virtually place the camera close to the position of the eyes of the three-dimensional head. Crossed polarizers on the camera and beam-splitter prevent the video feed camera from seeing past the beamsplitter while preserving the audience's reflection. The video from the aligned 3D display camera is transmitted to the the remote participant where it is shown on a geometrically calibrated projection screen. While the remote participants's view is not autostereoscopic, the screen is approximately at the typical distance of the audience members so that the visual disparity is approximately correct.

While the autostereoscopic horizontal-parallax-only nature of the display will generally produce accurate horizontal perspective, high or low vantage points may result in a less accurate vertical gaze direction. To correct the vertical perspective, we use marker-less face detection [Viola and Jones 2004] to track viewers based on the 2D video feed. In this way, the display's horizontal parallax provides binocular stereo with no lag as the viewers move their heads horizontally, while vertical parallax is achieved through tracking.

References

JONES, A., MCDOWALL, I., YAMADA, H., BOLAS, M., AND DEBEVEC, P. 2007. Rendering for an interactive 360 light field display. *ACM Transactions on Graphics 26*, 3 (July), 40:1–40:10.

VIOLA, P., AND JONES, M. J. 2004. Robust real-time face detection. *International Journal of Computer Vision 57*, 2, 137–154.

ZHANG, S., AND HUANG, P. 2006. High-resolution, real-time three-dimensional shape measurement. *Optical Engineering 45*, 12.

SIGGRAPH 2009, New Orleans, Louisiana, August 3–7, 2009.
ISBN 978-1-60558-726-4/09/0008