

TeleHuman2: A Cylindrical Light Field Teleconferencing System for Life-size 3D Human Telepresence

Dan Gotsch, Xujing Zhang, Timothy Merritt¹ and Roel Vertegaal

Human Media Lab
School of Computing
Queen's University
Kingston, Ontario, Canada

¹Dept. of Computer Science
Human Centered Computing
Aalborg University
Aalborg, Denmark

ABSTRACT

For telepresence to support the richness of multiparty conversations, it is important to convey motion parallax and stereoscopy without head-worn apparatus. TeleHuman2 is a “hologrammatic” telepresence system that conveys full-body 3D video of interlocutors using a human-sized cylindrical light field display. For rendering, the system uses an array of projectors mounted above the heads of participants in a ring around a retroreflective cylinder. Unique angular renditions are calculated from streaming depth video captured at the remote location. Projected images are retro-reflected into the eyes of local participants, at 1.3° intervals providing angular renditions simultaneously for left and right eyes of all onlookers, which conveys motion parallax and stereoscopy without head-worn apparatus or head tracking. Our technical evaluation of the angular accuracy of the system demonstrates that the error in judging the angle of a remote arrow object represented in TeleHuman2 is within 1 degree, and not significantly different from similar judgments of a collocated arrow object.

Author Keywords

Telepresence; cylindrical display; light field; hologram; 3D video; videoconference; motion parallax.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous;

INTRODUCTION

Current commercial video conferencing systems, like Skype, are still mostly limited to small displays. Some systems, like CISCO's TelePresence system, support life-sized images, but most conferencing solutions do not

convey images of remote interlocutors at sizes that are consistent with perceived social distance for natural discourse [1,7]. Most solutions are also limited to two-dimensional display. In multiparty scenarios, where there are multiple interlocutors conversing with multiple remote participants, 2D displays do not preserve the relative angles needed to be able to appropriately observe the body orientation and gaze direction cues necessary to support deixis and conversational turn taking [34]. To do so, systems should provide support for *motion parallax*, the ability to shift visual perspective from one interlocutor to another. Motion parallax can range from *discrete* viewports, one per interlocutor, to *continuous*, in which an interlocutor can freely move around to experience telepresence when exploring a remote scene, the latter still being a challenge with today's display systems [31]. Second, systems must support *stereoscopy*, the ability to view binocular images. Stereoscopy is important to be able to accurately judge angular depth information when at rest [11]. It is not supported by most commercial systems, and few research prototypes are capable of conveying stereoscopic cues without stereo glasses [6,11,27]. While both continuous motion parallax and stereoscopy are relatively easily supported in Augmented or Virtual Reality (AR/VR) systems, such as Microsoft's HoloLens [18], this comes at a cost of requiring a head-worn apparatus that obscures capturing of facial expressions and eye contact. Prior empirical work [4,16,34,35] has demonstrated the critical need to preserve such gaze directional and deictic body cues in support of multiparty conversational turn taking and deixis. We believe that for telepresence solutions to fully support the richness of multiparty conversations, it is necessary to capture and convey continuous motion parallax and stereoscopy at life size and within normal social proxemics, *without* requiring head-worn apparatus.

TeleHuman2: A Light Field Approach

In this paper, we present TeleHuman2, a life-sized hologrammatic telepresence system capable of conveying stereoscopy and continuous motion parallax around a cylindrical light field display without glasses or head tracking (*Note that we use the term hologrammatic to distinguish from a hologram, which is a light field generated by a laser interference pattern*). The system

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI 2018, April 21–26, 2018, Montréal, QC, Canada.
© 2018 ACM ISBN 978-1-4503-5620-6/18/04...\$15.00.
<https://doi.org/10.1145/3173574.3174096>

captures the remote interlocutor through a circular array of stereoscopic cameras. Their images are used to generate a relief map, which is sent over Ethernet to a network of System-on-Chips (SoCs) that each render a unique viewport, one for every 1.3 degrees of angle around the interlocutor. Each view is projected onto a retroreflective cylindrical surface through a large circular array of projectors, one per 1.3 degrees. Unique views are retro-reflected back into the left and right eyes of any number of participants standing underneath the projection ring, rendering 360 degrees of motion parallax when 275 projectors are slotted into the projector ring. Due to budget constraints, and as a proof of concept, our current prototype implements 45 of these, providing a total parallax of 59 degrees around the remote human, sufficient for a three person conversation between two local interlocutors and one remote person.

Contributions

Our system is, to our knowledge, 1) the first light field teleconferencing system to support life size stereoscopy and continuous motion parallax *around* a remote human, without requiring glasses or head tracking. 2) It allows *many* local participants, each with their own spatially accurate stereoscopic point of view. 3) Its high 720p resolution image is sufficiently bright even in full sunlight. 4) The system captures 3D surround video of the remote human using 2K stereo cameras in broad daylight, providing superior image quality without the infrared interference issues that are typical when using Kinect sensors [11]. 5) We contribute a technical evaluation of angular accuracy of the system that demonstrates that judging the angle of an arrow represented in TeleHuman2 is not significantly different in accuracy from judging the real object.

BACKGROUND

First, we discuss the background of light field display technologies. We then provide an overview of prior work on multiparty video conferencing solutions, limiting our review to those systems that supported multiple simultaneous viewports to provide varying levels of motion parallax and/or stereoscopy. We conclude with a brief discussion of empirical work.

Approaches to Light Fields

Light field displays visualize 3D scenes by recreating the field of light around a scene – both spatial and angular light information. Technologies for generating light fields include lenslet arrays [13], parallax barriers [10] and multiview projection [21]. Lenslet displays work by emitting light from clusters of pixels into corresponding angles using a large array of small lenses. Hirsch et al. [9] used a lenslet display to both capture *and* display 3D content. Hirsch et al. [8] implemented a display that reacts to incident light. Users can use this for relighting virtual objects. The disadvantage of lenslets is that they sacrifice spatial resolution for angular resolution. Parallax barriers block light such that some pixels are only visible to the left

eye, and some to the right eye [15]. The disadvantage of this technology is that images tend to be low resolution *and* low in intensity. Random hole displays improve on parallax barriers by avoiding repeating patterns [22]. Multiview projection systems can be back-projected with lenslet arrays [32], or front projected onto retroreflective materials that reflect light from differently angled projectors into the eyes of an onlooker [21]. One of the key benefits of the latter technology is that it preserves resolution by multiplexing the number of displayed images rather than re-distributing pixels. Next, we provide a more detailed discussion of the use of this technology in teleconferencing.

Multiview Teleconferencing Systems

Over the years, there has been steady incremental progress from discrete viewports towards the use of full projected light field displays in multiparty teleconferencing systems. Hydra [29] was one of the first systems to support multiple viewports through the use of multiple display/camera units. Eye contact could be supported by having the camera mounted close to the displays, above the eyes of the participant, and while neither continuous motion parallax nor stereoscopy were supported, participants could observe different sides of other participants' faces.

Twister [30] allowed an autostereoscopic view of a remote participant, with the local person standing inside a hollow tube. It, however, did not support multiple simultaneous views. One of the first systems to use multi-projection and retroreflective surfaces – an early but discrete form of light field – was Nguyen et al.'s MultiView [23]. It used three projectors placed in front of 3 participants, alongside 3 cameras to render a maximum of 3 spatially correct discrete viewports in a room-sized video conference. MultiView used a narrow-angle retroreflective projection surface, which bounces all light back in the direction of the projector. A vertical diffuser layer allows the projector to be vertically offset from the eyes of the users without a significant drop in brightness. The result is a drop in brightness of the projected image only along the horizontal dimension and only when participants are not within a few degrees of the projector axis. This ensures a different projected image is seen by each of the three participants. MultiView's projector array was not sufficiently dense to support stereoscopy.

Baker and Li [2] added cameras and projectors to increase the number of discrete viewports and provide limited stereoscopy. Their system involved 9 cameras and 9 pico projectors projecting 7 independent binocular viewports on a retroreflective screen. Our system improves on this by providing continuous motion parallax, and by using a cylindrical projection surface, which allows users to walk around the display. While Bolton et al.'s TeleHuman [11] did not support multiple simultaneous viewports, its cylindrical human-sized display provided continuous near-360 degree motion parallax through head tracking, as well as stereoscopy through shutter glasses. Maimone & Fuchs

[14] introduced the concept of using depth camera arrays for capturing real-time 3D surround video. The use of 3D surround video, in which a full 360 degree 3D model is generated every video frame, has the advantage of allowing a virtual camera to be moved freely around the live video. While Maimone & Fuchs [14] rendered 3D surround video on flat displays, TeleHuman rendered different perspectives 360 degrees around a cylindrical display. Pan et al. presented a hemi-circular video capture array with 11 PlayStation Eye cameras. Its 11 discrete viewports were projected on a spherical MagicPlanet display [25]. However, this display did not support multiple simultaneous viewports. Nagano et al. [21] discuss a flat anisotropic retroreflective surface with a vertical Luminit diffuser projected on by an array of 72 TI Pico projectors. However, the image was limited to a static 3D model, and no 3D capture system was present. Pan et al [27] introduced a cylindrical retroreflective display with a vertical diffuser and 9 projectors allowing the capture and presentation of 9 independent viewports.

Recently, the USC Shoah Foundation [33] presented a recorded light field that used cameras in a ball-like configuration. Their set up, however, was not used for teleconferencing as it involves recording with many high definition cameras in a dome shaped green room. Our approach differs from the above in three ways: 1) rather than using N capture cameras relaying video to N projectors, we employ depth cameras with perspective shifts, requiring only 1 camera per 15 projectors or 20 degrees of angle. 2) Our approach supports glasses-free stereoscopy and motion parallax in natural settings. 3) Finally, our system uses the TeleHuman [11] form factor, is human-sized, with neither projectors nor cameras obstructing interactions at social distance with the virtual human [1]. While our system is not currently bi-directional, bi-directionality is possible by mounting the depth cameras on the cylindrical display and projector ring rather than in a separate capture room.

Virtual and Augmented Reality Systems

Some shared virtual and augmented reality environments allowed the rendering of multiple viewports of 3D surround video of humans. Microsoft's Holoportation system captures 3D surround video of users using an array of depth cameras [18]. These live video models are then rendered onto a HoloLens [17] headset, allowing full motion parallax and stereoscopy. The drawback of these technologies is that they require users to wear head mounted displays that obscure their faces. This makes it difficult to capture surround video of participants during a teleconference. Some CAVE-like environments circumvented this by only requiring headsets for stereoscopy, rather than imaging, however, these are typically limited to providing a single user with a single (stereo) view. blue-c [6] combined 3D surround video capture and rendering from 8 camera views in a CAVE. Head tracking was used to generate a unique viewport around the surround 3D video model. Hao Li et al.

[12] tried to resolve the issue of headsets obscuring video capture of the face by animating a 3D model of the face using data obtained from strain gauges on the VR helmet. While this approach appears promising, it is susceptible to the Uncanny Valley effect, as users are very sensitive to unnatural motion in human faces [20].

Empirical Work

We limit this review to the body of work studying the accuracy of gaze and/or pointing angles using multiview and non-flat motion-parallax teleconferencing displays. Nguyen et al. [23] studied the effect of seating position on the perceived angle of gaze of an on-screen participant. Results suggested viewer angle relative to the screen did not affect accuracy significantly. They repeated this experiment for pointing gestures with a similar result. Bolton et al. [11] studied the independent effect of 2D display, motion parallax display and motion parallax plus stereoscopic display on the accuracy of estimating where a remote participant pointed or looked. Results suggested significant effects on accuracy of both continuous motion parallax and stereoscopy. Pan et al. [26] studied the effect of a flat multiview autostereoscopic random hole display on head gaze estimation errors. The display allowed for only 3 simultaneous viewports, but these were stereoscopic. As with [4], results suggested that the presence of motion parallax and stereoscopy significantly improved the accuracy with which observers were able to assess gaze direction. Pan et al. [24] studied the use of a spherical teleconferencing display with surround video capture on target angle perception. They asked participants to determine which target a recorded actor was looking at, comparing a variety of flat displays with a spherical display and a face-to-face condition. The spherical display was not multiview and not stereoscopic: instead the view shifted around the display based on the angular position of the participant. Findings were encouraging in that after face-to-face, the spherical display achieved the lowest mean error. Pan et al. [27] also measured the level of error in their cylindrical multiview telepresence system during an angle estimation task. They used 12 groups of 4 participants seated at different locations around the display, judging at which of 15 targets a previously recorded actor was looking. Results suggested that the absolute error between the observers' perceived target and the actual target angle was lowest when using their cylindrical multiview display.

DESIGN CONSIDERATIONS

Our overall goal in the design of TeleHuman2 was a system with a flexible and very dense number of viewports. When viewports are sufficiently dense, multiple interlocutors can perceive the remote participant stereoscopically, with continuous motion parallax, without any need for glasses or head tracking. The following design criteria were taken into consideration:

Form Factor – Providing the option for 360° motion parallax requires the use of a cylindrical display, preferably

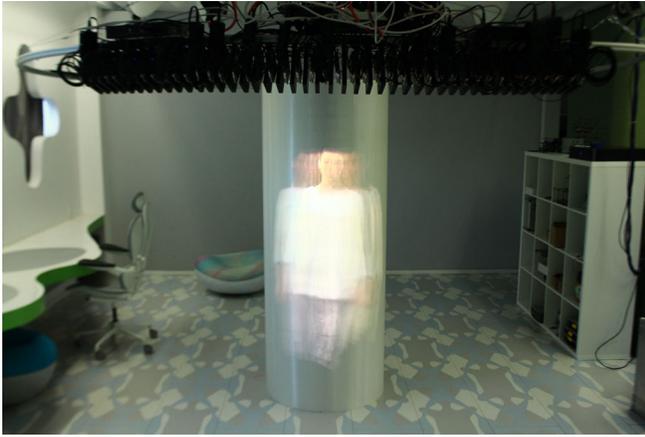


Figure 1. TeleHuman2 with projection ring and cylindrical retroreflective projection surface displaying a remote user.

proportionate to the human body. This enables users to explore different perspectives by simply walking around the display.

Glasses-free – We designed the system such that no stereo glasses of any kind would obstruct live capture of the face.

Many Viewports – We designed the system such that it would support multiple interlocutors in a conversation with one remote participant, with appropriate parallax and stereoscopy. This required *many* more viewports than have been demonstrated in the literature.

Resolution – Lenslet array and parallax barrier solutions repurpose display resolution into unique viewports. This reduces the spatial resolution of the display considerably. We opted for a retroreflective solution with an array of pico projectors providing full 720p resolution per viewport.

Brightness – Parallax barrier solutions to light fields also reduce the brightness of display considerably. This is also, to a lesser extent, true for lenslet arrays. Although pico projectors, at 32 lumens, are dim when projected on a typical diffuser, when using a focused retroreflector, all the light is reflected back to the eyes of the observer in a narrow band of approximately 1.3 degrees. The result is a very bright display that operates in full daylight.

Angular Resolution – Our criterion for angular resolution was the average interpupillary distance, with a mean of approximately 6.3 cm between pupils for adults [5]. We constructed a ring of projectors such that the distance between any 2 projectors was less than 6 cm at a comfortable social distance (Hall's *Social Space* 1.2m – 3.6m) [7].

Stereoscopy and Horizontal Motion Parallax – Both motion parallax and stereoscopy would be facilitated by a 1.3 degrees drop-off in brightness of a retroreflector, allowing each eye to only see one projected image at a time, corresponding to the projector that is closest to that eye. When the left eye is below one projector, and the right eye



Figure 2. Projector/Odriod array mounted on the ring.

below a second projector, images are perceived stereoscopically.

Vertical Parallax – Since the retroreflector projects light back to the projector, the light needs to be diffused vertically in order for there to be clearance between the location of the projector and the location of the user's head. This means it is not possible to provide motion parallax in the vertical dimension. We did not believe this to be an issue since most human movement during conversations is horizontal. However, we did make provisions for allowing the entire ring of projectors to be moved up and down.

Surround 3D Video Capture – Rather than using one camera for each projector, we decided to use depth cameras to generate partial 3D surround video images. This also allowed for adjustments to the vertical parallax by means of virtual camera adjustments, allowing for eye contact simulation. An additional benefit of using 3D surround video is that only a small number of cameras is needed, reducing the need to obstruct the display in symmetrical conditions. We decided to use a visible light stereo camera rather than a Kinect for superior image quality.

Freedom of Movement – Participants should be able to move within Hall's Social Distance (1.2m-3.6m) comfortably without losing stereoscopy or motion parallax. This is ensured by mounting the projectors in a circular array such that projectors need not be consecutive to accurately represent angles for the left and right eye. Whichever projector the eye happens to be aligned with provides the accurate angular representation.

Parallel Rendering – By providing each projector with its own processor, we were able to render as many viewports as are needed, in parallel, without time delay. Combining arrays of projectors with one stereo camera allowed simple shifts in the horizontal perspective to be calculated most efficiently using a system on a chip (SoC) via relief maps.

Asymmetrical Telepresence – While it is possible to design a symmetrical solution that allows simultaneous two-way capture and display, the current system was limited to providing only one directional telepresence for budgetary reasons. Bidirectional telepresence with our system however, only requires mounting stereo capture cameras on and around the cylindrical display, rather than in a separate remote capture room.

IMPLEMENTATION

Our implementation focused on the development of a light field cylindrical display using a multi-projector array. We will first discuss the implementation of the cylindrical display, after which we discuss the projector array, rendering subsystem, the capture subsystem and surround audio.

Cylindrical Light field Display

Figure 1 shows the cylindrical display deployed in TeleHuman (*Note that the rolling shutter of the camera introduces distortions in this image not visible to the user*). The display consists of a 195 cm tall hollow tube with a diameter of 75 cm made of 5 mm thick acrylic. The acrylic is coated with a first layer of retroreflective sheeting. This material has a very narrow retroreflective viewing angle of approximately 1.3 degrees vertically and horizontally, reflecting light back to the projector such that the brightness of the image drops off below 50% beyond 1.3 degrees from the projector. This is important to ensure stereo pair separation between projectors in the system and allows only eyes within 1.3 degrees of separation from the projector to see an image from that projector. Since we did not design for vertical parallax, and to allow the projector array to be raised above the heads of users, we applied a second sheet of a Brightview Technologies one-dimensional diffuser. By diffusing the light vertically, this layer helps maintain brightness of the reflected image when the projector is mounted above eye level.

Projector Array

Suspended over the cylinder is a circular rail with slots designed to hold a total of 275 projectors (see Figure 2). We used PicoPro laser pico projectors [3] because they do not require focusing, are small, thin, low cost, and provide a high 720p resolution. While they are only rated at 32 Lumens, because much of the light is directly reflected back at the user, the image appears bright even in daylight. A 3D printed mount connects to each pico projector, allowing it to slide into the circular rail's profile. Laser cut spacers

between projectors ensure a uniform distance of 4.5 cm between all projectors, which is less than the average interocular distance [5]. This allows any two consecutive projectors to function as a stereo pair. When a user stands directly below the projector rail the system has an angular resolution of 1.3 degrees. The rail has an inside radius of 180 cm and is suspended aligned with the top of the cylinder, at 195 cm height, providing about 18 cm clearance above the top of the average male's head. Projectors are powered over USB via a 60A 5V power supply per cluster of 15 projectors (covering ~20 degrees). We implemented 3 clusters of 15 projectors in this version, providing a total of 45 viewports with approximately 59 degrees of motion parallax around the cylinder. Adding more parallax only requires the addition of more projection clusters.

Render Array Hardware

Connected to each pico projector is an Odroid C1+ board that functions as its rendering engine (see Figure 2). As such, the projection rail functions as a parallel computer that can simultaneously render the subject from any angle. Each cluster of 15 Odroid boards is connected to one Gigabit Ethernet switch. This allows each Odroid to receive colour images and depth maps over UDP from the remote capture system. Note that we chose UDP for its efficiency as it requires each datagram to only be broadcast once. Each Odroid is connected to its pico projector using an HDMI Type D connector and powered from the cluster's central power supply over 5V USB cable. Average rendering speed is in the order of 10 frames per second (FPS).

Remote Capture System Hardware

Figure 3 shows a user in the remote capture room. The remote participant stands in a circle with a radius of 180 cm with 1 ZED camera per 20 degrees (see Figure 5). The ZED camera is a 2x2K stereo camera capable of generating a depth map and texture map with much higher resolution than the Kinect 2.0, in broad daylight, without interference when multiple cameras are used. Each ZED camera is connected to an Intel i5 computer running Ubuntu 16.04, with one high-end NVidia GTX1080 graphics card per ZED cam. This PC runs the ZED camera SDK which turns the stereo image into a depth map and RGB image. The images of each ZED cam are sent to the corresponding cluster of 15 Odroids at the receiving end, with the camera location corresponding to approximately the central projector in each cluster. The other projectors in the cluster render either negatively or positively rotated projections of this image, with 1.3 degree increments per projector. We implemented a capture system with 3 ZED cameras, providing 59 degrees of motion parallax video imaging around the participant. Increasing this parallax only requires adding more ZED cameras and graphics cards.



Figure 3. Remote capture room with ZED cameras and displays

Capture and Rendering Software

The ZED cameras capture stereo 2K images at 15 frames per second. These are used for depth calculations in CUDA 8.0 running on the NVidia GTX1080, via the ZED SDK. The background of the images is subsequently removed using an image subtraction algorithm, after which images are scaled to 1280 x 720 pixel resolution. The images are then compressed as JPEGs and sent through a UDP broadcast to the corresponding cluster of Odroids. Since the maximum size of a UDP packet is 64 Kilobytes, colour and depth images are conditionally split into multiple packets when oversized. On the Odroids, UDP packets are collected until both colour and depth data of each frame is received. Frames with missing packets are discarded. The colour and depth data are applied to an OpenGL rectangle which is rendered in 3D using a relief mapping shader, in a method similar to Policarpo et al. [28]. This rotates the video based on the angular distance from the center of the projector cluster (and thus the center of the ZED camera) to the associated projector, up to $\pm 10^\circ$. Images were also corrected for projection onto a curved surface.

Camera and Projector Calibration

The ZED cameras' position and orientation in space was measured using OpenCV's camera calibration module. As a calibration image, we used a 7 x 11 pattern of an asymmetric grid of black circles on a white background with 5 cm spacing. To further calibrate each camera's position, we incorporated the lens functions of the cameras as provided by the manufacturer. Calibrations were used to arrange the ZED cameras concentrically and were sent over to the projection ring for use in the Odroid rendering algorithm.

Projector Calibration

To align the large array of projectors, we designed an automated software calibration system based on [19], with modifications to allow for multiple projectors. To calibrate the projector array, we first placed a planar chessboard

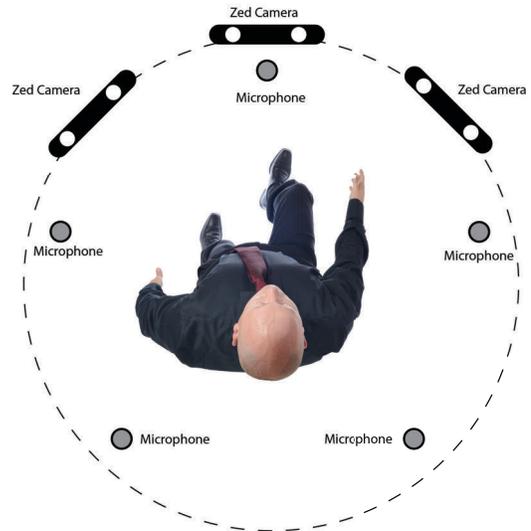


Figure 4. Diagram of remote capture system. Coverage is increased by adding 1 ZED camera every 20 degrees.

pattern marker one meter above the floor in the center of the projection ring. A webcam was placed above the marker looking directly at the chessboard pattern. From this perspective, an image is captured of the pattern with all projectors fully illuminated. A mapping between the real world coordinates of the chessboard corners and the webcam image coordinates was determined using OpenCV's findChessboardCorners() method. For each projector, images were captured as the projector displays a structured light sequence, where each pixel presents its projector image space coordinates, one bit at a time, over a sequence of images. As images were captured, their pixel values were compared against the global minimum and maximum intensities for that coordinate across all images. Once the entire coded light sequence had been captured, the images were thresholded using these global maxima and minima values and decoded into an image that mapped the projector coordinate space to that of the camera. A median sample of the decoded light maps was then taken for each chessboard corner location, determined in the first step. These samples provided a mapping between the known real-world coordinates of the chessboard and the projector image coordinates. Each of these mappings was then used to estimate the pose of the associated projector by triangulation using OpenCV's solvePNP method. The resulting translation and orientation matrices were applied to transform the virtual viewport used to render content for the associated projector.

Remote Participant Display

In the current implementation, the remote participant sees a 180 degree cropping of a 360° video capture of the local participants performed by a Ricoh Theta S camera mounted on top of the cylindrical display. The Theta 360 video stream is sent over the network to the remote capture room, where select parts of the stream are rendered on two 27" iMac displays (see Figure 3).

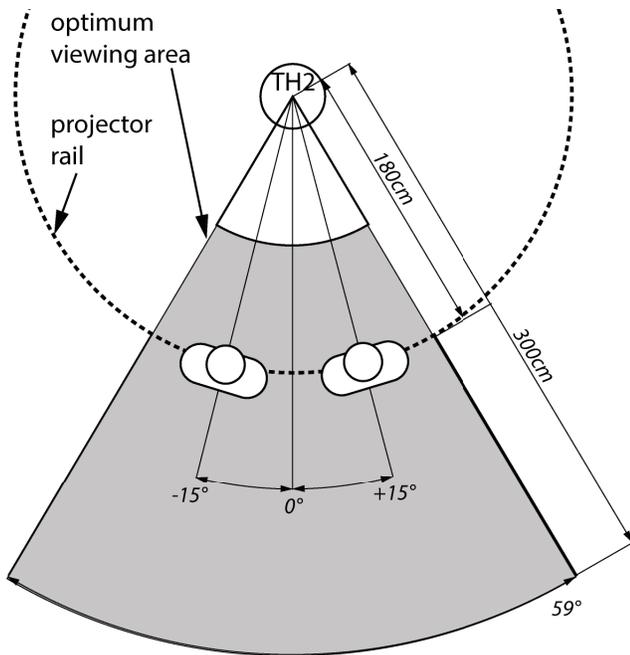


Figure 5. Overview of the experimental setup.

Audio Subsystem

A circular array of 5 speakers surrounding the projector ring provides surround sound to TeleHuman2. Sound is recorded through a ring of microphones surrounding the remote participant (see Figure 4). This means audio is directional, and when the remote participant turns to speak to another person, the sound appears from that direction. In our lab setup, audio is sent via MOTU IEEE 802.1 Audio Video Bridging (AVB) interfaces over Cat-6 Ethernet to a NAD T775 Dolby 7.1 amplifier that outputs to the surround speakers mounted in the ceiling above the TeleHuman2. Surround sound can, however, also be delivered over the Internet via multiple dedicated VOIP channels.

INITIAL STUDY: ANGULAR ACCURACY

We evaluated initial user experiences, focusing on establishing the *technical* accuracy of the system with regards to its ability to convey angular information.

Tasks and Questionnaire

In each session, we asked 2 participants to perform a collaborative pointing task while placed directly underneath the projection units, at 1.8 m from the center of the cylinder. The left participant was placed at -15 degrees from the center of the projector cluster, the right participant at +15 degrees from the center of the projection cluster (see Figure 5). Each participant was provided with a Bluetooth remote that controlled the angle of a motorized *physical arrow* placed remotely in the capture room (see Figure 6). In one condition, this physical arrow was rendered as a light field on the TeleHuman2. In the other condition, it was physically present in the room, at the location of the TeleHuman2 cylinder. In each condition, the left participant was asked to align the arrow such that it pointed at him or



Figure 6. Physical tripod with cylindrical arrow and servo.

herself. Next, the left participant was asked to point the arrow at the right participant. This task was repeated for the right participant, who was asked to point the arrow at herself and at the left participant. Participants were allowed to move their heads during this task, but not their body.

Questionnaire

After this task, and while standing underneath the projector array, a live remote human was projected on the cylinder. Participants were asked to rank their stereoscopic perception of the remote human by indicating agreement using a Likert scale of 1 (strongly disagree) to 5 (strongly agree) on the question “*I am experiencing an accurate three dimensional image of a person*”. Next, participants were asked to walk around the 59° projection arc, after which they were asked to rank their agreement with the following statement using a Likert scale of 1 (strongly disagree) to 5 (strongly agree): “*I am experiencing an accurate rendition of motion around a person*”. This procedure was repeated at 3m away from the center of the cylinder. Participants were also asked for general comments on the system.

Factors

We used two factors, each with two levels: 1) Physical presence of the arrow; and 2) pointing at self vs. pointing at the other person.

Apparatus and Measurements

The physical arrow was made of a poster tube, 43 cm in length and 10 cm in diameter (see Figure 6). The arrow was mounted on an axle placed on a servo controlled by an Arduino. This assembly was placed on a tripod at 1 m from the floor in both the remote and local room. In the remote

	Error pointing at self (°)	Error pointing at other (°)
Real	0.33 (0.51)	0.33 (0.62)
Light Field	-0.33 (0.50)	0.67 (1.58)

Table 1. Mean error (std. err) in degrees of arc pointing the arrow at self and the other participant, for both real and light field conditions.

condition, the arrow was captured in 3D using 3 ZED cameras to produce 3 3D relief maps, one for each 15-projector cluster. Users used the buttons on an Apple Bluetooth mouse as a remote to rotate the arrow left or right during the alignment task. Measurements of the physical arrow's rotation were performed by logging the angles of the servo via the Arduino serial port.

Participants

We used 12 participants, 4 of which were female, with a mean age of 24.3 years.

Experimental Design

We used a within-subjects design in which all participants experienced all conditions. For counterbalancing, the order of left and right participant presentation of tasks was alternated at every trial, as was the order of presentation of conditions.

RESULTS

Table 1 shows the resulting angular error in degrees for our two conditions. We performed a two-way ANOVA to simultaneously test differences between all conditions. Results show that the mean error in pointing the arrow was less than 1° in all cases. There were no significant differences between real and light field (virtual) conditions ($F(1,44)=0.033$, *n.s.*). There were also no significant differences between pointing at self and pointing at another person ($F(1,44)=0.294$, *n.s.*). Finally, there was no significant interaction effect.

Questionnaire Data

Mean rankings of stereoscopy and motion parallax are presented in Table 2. A Wilcoxon signed-rank test showed differences between rankings for stereoscopy between 1.8 m and 3 m distance were not significant ($Z=-0.45$, *n.s.*). Differences in rankings for motion parallax, however, were significantly better at 1.8 m distance ($Z=-2.53$, $p<0.05$).

DISCUSSION

We now discuss results from our initial user observations, in light of limitations of the current design.

Angular Accuracy

Results show that participants were able to point the arrow at themselves very accurately, to within 1 degrees of arc, in

	"Accurate 3D Image"	"Accurate Rendition of Motion"
1.8 Meters	3.8	4.4
3 Meters	3.7	3.8

Table 2. Mean scores, out of 5, of agreement with the statements "I am experiencing an accurate 3D image of a person" and "I am experiencing an accurate rendition of motion around a person".

both real and light field conditions. Note that this is better than the actual angular resolution of the system, due to stereoscopy and averaging over subjects. This means the total angular error of the light field capture and projection system is at least within this margin, minus any error attributable to the human perception component. Update rate and blur may have affected these results, which means they qualify as a lower bounds.

Stereoscopy

TeleHuman2 appears to provide acceptable stereoscopy for multiple simultaneous onlookers: Scores for stereoscopy rankings equalled 4 in 9 out of 12 observations at 1.8 m distance. There also appears to be tolerance with regards to stereo perception within normal social interaction distances. This is because at 3 meters, the reflected image space is still approximately 6 cm apart, thus maintaining stereoscopy. This is why projectors were placed at 4.5 cm distance, closer than the interpupillary distance, on the ring. Note that projector alignment and calibration is critical for achieving stereoscopy: small aberrations in angle and throw distance create large and observable shifts in the left/right image.

Motion Parallax

Mean scores for motion parallax were high, particularly at 1.8 m distance. This suggests participants experienced continuous motion parallax throughout the 59 degrees of movement along the projector ring. Unlike stereoscopy, motion parallax was experienced as less the further the participant was away from the cylinder. Participants also reported blurring on the extreme angles of motion parallax. This was likely due to the use of relief maps for rendering the graphics on the SoCs: At extreme angles artifacts of the relief mapping become visible.

Vertical Brightness Drop-off and Video Quality

At 10 FPS, all participants deemed the system's refresh rate too low. This is caused by the use of underpowered SoCs for rendering. We established the latency between capture and projection at approximately 200 msec. Some participants commented they observed a ghosting effect. This was caused by imperfect retro-reflection causing crosstalk between projectors and is visible in Figure 1. The bleed from projector to projector can be removed with improvements in the retroreflective material, or by using

face-tracking to turn off projectors that are not seen by the user(s).

Projector Array

Participants appeared positive about the hologrammatic effect provided by the system. Further work may be necessary to improve upon framerate and the capture system. One limitation of the present prototype is that motion parallax is restricted to 59 degrees. Future versions would need to increase the number of projectors to 275 to achieve 360 degrees of motion parallax around the ring. Frame rate is presently limited by the computing power of the Odroids. With the current parallel computing design framerates can be improved by introducing more powerful SoCs when they become available. Alternatively, each projector could be connected to its own dedicated graphics card. The latter solution would allow 30 frames per second operation, but comes at a considerable expense.

Number of Participants and Bidirectional Use

Although the present design does not reflect this, the system is entirely capable of representing as many users in the pod as is needed. The limiting factors are the size of the projection screen, the form factor chosen, and the size of the capture room. We chose a cylindrical form factor that fits one person, but with a larger radius projection ring one can create a cylinder that fits two or more people. The limitations of the one-way system presented in this paper were largely due to budget constraints: placement of ZED cameras on top of and around the projection cylinder allow a full two-way system with TeleHuman2 displays in both locations. Movement constraints can be relaxed by placing ZED cameras further away and by using a larger capture room.

CONCLUSIONS

We presented TeleHuman2, a remote telepresence system that conveys full-sized 3D video of remote participants via a cylindrical light field display. For rendering, the system uses a large array of projectors mounted above the heads of participants in a ring around a human-sized horizontally retroreflective cylinder. Each projector has its own renderer calculating a viewport from a 3D relief map provided by the video capture system at the remote location. This capture system consists of an array of visible light stereoscopic ZED cameras connected to a PC. Locally, each projector image is retro-reflected back for each angle into the eyes of local participants, conveying correct continuous motion parallax and stereoscopy to multiple participants without any need for head-worn apparatus or head tracking. Our technical evaluation of angular accuracy of the system demonstrates that the error in judging the angle of an arrow object represented in TeleHuman2 is within 1 degree, and not significantly different from that of the real arrow object.

REFERENCES

1. Michael Argyle and Janet Dean. 1965. Eye-Contact, Distance and Affiliation. *Sociometry* 28, 3: 289. <http://doi.org/10.2307/2786027>
2. Harlyn Baker and Zeyu Li. 2009. Camera and projector arrays for immersive 3D video. *Proceedings of the 2nd International Conference on Immersive Telecommunications*: 23.
3. Celluon. PicoPro Projector. Retrieved from http://www.celluon.com/products_picopro_overview.php
4. Milton Chen. 2002. Leveraging the asymmetric sensitivity of eye contact for videoconference. In *Proceedings of the SIGCHI conference on Human factors in computing systems Changing our world, changing ourselves - CHI '02*, 49. <http://doi.org/10.1145/503376.503386>
5. Neil A. Dodgson. 2004. Variation and extrema of human interpupillary distance. 36–46. <http://doi.org/10.1117/12.529999>
6. Markus Gross, Silke Lang, Kai Strehlke, et al. 2003. blue-c: a spatially immersive display and 3D video portal for telepresence. In *ACM SIGGRAPH 2003 Papers on - SIGGRAPH '03*, 819. <http://doi.org/10.1145/1201775.882350>
7. Edward T. (Edward Twitchell) Hall. 1992. *The hidden dimension*. Peter Smith Pub.
8. Matthew Hirsch, Shahram Izadi, Henry Holtzman, and Ramesh Raskar. 2013. 8D: Interacting with a relightable glasses-free 3D display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 2209. <http://doi.org/10.1145/2470654.2481304>
9. Matthew Hirsch, Douglas Lanman, Henry Holtzman, and Ramesh Raskar. 2009. BiDi screen: a thin, depth-sensing LCD for 3D interaction using light fields. In *ACM SIGGRAPH ASIA 2009 Art Gallery & Emerging Technologies: Adaptation on - SIGGRAPH ASIA '09*, 62. <http://doi.org/10.1145/1665137.1665183>
10. Frederic E. Ives. 1903. Parallax stereogram and process of making same.
11. Kibum Kim, John Bolton, Audrey Girouard, Jeremy Cooperstock, and Roel Vertegaal. 2012. TeleHuman: Effects of 3D Perspective on Gaze and Pose Estimation with a Life-size Cylindrical Telepresence Pod. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*, 2531–2540. <http://doi.org/10.1145/2207676.2208640>
12. Hao Li, Laura Trutoiu, Kyle Olszewski, et al. 2015. Facial performance sensing head-mounted display. *ACM Transactions on Graphics* 34, 4: 47:1-47:9. <http://doi.org/10.1145/2766939>
13. Gabriel Lippmann. 1908. Epreuves reversibles. Photographies integrals. *Comptes-Rendus Academie des Sciences* 146: 446–451.
14. Andrew Maimone and Henry Fuchs. 2011. Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras. In *2011 10th IEEE International Symposium*

- on *Mixed and Augmented Reality*, 137–146. <http://doi.org/10.1109/ISMAR.2011.6092379>
15. Diego Martinez Plasencia, Abhijit Karnik, Jonatan Martinez Muñoz, and Sriram Subramanian. 2014. Portallax: Bringing 3D displays capabilities to handhelds. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14*, 145–154. <http://doi.org/10.1145/2628363.2628392>
 16. Michael Argyle and Mark Cook. 1976. *Gaze and Mutual Gaze*. Cambridge University Press. Retrieved from <http://www.cambridge.org/ae/academic/subjects/psychology/psychology-general-interest/gaze-and-mutual-gaze>
 17. Microsoft. HoloLens. Retrieved from <https://www.microsoft.com/microsoft-hololens/>
 18. Microsoft Research. Holoportation. Retrieved from <https://www.microsoft.com/en-us/research/project/holoportation-3/>
 19. Daniel Moreno and Gabriel Taubin. 2012. Simple, Accurate, and Robust Projector-Camera Calibration. In *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, 464–471. <http://doi.org/10.1109/3DIMPVT.2012.77>
 20. M Mori. 1970. Bukimi no tani [the Uncanny Valley]. 7: 33–35.
 21. Koki Nagano, Andrew Jones, Jing Liu, et al. 2013. An autostereoscopic projector array optimized for 3D facial display. In *ACM SIGGRAPH 2013 Emerging Technologies on - SIGGRAPH '13*, 1–1. <http://doi.org/10.1145/2503368.2503371>
 22. Andrew Nashel and Henry Fuchs. 2009. Random Hole Display: A non-uniform barrier autostereoscopic display. In *2009 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, 1–4. <http://doi.org/10.1109/3DTV.2009.5069665>
 23. David Nguyen and John Canny. 2005. MultiView. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05*, 799. <http://doi.org/10.1145/1054972.1055084>
 24. Ye Pan, Oyewole Oyekoya, and Anthony Steed. 2015. A Surround Video Capture and Presentation System for Preservation of Eye-Gaze in Teleconferencing Applications. *Presence: Teleoperators and Virtual Environments* 24, 1: 24–43. http://doi.org/10.1162/PRES_a_00213
 25. Ye Pan and Anthony Steed. 2012. Preserving gaze direction in teleconferencing using a camera array and a spherical display. In *2012 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, 1–4. <http://doi.org/10.1109/3DTV.2012.6365433>
 26. Ye Pan and Anthony Steed. 2016. Effects of 3D perspective on head gaze estimation with a multiview autostereoscopic display. *International Journal of Human-Computer Studies* 86: 138–148. <http://doi.org/10.1016/j.ijhcs.2015.10.004>
 27. Ye Pan, Anthony Steed, Ye Pan, and Anthony Steed. 2014. A gaze-preserving situated multiview telepresence system. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, 2173–2176. <http://doi.org/10.1145/2556288.2557320>
 28. Fábio Policarpo, Manuel M. Oliveira, and João L. D. Comba. 2005. Real-time relief mapping on arbitrary polygonal surfaces. In *Proceedings of the 2005 symposium on Interactive 3D graphics and games - SI3D '05*, 155. <http://doi.org/10.1145/1053427.1053453>
 29. Abigail Sellen, Bill Buxton, and John Arnott. 1992. Using spatial cues to improve videoconferencing. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '92*, 651–652. <http://doi.org/10.1145/142750.143070>
 30. K. Tanaka, J. Hayashi, M. Inami, and S. Tachi. 2004. TWISTER: an immersive autostereoscopic display. In *IEEE Virtual Reality 2004*, 59–278. <http://doi.org/10.1109/VR.2004.1310056>
 31. Dongdong Teng, Yi Xiong, Lilin Liu, and Biao Wang. 2015. Multiview three-dimensional display with continuous motion parallax through planar aligned OLED microdisplays. *Optics express* 23, 5: 6007–19. Retrieved September 7, 2016 from <http://www.ncbi.nlm.nih.gov/pubmed/25836825>
 32. James Tompkin, Samuel Muff, James McCann, et al. 2015. Joint 5D Pen Input for Light Field Displays. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology - UIST '15*, 637–647. <http://doi.org/10.1145/2807442.2807477>
 33. USC Shoah Foundation. New Dimensions in Testimony. Retrieved from <http://sfi.usc.edu/collections/holocaust/ndt>
 34. Roel Vertegaal and Yaping Ding. 2002. Explaining effects of eye gaze on mediated group conversations: In *Proceedings of the 2002 ACM conference on Computer supported cooperative work - CSCW '02*, 41. <http://doi.org/10.1145/587078.587085>
 35. Roel Vertegaal, Ivo Weevers, Changuk Sohn, and Chris Cheung. 2003. GAZE-2. In *Proceedings of the conference on Human factors in computing systems - CHI '03*, 521. <http://doi.org/10.1145/642611.642702>