LightBee: A Self-Levitating Light Field Display for Hologrammatic Telepresence

Xujing Zhang, Sean Braley, Calvin Rubens, Timothy Merritt[1] and Roel Vertegaal
Human Media Lab
Queen's University
Kingston, Ontario, Canada
[1]Human Centered Computing
Aalborg University
Aalborg, Denmark

ABSTRACT
LightBee is a novel "hologrammatic" telepresence system featuring a self-levitating light field display. It consists of a drone that flies a projection of a remote user's head through 3D space. The movements of the drone are controlled by the remote user's head movements, offering unique support for non-verbal cues, especially physical proxemics. The light field display is created by a retro-reflective sheet that is mounted on the cylindrical quadcopter. 45 smart projectors, one per 1.3 degrees, are mounted in a ring, each projecting a video stream rendered from a unique perspective onto the retroreflector. This creates a light field that naturally provides motion parallax and stereoscopy without requiring any headset nor stereo glasses. LightBee allows multiple local users to experience their own unique and correct perspective of the remote user's head. The system is currently one-directional: 2 small cameras mounted on the drone allow the remote user to observe the local scene.

CCS CONCEPTS
• Hardware>Emerging Technologies>Emerging interfaces.

KEYWORDS
Telepresence; light field; quadcopter; cylindrical display; glasses-free 3D; projector array.

ACM Reference Format:
Xujing Zhang, Sean Braley, Calvin Rubens, Timothy Merritt and Roel Vertegaal. 2019. LightBee: A Self-Levitating Light Field Display for Hologrammatic Telepresence. In 2019 CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland, UK. ACM, New York, NY, USA. 11 pages. https://doi.org/10.1145/3290605.3300242

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300242

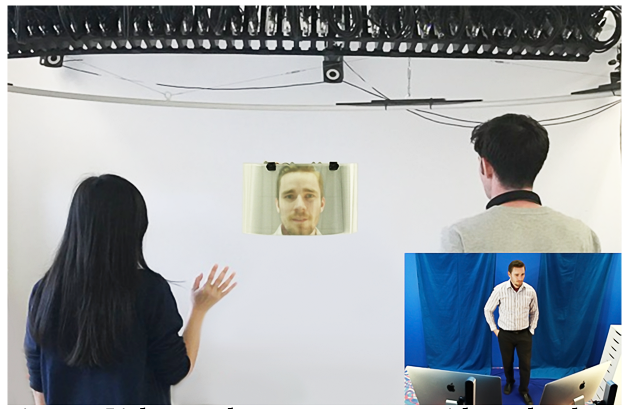

Figure 1. LightBee telepresence system with two local users viewing the drone-based light field display to communicate with a person in a remote capture room (inset).

1 INTRODUCTION
Over the past decades, research on teleconferencing systems has enabled the transmission of essential non-verbal communication cues between local and remote participants over large distances [4,20,28,30,38,43,46]. Prototypes such as TeleHuman [20] convey body orientation cues, facial expressions and eye contact between local and remote users, making it easier to, e.g., take turns in multiparty remote conversations [45]. Displays with head-tracking and stereo glasses now allow both motion parallax and stereoscopy to be conveyed, in life size [13,20]. However, these prototypes do not allow a remote participant to move around a local space, because they focus on the use of static projection systems.

Spatial movement of remote users serves as an important non-verbal cue, one that can enhance social presence. Social presence was defined by Short et al. [39] as reflecting the extent to which a person is perceived as salient. Movements can also convey emotions independently of representation [25]. Interaction distance, specifically, can affect the interlocutors' relationship positively or negatively [10]. Proximity [2] can signal mutual intimacy, and plays a role in the turn taking process [13,38]. Movement also allows participants to monitor the progress of a conversation. For example, by moving away and turning the body, participants may signal the end of a conversation. While augmented reality headsets, such as Microsoft's HoloLens, can create an illusion of movement of participants in their surroundings, these require head mounted displays that make it difficult to capture facial features [31]. As such, telepresence systems that accurately convey physical movement, specifically of the remote user's head, still represent a technical challenge.

1.1 Contribution

In this paper, we present LightBee, a telepresence drone that displays a 3D hologrammatic representation of the remote user's head that moves with the actual head movements of the remote user (see Figure 1). LightBee consists of a self-levitating retroreflective cylinder that reflects images projected on it by a ceiling-mounted radial array of 45 smart projectors in a narrow band of approximately 1 degree of angle. Simultaneous projection of the image of the remote person's head from every degree of angle creates a light field display, which we refer to as "hologrammatic" because an actual hologram only refers to a light field created through laser interference patterns. 3D surround video and movements of the head are captured in the remote user's capture room using an array of stereoscopic ZED video cameras (Figure 1, inset). These are conveyed across a gigabit Ethernet connection as depth maps to the projection rig. The smart projectors then act as a parallel computer, each calculating its own angular view.

The main contribution of this systems paper is a novel self-levitating light field display for use in telepresence. Our contributions over and above existing work are as follows: Our system, to our knowledge, 1) is the first teleconferencing system to feature a self-levitating light field display; 2) the first drone-based display to provide motion parallax and stereoscopy simultaneously to multiple users; 3) the first robotic telepresence system to accurately represent the physical movement of the remote user's head to local participants; 4) due to the drone-based propulsion system, the display moves more quickly than previous wheeled and blimp-based teleconferencing robots.

2 BACKGROUND

We will first discuss prior work in the area of light field display technologies. Next, we will provide an overview of light field displays for use in multiparty video conferencing, and the use of telepresence robots for mobility. Note that the focus of this review is on embodied *multiview* group telepresence systems.

2.1 Approaches to Light Fields

Light field displays visualize 3D scenes by recreating the angular rays of light around a scene. A light field display emanates light in *many* different angles, such that they generate *many* different views. There are three main ways to create light field displays: Using parallax barriers [16]; using integral imaging, also known as microlens arrays [23]; and using multiview retroreflective projection [27]. A parallax barrier display uses an array of pinholes or slits to control what pixels are seen from a particular angle. An integral imaging display uses microlens array overlays that bend pixels into specific angles. Both methods lower the resolution of a display since a user only sees a subset of pixels from any one direction. Multiview projection systems typically use narrow-band retroreflective materials [32] that reflect the light from a set of angled projectors directly back into the eyes of the user. Two key benefits of this technology are that 1) it supports full resolution for all views, and 2) it generates a much brighter image. Next, we will discuss previous use of multiview projection in teleconferencing.

2.2 Multiview Projection Teleconferencing

There has been a wealth of approaches to the development of multiview projection displays for teleconferencing. Multiview [29] was one of the first systems to use multi-projection and a retroreflective screen to provide multiple viewports. Jones et al. [18] proposed a bi-directional teleconferencing system that featured a spinning anisotropic display surface. This display surface reflected rays of light from a high-speed projector into different viewing positions. The system provided a 180° field of view of the head of a remote user. This solution relied on the use of moving parts, and was not ideal for our drone. Jones et al. [17] provided a solution containing an anisotropic screen and a curved projector array. The screen reflected the light rays from each projector as a vertical strip, and the strips were blended using a 1° horizontal diffuser. The prototype provided vertical parallax for multiple participants by means of head tracking. Pan et al. [32] designed a display that applied the full resolution of projectors to a cylindrical retro-reflective screen. The screen reflected each angular projection back into a narrow viewing zone. However, the density of projectors was insufficient to provide a full stereoscopic light field around the cylindrical surface. All of the above approaches provided corrected eye contact in one-to-many teleconferencing scenarios. However, the position of the screens was fixed, meaning remote users could not move around the room to communicate physical proxemics to local users. TeleHuman2 [11] is a telepresence system that uses a circular array of 45 720p projectors to front-project a 3D representation of a person into a life-size cylinder with a retroreflective screen. We were inspired by its design as it allows the projection technology to be located away and around a light weight projection surface. Note that none of the above examples were mobile.

2.3 Telepresence Robots

PRoP [33] was one of the first telepresence robots. It consisted of a camera at 1.5 meters from the floor, an LCD screen right below the camera, a backbone and wheels. The basic construction of subsequent telepresence robots is similar, with varying display designs. Some commercial robots [15,8] have a "Head Agnostic" design, where phones and tablets are interchangeably used as the display. Some robot displays provide depth cues beyond a flat 2D display. Animatronic Shader Lamps Avatars [22] projected a user's face on a life-size Styrofoam head. TELESAR4 [41] featured a retroreflective material as its main display. Local users wore a near-eye projector to project images from different angles on its surface, allowing them to perceive motion parallax.

2.3.1 Remote Control

In terms of the control of telepresence robots, approaches include joystick, keyboard, tablet control, body tracking and semi-autonomous control. PRoP [33] was driven using a joystick. Double 2 [8] could be moved through keyboard or touch input. It also allowed remote users to adjust the height of the screen and the camera. Higuchi et al. [14] proposed an unmanned aerial vehicle (UAV) telepresence robot that used the user's head input to control the UAV's aerial movement. To help users to avoid obstacles, some telepresence robots apply assisted navigation strategies that are beyond the scope of this paper [42,40,19]. Blimps [44] have been used for telepresence purposes, allowing remote users to move unencumbered through the environment. DisplayDrones by Gomes et al. [9] featured a small drone with a visual representation of a remote Skype user. However, this display did not provide 3D depth cues.

2.3.2 Social Cues and Mobility

Researchers have provided telepresence robots with varying degrees of mobility, which can aid in the establishment of social presence. Choi et al. [5] made a wheeled robot that could mimic a remote user's body orientation and movement. They found that the robot's movement affected interpersonal judgements. MeBot [1] and Animatronic Shader Lamps Avatars [22] mapped a remote user's head motion to the head display of the robot in order to convey gaze awareness. ScalableBody [24] allowed users to adjust the height of the robot to maintain the correct eye level, and found this enhanced engagement. Rae et al. [37] found that the *height* of a telepresence robot influenced the local user's impression of dominance from the interlocutor. However, conveying lifelike head movement is still a technical challenge with telepresence robots due to their comparatively slow motor systems.

3 DESIGN RATIONALE

Our overall goal in the design of LightBee was to achieve a lightweight light field display capable of flight that still featured a very dense number of viewports. This allows multiple interlocutors to perceive the remote participant stereoscopically, with continuous motion parallax, without the need for glasses or head tracking. The following design criteria were taken into consideration:

3.1 Drones and Head Representation

There are good reasons to use a drone for telepresence displays. Drones potentially help a display look and behave more human, by adding smooth physical movement. Although this is not a universal requirement, one benefit of using drones is that they do not require structural support, offering no exposed mechanical parts that impact human likeness. Drones, if properly powered and tracked, have no limitations in terms of mobility, particularly in the 3rd dimension. Compared with wheeled robots, the 3D spatial location of a drone is readily controllable, allowing drone displays to smoothly replicate a remote user's behavior. Blimp-based displays have not been able to match the speed of human head movements [45], and the size of the envelope must be very large in order to provide adequate lift. A drone-based solution has the potential to move very quickly to match natural head movements and the display size can be chosen to match the head size.

We chose to represent only the remote user's head, mostly for practical reasons. Flying a screen large enough for full body projection is not possible with current payloads of indoor drones. We chose representing the head because it communicates facial expressions and gaze, and is a rich source of non-verbal information. Rather than relying on animation, we chose to use 3D volumetric video of the head. This helped avoid the Uncanny Valley effect common in avatar-style animations [26]. Accurate eye contact and head orientation images are automatically provided by the rendition of 3D volumetric video via the light field projection, and can be adjusted using calibration of the remote capture system.

3.2 Using Head Control

Driving a robot requires attention [36], and this influences the efficiency of communication [21]. However, using the remote user's head movement as input allows for intuitive control over the local drone [14] as head gestures belong to the same schema that are used to control looking behaviour [34]. The synchronization of the drone with the remote user's head motion also serves as an important nonverbal cue to local participants [25]. In our design, the 3D location of LightBee is controlled directly by the remote 3D head location to mimic the proxemics that take place in collocated conversation. All movements of the head model are mapped to realistically convey both direction and proximity of the remote person, as well as eye contact. When the head moves further away it becomes smaller, as does the drone surface.

3.3 Freedom of Movement and Manual Control

An important requirement was that participants are able to move comfortably within Hall's Social Distance (1.2m-3.6m) [13] without losing stereoscopy or motion parallax. When the LightBee is out of range of the projectors, users rely on manual control. Note that the latter method of control is beyond the scope of this paper: when the drone is not projected on or tracked, users are, however, able to explore the remote space using two drone-mounted cameras.

3.4 Form Factor

To provide a potential 360° horizontal viewing area, we decided to use a cylindrical screen form factor. This has the additional benefit of allowing the quadcopter to be mounted hidden within the cylinder, while providing space for airflow. The size of the cylinder would need to approximate that of the average human head.

3.5 Glasses-Free Projected Light Field

One of our main goals was to free users from wearing any helmets or glasses in order to increase the users' comfort level and their ability to see one another. For this purpose, we chose to project a light field on the drone's surface using an exterior projector array. Another method would have been to generate a light field using spinning LED strips on the surface of the drone [6], but aside from negative effects on flight dynamics there are other considerable engineering costs associated with this method. To render a high-quality light field, a powerful onboard computer is required, with associated high-bandwidth network boards that are beyond the current payload of interior drones. Our approach of combining a self-levitating retro-reflective screen with an external light field projector array provided a low weight display solution that removed the need for head tracking, shutter glasses or other complex head-worn apparatus.

3.6 Simultaneous Stereoscopy and Horizontal Motion Parallax

Although both stereoscopic and motion parallax information provide effective depth cues [12], the combination of these cues improves the precision of depth perception [3]. We designed a ring of tightly packed external projectors such that multiple local interlocutors would be provided with different stereoscopic views through the alignment of their left and right eyes with different projector units. They would be provided with simultaneous motion parallax as the eyes align to the different images while moving along the projector array. Since most human movement during conversations is horizontal, and to avoid occlusion of the display by a two-dimensional array of projectors, we decided to limit the light field to horizontal parallax only. This allows the projectors to be mounted above the users' heads without impacting the horizontal 3D views.

3.7 Angular Resolution

To provide stereoscopy to each local interlocutor, the angular resolution of the cylindrical display had to be smaller than the average interpupillary distance (6.3 cm for adults [7]). We designed the projector ring such that the distance between two viewports would provide stereoscopy at a comfortable social distance (1.2m – 3.6m [13]).

3.8 Asymmetrical telepresence

While the system could, in principle, be designed such that it is bi-directional, we focused on the light field experience delivered by the drone, limiting the experience of the remote person to two displays showing images from two on-board drone cameras. As such, we delivered a one-way illusion of telepresence while maintaining the ability of the remote user to fly through the space independently of the light field representation or local positioning trackers.

3.9 Surround 3D Video Capture

Interference is an issue when capturing volumetric 3D video using infrared structured light depth cameras. To avoid this, we chose to apply stereo cameras that operate in visible light. To reduce computational complexity and network load, images would need to be sent from each camera to a cluster of projectors, each augmented with computational ability to render its own custom view.

Figure 2. The LightBee prototype viewed from above. Special rotor blades direct airflow through the cylinder.

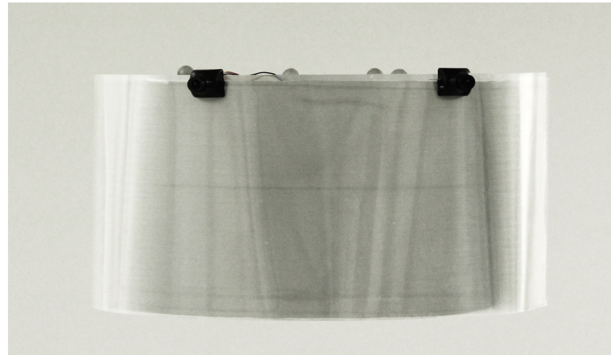

Figure 3. LightBee prototype with two cameras and retroreflective surface.

4 IMPLEMENTATION

Next, we discuss the implementation of the self-levitating cylindrical display, which includes the quadcopter hardware, the cylindrical retroreflective screen, and the flight control system. After that, we discuss the light field rendering approach, the remote capture system, and how the quadcopter-captured video is transmitted.

4.1 Flying Cylindrical Retroreflector

Figures 2 and 3 show the LightBee prototype, a quadcopter of our own design with a hollow retroreflective cylindrical surface. The quadcopter is outfitted with 4 brushless motors, a Micro MultiWii flight controller board and an ESP-8266 WiFi breakout board. The flight controller is a customized version of the MultiWii software modified to work with the brushless motors, and with our custom communications protocol. Power is provided by a 1300 mAh battery that sustains approximately 4 minutes of flight time at present. There are two small cameras mounted on the top of the screen to provide visual feedback to the remote user.

4.2 Size and Aerodynamics

The drone is 38 cm in diameter and 22 cm in height. We made the diameter larger than that of an average human head for aerodynamic purposes. This is because the walls of the cylindrical screen act as a drag surface for the air moved by the propellers, decreasing the effective thrust output. As the height of the screen increases, the air pressure above the quadcopter decreases. The resulting design was a compromise between average head size and optimal thrust capacity. Four angled 9 cm propellers power the flight (Figure 2). The screen materials consist of a microbead retroreflective cloth wrapped around a cylindrical film, with a one-dimensional diffuser layer. The diffuser (RPC Photonics™ EDF-L1) features a 56° vertical scatter angle, which allows for retroreflective projection without a significant drop in brightness when the quadcopter moves up or down.

4.3 Flight Control

On top of the quadcopter, reflective motion capture markers allow tracking of position and rotation by a Vicon MX Motion Capture System. Coordinates are provided to the flight controller at 100 Hz via the Vicon Tracker 3.4 software. We designed a control system to calculate the waypoints of the quadcopter, as well as to govern its behavior as it flies between waypoints. The waypoints are set based on the remote user's head movements. These head movements are tracked at 60Hz by a Kinect v2.0 through the Kinect for Windows SDK 2.0 and transmitted through a network to the flight control system. The quadcopter can automatically fly to these waypoints, through Proportional, Integral and Differential (PID) control loops, which adjust the quadcopter's thrust along the x, y and z axes.

4.4 Projector Array

To hold the projector array, a circular rail of 180 cm radius was suspended from the ceiling at a height of approximately 195 cm. Mounted underneath this rail are 45 PicoPro 720p focus-free laser projectors that each project a different angular image (see Figure 4). Each projector image is retro-reflected back into the eyes of users by the drone surface, producing both stereoscopy (left and right eye catching a different projector image) and motion parallax (eyes moving between projector images) to local users. Since all the light from each projector is retro-reflected back into a very narrow angle, the 32 Lumens of each projector produces a sufficiently bright image even in broad daylight. To provide stereoscopy for users standing directly below the projector rail as indicated in Figure 4, projectors are spaced approximately 4.5 cm apart horizontally, slightly closer than the average inter-pupillary distance [7]. The projector array provides nearly 59º of motion parallax, which can be extended by adding more projectors.

Figure 4. Projector rail with 45 projectors, each with an Odroid rendering board.

4.5 Light Field Rendering

We distributed the rendering for each viewport to 45 Odroid C1+ boards, one connected to each projector. Depth images are captured by stereo cameras and broadcast by the capture system over UDP. The software running on each Odroid receives the color and depth images from the remote system to produce a relief map. The rendering software then renders a viewport at 720p for each projector by calculating the offset of its projector in this relief map [35], at ~10 FPS. This image is subsequently mapped and projected onto the drone surface. One nice benefit of this method is that the hologram appears rendered in the correct location regardless of the exact location of the projection surface. Colour and depth data of the remote head are applied to an OpenGL rectangle which is rendered in 3D using a relief mapping shader, in a method similar to Policarpo et al. [35].

4.6 Remote Capture System

Figure 5 shows the remote capture room. Here, the user is able to move within a circle of 3 ZED cameras, 1 ZED camera per 20 degrees, each capturing a 3D volumetric video of the head. The resolution of a ZED camera is 2 x 2K, and the frame rate of stereo images is 15 frames per second. Unlike the Kinect, ZED cameras do not suffer from interference when multiple cameras are used and can be used in bright daylight. Each ZED camera is connected to a high-end NVidia GTX 1080 graphics card on an Intel i5 computer. Our software is implemented on Ubuntu 16.04 to capture the user's color and depth images, with background subtraction. Each ZED camera is in charge of producing a relief map for a cluster of 15 projectors. The 3 ZED cameras in the remote location provide approximately 59 degrees of

Figure 5. The capture room with 3 ZED cameras, and two displays showing images from the drone cameras.

motion parallax. Extending motion parallax only requires adding more ZED cameras and processing power.

4.7 Quadcopter Video

Two small 26*x*26 mm cameras are mounted 60° apart on top of the quadcopter (see Figure 3). Video from these cameras is sent via two on-board video transmitters to a local computer. The video streams are encoded and sent across the network to the remote capture room, where they are displayed on two monitors spaced approximately 60° apart. Since the setup is not symmetrical, the remote user can only see images from each drone camera on these monitors. Although the images move around with the drone camera, the monitors themselves are in a static location.

5 INTERACTIONS

We designed for a limited number of interactions with the LightBee. Most of these center around the correct conveyance of head-related cues:

Head Proximity – By moving the head forward or backwards, proximity cues can be conveyed. Because the drone is a physical object, when it moves closer towards you it provides a more powerful cue than does video.

Head Orientation – Head orientation is mostly conveyed through the light field, by rendering the head from an angle. It can convey interest in a person.

Eye Contact – Eye contact is conveyed and corrected for by rotating the head model and positioning the drone such that it is in line with eye gaze by the onlookers. The use of a stereo camera means there is no noticeable parallax.

Speech – Speech is conveyed using a Dolby surround sound speaker system in the ceiling.

6 APPLICATION SCENARIOS

The LightBee system can be used in a variety of applications, including entertainment, remote education and telepresence. In this paper, we focused on its use as a telepresence robot. We further examine its utility using the following two user scenarios, future envisionments that highlight key features and benefits of LightBee's unique method of remote telepresence.

6.1 Telepresence for Business

A majority of business meetings are still conducted inside specific meeting rooms that are equipped with video conferencing platforms. It is not inconceivable that, in the future, such platforms would include light field projection, as per the following scenario:

When business people collaborate from different locations while working on projects, there is often a need for sharing information about the space, or to demonstrate an artifact or experience that takes place in the built environment. When collocation is not possible, or when users need to visit multiple sites, LightBee can be used to set up a telepresence session, which enables a person to be invited in to explore the physical room remotely. This is especially useful in situations where buildings need to be simultaneously inspected and discussed, for example, in real estate, architectural design or construction (where drones are already widely applied for this purpose).

Hologrammatic telepresence via drones reduces the need to travel, while still providing many of the benefits of non-verbal communication, making it easier to hop between job sites. With the front facing cameras and ability to fly over obstacles, the remote participant is able to inspect and navigate within a remote job space, even when it is not outfitted with a tracking system. This emphasizes the fluidity of drone applications as compared to traditional conferencing systems. In this scenario, a remote architect is visiting a job site to discuss details of the build quality with his clients. When ready to remotely join a face-to-face meeting, the remote user navigates his LightBee to a meeting room in a building next to the job site, which is outfitted with a projection ring and tracking system. As the meeting starts, the projector array begins projecting on the LightBee's retroreflector to provide local users with a synchronized three-dimensional view of the remote participant's face. The local participants engage in a conversation about the building project. The head movements of the remote participant and eye contact convey non-verbal cues, which help the local participants understand to whom the remote person is talking, facilitating turn taking. In collocated conversation, interlocutors often use such pronouns as "you" or "that", which require disambiguation. The remote participant simply orients his head to refer to the architectural model under discussion. Such head pointing is naturally supported by the LightBee's light field imaging, facilitating deixis towards the model without requiring verbal disambiguation. The remote participant moves closer to one of the clients in an attempt to emphasize that the design needs alterations. The client responds to the physical proxemics of the LightBee, and agrees to the request. The co-workers finalize the decision on the action plan and schedule, and wave while the remote user turns around and flies off to inspect another part of the building site.

6.2 TelePresence for Entertainment

Large performance venues are now routinely equipped with powerful projection technologies as a main entertainment feature (see, e.g., Olympic opening ceremonies). The addition of large and high-powered projector arrays would allow floating holograms to be rendered across a wide range of drone movement and across large distances inside a stadium, as per the following scenario:

A DJ is playing a set in an indoor stadium with 10,000 people. Specific areas in the upper levels of the stadium have been outfitted with 360 degree projector arrays, and the stage has been outfitted with a volumetric video capture array. Normally, the DJ is confined to the stage, standing behind her console from which she spins her tunes. With multiple copies of herself on LightBees hovering throughout the stadium, her audience feels much more connected. As the DJ dances, the LightBees hop up and down to the beat, whipping the crowd up to take selfies with the drones.

The LightBee replicas are not limited to one stadium, and the DJ is able to sell out several more venues that night in different cities where she is only represented by the telepresence drones. Not only does this allow her to be in multiple places at once, her revenue is significantly increased due to the ability to perform in multiple simultaneous shows, at a significantly reduced travel cost.

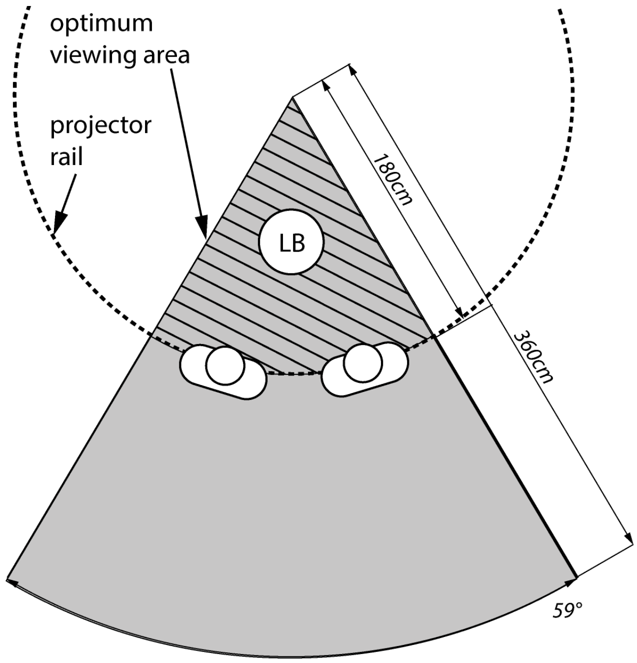

Figure 6. Overview of active area with LightBee (LB) and projector rail. Optimum viewing area is indicated in grey, while optimal flight area is hatched.

7 INITIAL USER EXPERIENCES

In order to gain some initial insight into how users might experience interactions via the LightBee system, we performed a preliminary evaluation in which we asked 6 participants (2 female and 4 male) between 25 and 30 years of age to engage in a triadic conversation for 3 minutes. In each conversation two local participants conversed with one remote person represented on a LightBee. The remote person was an actor who made sure to turn his head toward each local interlocutor while speaking to him or her. The actor also moved toward each local user when speaking, and away when not speaking. In Figure 6, a plan view diagram illustrates the area in which participants (gray) and the LightBee (hatched) were allowed to move. To begin each session, participants were briefed about the system and the confines of the active space, which was between 180 and 360 cm from the center of the projection ring. After each session, the participants were asked to reflect upon their experience, after which they were interviewed by the experimenter. All participants were successfully able to interact with the remote person and overall responses were positive. We now provide details about the initial impressions of the system and comment on aspects that appeared to affect user experience.

7.1 Results

Most participants noted that the movements of LightBee helped them facilitate turn taking and management of the conversation. P4 claimed that *"The movement (of the LightBee) gave me a better idea of what the remote person was doing."* P6 claimed that the LightBee's movement was *"easy to understand and smooth."*

Three participants commented on their experience of proxemic movements by the LightBee. P1 claimed that he *"felt a kind of pressure when it approached me"* that helped him to know that the remote person was engaging with him. P6 also noted, *"the eye contact was really helpful, like in a class or meeting."* P3 claimed that he wanted the remote person to always face him when talking. P2 noted that it was convenient to *"show the person around"* by moving herself and P3 claimed that it was *"cool to look around the person by moving around (the LightBee)."* P1 stated that he *"did not move around when talking to the other person"*, yet seemed to appreciate LightBee's movements towards him.

Some participants commented on the 'missing body' of the remote person, yet after further enquiry suggested this did not necessarily negatively impact their conversation. Perhaps the relatively high resolution of the head projection and corresponding head movements of the remote person led to an Uncanny Valley effect, with participants expecting there to also be hand movements and body cues [26]. Two participants complained about the noise of the drone. P2 did not like *"how loud the propellers are"*. None of the participants commented on the lack of vertical parallax.

8 DISCUSSION AND FUTURE DIRECTIONS

There are, of course, numerous issues with the use of a drone for telepresence. The system is, at present, asymmetrical, i.e., the remote user does not obtain a similar sense of the proximity or direction of the local users. This can complicate a correct mapping of eye gaze direction if the remotely perceived location of local users does not correspond to their correct local angle. This can easily be solved by making the system bi-directional (i.e., drones in both locations). LightBee currently supports only indoor conversations due to limitations in the tracking system and projector array. The light field and cameras mounted on the drone are currently limited to a < 60-degree field of view. This is, however, only due to budget constraints and can be addressed by adding more cameras and projectors. In the current implementation, the propellers of the drone create a clearly audible noise. We expect to address this in future versions using anti-sound noise-cancelling speakers, or by asking users to wear Bluetooth earbuds. Additionally, improved flight dynamics and rotor design can significantly cut down on noise generated by the propellers. Battery life is another limitation, with flight time constrained to about 4 minutes. This is mostly due to the weight of the drone in relation to its limited battery capacity. A better mechanical design of the rotors and air flow dynamics would increase efficiency, for example, by using ducted propellers. Our current system only captures the representation of a single remote user. That said, multiple remote users can fly together provided each user is represented by their own LightBee. Head controlled movements are currently limited to within the remote capture space. However, the drone can be flown using manual control beyond the confines of this space. After we improve the engineering and design of the drone, we plan to conduct a more formal study of the potential influence of LightBee on telepresence experience.

9 CONCLUSIONS

We presented LightBee, a hologrammatic self-levitating telepresence drone that conveys a 3D image of the remote participant's head via a cylindrical light field display. The light field display consists of a retroreflective screen mounted on a cylindrical quadcopter. This screen is projected on by a circular array of 45 projectors mounted above the heads of local interlocutors. Each projector has its own renderer that, in parallel, calculates a viewport from a 3D relief map captured at the remote location. This capture system comprises an array of visible light stereo cameras connected to a PC. The remote user's head movement is tracked and transmitted to control the location of the drone. Our light field display provides continuous motion parallax and stereoscopy to multiple local interlocutors without any need for head-worn apparatus. It allows correct horizontal gaze awareness between local and remote participants. Moreover, the movement of the display provides the remote user with the ability to explore the local environment, even when not projected upon. We discussed results from initial evaluations with users indicating that the telepresence system could support interactions between remote users, and concluded with initial insights into areas of improvement.

ACKNOWLEDGEMENTS

We thank Daniel Gotsch for his help in implementing the light field rendering software.

REFERENCES

[1] Sigurdur O. Adalgeirsson and Cynthia Breazeal. 2010. MeBot: a robotic platform for socially embodied presence. *In Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction (HRI '10).* IEEE Press, Piscataway, NJ, USA, 15-22.

[2] Michael Argyle and Janet Dean. 1965. Eye-Contact, Distance and Affiliation. *Sociometry* 28, 3: 289. https://doi.org/10.2307/2786027

[3] Marius Braun, Ulrich Leiner, and Detlef Ruschin. 2011. Evaluating motion and binocular parallax as depth cues for autostereoscopic displays. of *SPIE, the International Society for Optical Engineering*. Society of Photo-Optical Instrumentation Engineers.

[4] William A. S. Buxton, Abigail J. Sellen, and Michael C. Sheasby. 1997. Interfaces for multiparty videoconferences. *Video-mediated.*

[5] Mina Choi, Rachel Kornfield, Leila Takayama, and Bilge Mutlu. 2017. Movement Matters: Effects of Motion and Mimicry on Perception of Similarity and Closeness in Robot-Mediated Communication. *In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17*). ACM, New York, NY, USA, 325-335. https://doi.org/10.1145/3025453.3025734

[6] Docomo Drone Project. Retrieved September 14, 2017 from https://phys.org/news/2017-04-world-spherical-drone.html

[7] Neil A. Dodgson. 2004. Variation and extrema of human interpupillary distance. *Electronic imaging 2004.* http://doi.org/10.1117/12.529999

[8] Double 2. Retrieved September 14, 2017 from https://www.doublerobotics.com/double2.html

[9] Antonio Gomes, Calvin Rubens, Sean Braley, and Roel Vertegaal. 2016. BitDrones: Towards Using 3D Nanocopter Displays As Interactive Self-Levitating Programmable Matter. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*: 770–780. https://doi.org/10.1145/2858036.2858519

[10] Rita Gorawara-Bhat, Mary Ann Cook, and Greg A. Sachs. 2007. Nonverbal communication in doctor-elderly patient transactions (NDEPT): Development of a tool. *Patient Education and Counseling* 66, 2: 223–234. https://doi.org/10.1016/j.pec.2006.12.005

[11] Dan Gotsch, Xujing Zhang, Tim Merritt, and Roel Vertegaal. 2018. TeleHuman2: A Cylindrical Light Field Teleconferencing System for Life-size 3D Human Telepresence. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM, New York.

[12] Maureen Graham and Brian Rogers. 1982. Simultaneous and Successive Contrast Effects in the Perception of Depth from Motion-Parallax and Stereoscopic Information. *Perception* 11, 3: 247–262. https://doi.org/10.1068/p110247

[13] Edward Hall. 1969. *The Hidden Dimension : man's use of space in public and in private.*

[14] Keita Higuchi and Jun Rekimoto. 2013. Flying head: a head motion synchronization mechanism for unmanned aerial vehicle control. *In CHI '13 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '13). ACM, New York, NY, USA, 2029-2038. DOI: https://doi.org/10.1145/2468356.2468721

[15] iRobot. Retrieved September 14, 2017 from http://www.irobot.com/filelibrary/pdfs/hrd/ava/Ava-Brochure-May2-2011.pdf

[16] Frederic E. Ives. 1903. Parallax stereogram and process of making same.

[17] Andrew Jones, Koki Nagano, Jing Liu, Jay Busch, Xueming Yu, Mark Bolas, Paul Debevec. 2014. Interpolating vertical parallax for an autostereoscopic three-dimensional projector array. *Journal of Electronic Imaging.*

[18] Andrew Jones, Magnus Lang, Graham Fyffe, Xueming Yu, Jay Busch, Ian McDowall, Mark Bolas, and Paul Debevec. 2009. Achieving eye contact in a one-to-many 3D video teleconferencing system. *ACM Transactions on Graphics* 28, 212: 1. https://doi.org/10.1145/1531326.1531370

[19] Handry Khoswanto, Petrus Santoso, and Resmana Lim. 2016. Odometry Algorithm with Obstacle Avoidance on Mobile Robot Navigation. *In Proceedings of Second International Conference on*

Electrical Systems, Technology and Information 2015 (ICESTI 2015), pp. 155-161. Springer, Singapore.

[20] Kibum Kim, John Bolton, Audrey Girouard, Jeremy Cooperstock, and Roel Vertegaal. 2012. TeleHuman. *In Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, 2531. https://doi.org/10.1145/2207676.2208640

[21] Min Kyung Lee and Leila Takayama. 2011. "Now, i have a body": uses and social norms for mobile remote presence in the workplace. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11). ACM, New York, NY, USA, 33-42. https://doi.org/10.1145/1978942.1978950

[22] Peter Lincoln, Greg Welch, Andrew Nashel, Adrian Ilie, Andrei State, and Henry Fuchs. 2009. Animatronic Shader Lamps Avatars. *In Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality* (ISMAR '09). IEEE Computer Society, Washington, DC, USA, 27-33. http://dx.doi.org/10.1109/ISMAR.2009.5336503

[23] Gabriel Lippmann. 1908. Epreuves reversibles. Photographies integrals. *Comptes-Rendus Academie des Sciences* 146: 446–451

[24] Akira Matsuda, Takashi Miyaki, and Jun Rekimoto. 2017. ScalableBody: a telepresence robot that supports face position matching using a vertical actuator. In *Proceedings of the 8th Augmented Human International Conference* (AH '17). ACM, New York, NY, USA, , Article 13 , 9 pages. DOI: https://doi.org/10.1145/3041164.3041182

[25] Rachel McDonnell, Sophie Jörg, Joanna McHugh, Fiona N. Newell, and Carol O'Sullivan. 2009. Investigating the role of body shape on the perception of emotion. *ACM Transactions on Applied Perception* 6, 3: 1–11. https://doi.org/10.1145/1577755.1577757

[26] Masahiro Mori. 1970. The uncanny valley. IEEE Robotics & Automation Magazine. 19 (2): 98-100.

[27] 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

[28] Carman Neustaedter, Gina Venolia, Jason Procyk, and Daniel Hawkins. 2016. To Beam or Not to Beam: A Study of Remote Telepresence Attendance at an Academic Conference. *In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16). ACM, New York, NY, USA, 418-431. DOI: https://doi.org/10.1145/2818048.2819922

[29] David Nguyen and John Canny. 2005. MultiView: spatially faithful group video conferencing. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*: 799–808. https://doi.org/10.1145/1054972.1055084

[30] Ken-Ichi Okada, Fumihiko Maeda, Yusuke Ichikawaa, and Yutaka Matsushita. 1994. Multiparty videoconferencing at virtual social distance: MAJIC design. *CSCW '94: Proceedings of the 1994 ACM conference on Computer supported cooperative work: 385–393*. https://doi.org/http://doi.acm.org/10.1145/192844.193054

[31] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. 2016. Holoportation: Virtual 3D Teleportation in Real-time. *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*: 741–754. https://doi.org/10.1145/2984511.2984517

[32] Ye Pan and Anthony Steed. 2014. A gaze-preserving situated multiview telepresence system. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14). ACM, New York, NY, USA, 2173-2176. DOI=http://dx.doi.org/10.1145/2556288.2557320

[33] Eric Paulos and John Canny. 1998. PRoP: Personal Roving Presence. *In Proceedings of the Conference on Human Factors in Computing Systems.* Los Angeles, CA, 296–303

[34] Ekaterina Peshkova, Martin Hitz, and Bonifaz Kaufmann. 2017. Natural Interaction Techniques for an Unmanned Aerial Vehicle System. *IEEE Pervasive Computing.*

[35] 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*, 155–162. https://doi.org/10.1145/1053427.1053453

[36] Irene Rae, Bilge Mutlu, and Leila Takayama. 2014. Bodies in motion: mobility, presence, and task awareness in telepresence. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14). ACM, New York, NY, USA, 2153-2162. https://doi.org/10.1145/2556288.2557047

[37] Irene Rae, Leila Takayama, and Bilge Mutlu. The influence of height in robot-mediated communication. *In Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pp. 1-8. IEEE Press, 2013.

[38] 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), Penny Bauersfeld, John Bennett, and Gene Lynch (Eds.). ACM, New York, NY, USA, 651-652. DOI=http://dx.doi.org/10.1145/142750.143070

[39] John Short, Ederyn Williams, and Bruce Christie. 1976. *The social psychology of telecommunications.*

[40] Mohammed Tahri Sqalli, Kyoichi Tatsuno, Koichi Kurabe, Hiroto Ando, Hideharu Obitsu, Ryoma Itakura, Takatoshi Aoto, and Katsumi Yoshino. 2016. Improvement of a tele-presence robot autonomous navigation Using SLAM algorithm. *In Micro-NanoMechatronics and Human Science (MHS)*, 2016 International Symposium on, pp. 1-7. IEEE.

[41] Susumu Tachi, Kouichi Watanabe, Keisuke Takeshita, Kouta Minamizawa, Takumi Yoshida, and Katsunari Sato. 2011. Mutual telexistence surrogate system: Telesar4-telexistence in real environments using autostereoscopic immersive display. *In Intelligent Robots and Systems (IROS)*, 2011 IEEE/RSJ International Conference on, pp. 157-162. IEEE.

[42] Leila Takayama,Eitan Marder-Eppstein, Helen Harris, and Jenay M. Beer. 2011. Assisted driving of a mobile remote presence system: System desiagn and controlled user evaluation. *In Robotics and Automation (ICRA)*, 2011 IEEE International Conference on, pp. 1883-1889. IEEE,https://doi.org/10.1145/3064663.3064707

[43] John C. Tang and Scott Minneman. 1991. VideoWhiteboard: Video Shadows to Support Remote Collaboration. *Proceedings of the SIGCHI* conference *on Human factors in computing systems Reaching through technology* - CHI '91: 315–322. https://doi.org/10.1145/108844.108932

[44] H. Tobita, S Maruyama, and T Kuzi. 2011. Floating avatar: telepresence system using blimps for communication and entertainment. *CHI'11 Extended Abstracts on Human.*

[45] 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. https://doi.org/10.1145/587078.587085

[46] Roel Vertegaal, Ivo Weevers, Changuk Sohn, and Chris Cheung. 2003. GAZE-2: conveying eye contact in group video conferencing using eye-controlled camera direction. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '03). ACM, New York, NY, USA, 521-528. DOI=http://dx.doi.org/10.1145/642611.642702