Design and Implementation of a Multi-person Fish-Tank Virtual Reality Display

Dylan Brodie Fafard
dylan.fafard@usask.ca
University of Saskatchewan

Qian Zhou
qzhou@ece.ubc.ca
University of British Columbia

Chris Chamberlain
chris.chamberlain@usask.ca
University of Saskatchewan

Georg Hagemann
georg_hagemann@gmx.net
University of British Columbia

Sidney Fels
ssfels@ece.ubc.ca
University of British Columbia

Ian Stavness
ian.stavness@usask.ca
University of Saskatchewan

Figure 1: Our multi-person spherical FTVR display. From left to right (users' views inset): a single person viewing a fishbowl scene with perspective-correction and stereo, two people collaborating with two independent perspective-corrected views of the same scene, and a third person using a tracked mobile screen to get a perspective-corrected augmented reality view.

ABSTRACT

A mixed reality experience with a physical display, that situates 3D virtual content within the real world, has the potential to help people work and play with 3D information. However, almost all of such "fish tank virtual reality" (FTVR) systems have been isolated to a single-person experience, making them unsuitable for collaborative tasks. In this paper, we present a display system that allows two people to have unobstructed 3D perspective views into a spherical display while still being able to see and talk to one another. We evaluated the system through qualitative observation at a four-day exhibition and found it was effective for providing a convincing, shared 3D experience.

CCS CONCEPTS

• **Human-centered computing → Collaborative interaction**;
• **Computing methodologies → Mixed / augmented reality**;

VRST '18, November 28-December 1, 2018, Tokyo, Japan

ACM ISBN 978-1-4503-6086-9/18/11…$15.00
https://doi.org/10.1145/3281505.3281540

KEYWORDS

Fish tank virtual reality, collaboration, co-location, stereo, 3D displays, spherical displays

ACM Reference Format:
Dylan Brodie Fafard, Qian Zhou, Chris Chamberlain, Georg Hagemann, Sidney Fels, and Ian Stavness. 2018. Design and Implementation of a Multi-person Fish-Tank Virtual Reality Display. In *VRST 2018: 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18), November 28-December 1, 2018, Tokyo, Japan.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3281505.3281540

1 INTRODUCTION

One of the main challenges of virtual and mixed reality technology is to create a shared multi-person 3D interactive experience in the same place at the same time. In headset virtual reality, without accurate, physically based avatars, multiple people interacting in the same 3D space are not able to see each other and have a difficult time agreeing upon and pointing at the same 3D location. Fish tank virtual reality (FTVR) enables the experience of seeing the virtual world situated in the real world, and thus provides the building block for multiple people to see each other while interacting with virtual 3D content. However, because FTVR uses viewpoint tracking and stereo depth cues for representing 3D objects, it has traditionally been restricted to a single viewer.

To address this challenge, we have created a multi-person FTVR display. We describe the hardware and software infrastructure we developed for the highest-fidelity volumetric FTVR display that has been reported to date. The system supports perspective-corrected rendering in stereoscopic view for a single viewer, in mono for two viewers, and with additional mobile screens for extra viewers. It integrates high precision multi-projector calibration, perceptual viewpoint calibration, high fidelity motion tracking and active flicker glasses to create a high framerate, high-resolution spherical display capable of presenting a multi-person mixed reality experience. The hardware components are modestly priced and the software is open-source, thus it is possible to replicate the system for research studies on interaction and perception with high-fidelity FTVR displays. We successfully demonstrated our display at a four-day exhibition with multiple people cooperating in a "find the fish" game, competing in a 3D ping-pong game or working together in a cooperative apple grabbing game.

We first discuss some of the related work and technologies used in different mixed and virtual reality systems. We then describe the hardware and software used to create a multi-projector spherical display. We also explain how motion tracking is used to enable viewpoint and device tracking. Next, we discuss the different uses of the tracking system with active flicker glasses, displays and input devices to support multi-person interaction modes. We then present some of the cooperative and competitive demonstration applications to illustrate how people can work together in this novel mixed reality experience. Finally, we discuss people's experience playing together in virtual 3D and consider possible directions to improve the experience.

2 RELATED WORK

FTVR was originally proposed with a single desktop display [25]. The important finding of the original FTVR studies was a comparison of different visual cues. For a variety of 3D tasks, they found that while head-tracking and stereo cues together were best, head-tracking alone resulted in better performance compared to stereo cues alone [1, 25, 26]. This initial finding motivated many follow-on FTVR displays that neglected stereo cues so that they did not require any headgear or glasses [20].

Multi-screen FTVR displays: A significant extension of the FTVR concept used multiple screens to construct multi-sided FTVR displays. The first multi-screen FTVR displays used three flat LCD panels arranged into convex corner [13] or three projectors projecting into a concave corner [9]. The advantage of the multi-screen display is that it allowed for a larger range of head movement around the screen and therefore enhanced the motion-parallax 3D cues that were found important in the early single-screen displays. The concept was further refined with screens on five sides of a box to give the illusion of an enclosed volumetric display allowing a viewer to view all sides of the 3D objects "inside" the display [22, 23]. Despite the increase in motion-parallax cues and fidelity offered by these modern variants of the FTVR display, the FTVR experience has remained limited to a single person experience.

Spherical displays: A limitation of cubic displays, constructed with flat screen panels, is the large seams that inevitably exist on the edges of the cube. The seam size is a function both of the screen bezel and the screen thickness, since screens are abutted at right angles facing outward. Alternatively, a seamless enclosed display can be realized by projection onto a convex display surface. Projection-based cubic and cylindrical displays have been proposed [11, 14], but spherical surfaces are the most popular shape for such seamless displays. Spherical displays have also been shown to be effective for presenting virtual avatars and avoided negative bias associated with rendering on flat screens [18]. Spherical displays have been realized in a variety of designs, including using a single projector [7], a projector and mirror [6], front projection [4], and multiple projectors [2, 10, 29]. Using multiple projections has a number of advantages over other designs, notably by having scalable resolution (by adding more projectors) and achieving relatively uniform resolution over the display surface as compared to optically warping a single projector to cover the surface. The challenge of the multi-projector approach is the the projectors must be calibrated to each other and blended to provide a smooth transition among adjacent overlapping projection regions. A number of previous studies have addressed the problem of calibrating projectors on curved projection surfaces [19, 21, 28].

Surface vs. FTVR rendering: One approach to multi-person visualization and interaction on a spherical display uses a rendering that appears on the surface of the sphere such as done with [3]. This approach to multi-person interaction is limited to 2D interaction as the displayed content appears as if on the display surface. Typical uses of this approach are to interact with geographic information on the earth as the earth's surface fits naturally on the surface of the sphere. Although the content is not 3D, as there is no stereo rendering or motion parallax cues, the surface rendering does allow multiple people to see the same 2D scene. However, for FTVR, the image needs to be rendered relative to the view point of the user including right and left views for stereo. Because of this need, the calibration requirements and rendering approach for multiple users is more challenging. In particular, the multi-camera display calibration requires the ability to render multiple images at the same time to the spherical surface accurately, track multiple users simultaneously with low latency and have a mechanism to ensure the correct view goes to the correct user at the right time.

Collaboration in VR: Collocated display. It is possible to use local or remote networks for multi-person cooperation in a shared virtual environment [5, 8, 16]; however, interactions are restricted to each individual's display instead of a single, shared display. While users each have their own display and can see what the other user is doing in that area, they do not see how their real world gestures and interactions relate to the location in the virtual scene. Therefore, interaction is always indirect. Recently, multi-person FTVR displays have been proposed for flat [12, 17], concave [15], and cylindrical displays [11] that allow collocated participants to interact together in front of the display. Similarly, in the spherical display described in this paper, the display is shared and the 3D scene is rendered with respect to the real-world location of each user so that they can agree upon where things are in the virtual 3D scene. Thus, users can point in the real-world into the virtual world and the other user can tell where they are pointing. Users are also able to use deictic gestures to indicate references, such as 'here', 'there', 'this'

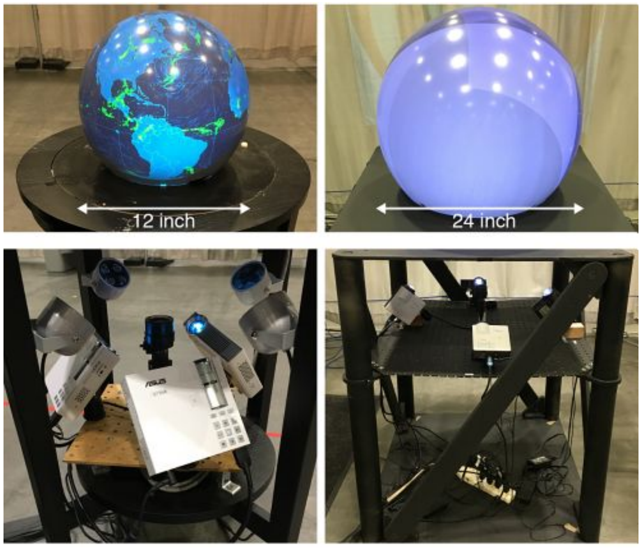

Figure 2: A 12 inch diameter spherical display built with three projectors (left) and 24 inch diameter spherical display built with four projectors setup (right). The 24 inch sphere is shown before screen calibration and therefore the projection overlap is visible, while the 12 inch sphere shows a globe visualization after calibration so that the overlap between projectors is smoothly blended together.

and 'that', and have the other user know where to look into the 3D scene. The spherical display described in this paper currently is the only volumetric FTVR display system that supports this approach to multi-person interaction while at the same time allowing for 360 degree movement around and above the virtual 3D space. Further, in the approach described here, secondary displays, such as tablets, may also be tracked. Thus, the virtual 3D scene can be rendered onto these displays enabling additional users to be collocated with the spherical display and still enable both direct and indirect interaction into the virtual scene.

3 MULTI-PERSON FTVR SPHERICAL DISPLAY SYSTEM

This section details the elements used to create the multi-person FTVR spherical display. The description includes the components for the spherical display including calibration, the techniques for supporting multiple users and the graphics rendering approach.

3.1 Spherical display

Our display uses rear-projection of multiple projectors onto a spherical surface. The inner surface of a plexiglass sphere is coated with a translucent projection paint (created by B Con Engineering, Inc.). A hole in the bottom of the sphere allows for rear-projection onto the inner surface of the sphere from projectors mounted below as seen in Figure 2. The hole size is a trade-off between projector coverage, projector placement and roundness of the final display. We found that a hole size of 75% of the diameter worked well for the projectors we use. We used two surface configurations: a 12 inch diameter sphere with a 9 inch diameter hole (Figure 2, left), and a 24 inch sphere with an 18 inch diameter hole (Figure 2, right).

Stereo projectors: For a compact design, the display uses short-throw mini-projectors. To render stereo views for a single user, and two multiplexed views for two users, the projectors require a refresh rate of at least 120Hz. Fortunately, there are a few options for small stereo projectors, such as the Optoma GT750ST (Optoma USA, Fremont, CA). Synchronization of the projectors and stereo glasses requires special consideration and is discussed in Section 3.2.

Chassis & projector placement: A simple chassis for a spherical display can be made from a table with a lower shelf. A circular hole is cut in the table top, and the projectors and camera are mounted to the lower shelf. The only requirements for projector placement are that they cover the spherical surface as much as possible and each projector overlaps its neighbor by ~10%. There are some potential issues with projecting onto a curved surface. Consumer projectors have a flat focal plane and a shallow depth-of-field; therefore the corners of the projector when projected onto the sphere can be a bit out of focus. Also, the perceived light intensity of the projectors is somewhat view dependent, which can cause variation in projector brightness as one moves around the display. In practice, we have found that these optical issues are not noticeable, particularly for scenes that do not have a bright background.

The optimal number of projectors depends on the size of the surface and the desired surface resolution. Covering the surface of the sphere with more projectors means that each projectors' visible patch on the surface will be smaller and therefore the effective resolution will be higher. We have designed two effective configurations: a 12 inch diameter sphere with three projectors (Asus P2B, 1280 x 800 resolution, 60Hz) at ~63.49 pixels per inch (PPI), and a 24 inch diameter sphere with four projectors (Optoma GT750ST, 1024x768 resolution at 120Hz) at ~34.58 PPI. PPI (without considering overlapping pixels) was computed using a Monte Carlo method with over three million samples. The multi-projector design allows the surface resolution to be scaled up by adding additional projectors. The same approach could also be used to increase brightness by increasing the overlap among projectors.

Multi-projector calibration: Creating a spherical display using multiple projectors requires calibrating all the projectors to form a coherent, spherical rendering that blends the brightness of each pixel to ensure the overlapping areas are invisible. To do this, we adopt the camera-based multi-projector calibration approach proposed by [28]. As pictured in Figure 2 and diagrammed in Figure 3, a single camera is placed at the bottom of the display to observe projected patterns from each projector. The 3D position of projected patterns are triangulated by coupling each projector with the same camera as a stereo pair. This process is repeated for each pair and the projected features are used to recover the sphere pose. These recovered parameters are further refined using a nonlinear optimization to minimize re-projection error. Finally, the 3D position of each pixel is recovered on the display surface using ray-sphere intersection. The automatic calibration approach can achieve sub-millimeter accuracy on the 12 inch sphere, and 1-2 millimeter accuracy on the 24 inch sphere. A visual depiction of the calibration results are provided in Figure 4.

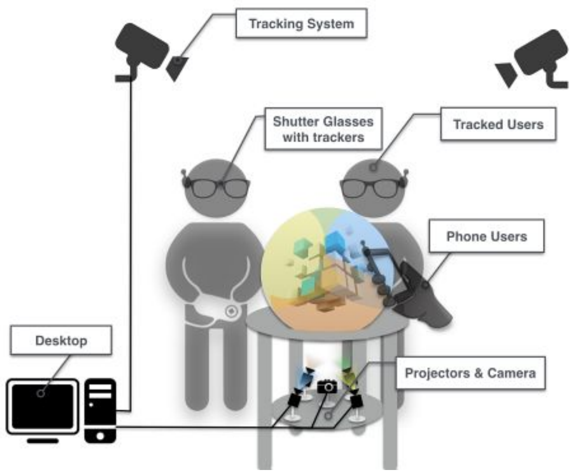

Figure 3: Overview of our system illustrating the workstation, tracking system, multiple projector spherical display, two tracked participants using shutter glasses to receive perspective-corrected views, and an additional participant using a tracked mobile screen to get a perspective-corrected view on the mobile device.

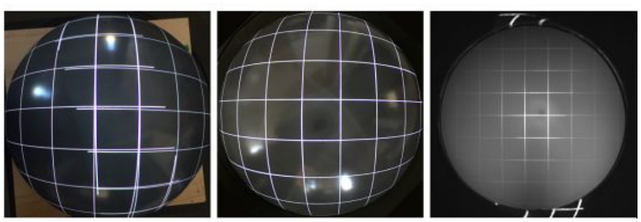

Figure 4: Top view of the spherical display before multi-projector calibration (left) showing that projected grid lines from different projectors are misaligned. Top view after the calibration procedure (middle) showing that the projected grid lines from different projectors are now well aligned and appear like continuous lines. View from the calibration camera below the sphere (right) showing that the grid pattern is correctly calibrated and rectified for its perspective.

Computing: The primary computing requirement of the system is to simultaneously output to multiple projectors. A non-stereo spherical display can be run using a video splitter connected to a high-resolution output. A stereo spherical display requires a workstation that can output frame-synchronized video. An NVIDIA Quadro P6000 graphics card was used to output and synchronize four projectors for the large stereo sphere, but multiple Quadro cards could be frame-synchronized up to 16 outputs.

3.2 Multi-view FTVR

Multi-view support: To support multiple views, we used active shutter glasses for our setup that are controlled wirelessly with a radio-frequency (RF) signal (XPAND 3D Glasses Lite RF, XPANDVISION, Beaverton, OR). For the single-person mode, the active stereo glasses allow us to alternate between left-eye and right-eye views to provide stereoscopic rendering. For the two-person mode, rather than shuttering the left and right eyes of a single pair of glasses, we shutter both eyes between two pairs of glasses. By time multiplexing views with shutter glasses, we provide a perspective-corrected view into the spherical display for each viewer, even when both viewers are standing close together. An observer would see both rendered views overlapping (Figure 1, middle), but, for the active players, the opposite player's view is filtered out by the shutter glasses (Figure 1, middle insets). The capability for two people to experience and collaborate within the same virtual scene and display is, to our knowledge, unique to our spherical FTVR setup. By using 120Hz projectors, we are limited to providing 60Hz non-stereo views to each person in two-player mode, but higher refresh-rate and brighter projectors would allow us to multiplex more viewpoints, or provide stereoscopy to multiple viewers. It also may be possible to use circularly passive polarized stereo glasses; however, we have not tested whether the spherical surface transmits the polarization sufficiently well. We also support tracked mobile screens, using compatible Android devices, so that additional participants can view into the 3D scene within the spherical display in an AR type experience.

Stereo synchronization. The NVIDIA Quadro graphics card generates a hardware synchronization signal which is used to control the stereo glasses. Typically, this sync signal is used as input to frame-synchronized projectors to ensure that the projectors' vertical-sync precisely matches the glasses shutter in order to avoid disruptive "cross-talk" between left and right views. Our inexpensive mini-projectors, while able to refresh at 120Hz, do not have the hardware frame synchronization input that is found on much more expensive stereo theater projectors. We found through experimentation, however, that the frame generation between projectors of the same model is very consistent, and therefore the sync signal from the graphics card only needs to be time delayed to align with one projector's frame refresh and no cross-talk is observable. We measured the projectors refresh timing using an oscilloscope and tuned the sync signal delay using a simple Arduino microcontroller. For the Optoma projectors, we used a 7.683 μs delay to synchronize with the graphics card.

Head tracking: FTVR requires tracking a user's viewpoint relative to the display screens. This could be accomplished by eye or face tracking from video, but for robust and low latency tracking it is often done using marker-based head tracking. We used the OptiTrack (NaturalPoint Inc., Corvallis, OR) optical tracking system with passive markers attached to the stereo glasses for head tracking and active markers on handheld wands for manipulating virtual pointers and objects within the display.

Viewpoint calibration: When using head-tracking, the location and orientation of the display relative to the tracking coordinate system and the participant's actual viewpoint relative to the head marker must be measured or calibrated. Fiducial markers could have been used to provide these calibrations, but then the system would only be compatible with tracking systems that support fiducial markers, and the markers would need to be placed on the display surface either temporarily during a calibration (which can be often

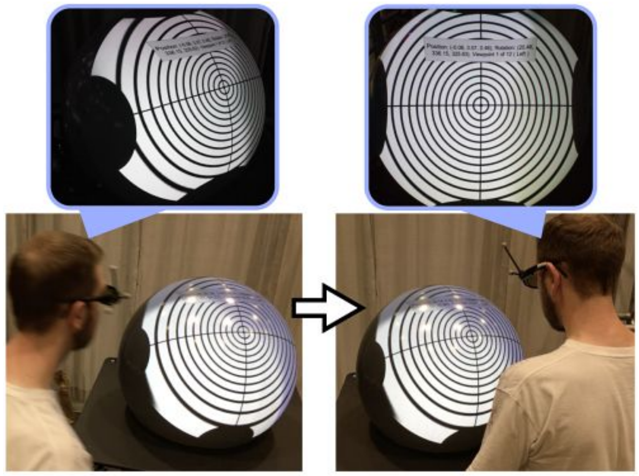

Figure 5: Our viewpoint calibration procedure: the participant starts at the wrong viewpoint (left) so the projected target pattern appears distorted from their viewpoint (left inset). The participant moves around the sphere (right), until the target pattern appears undistorted from their perspective (right inset).

if the display or tracking system is bumped) or permanently (which would obstruct the spherical display). For a calibration procedure without fiducial markers on the display, we have extended the viewpoint calibration approach proposed by [24] to work with spherical display surfaces, stereo viewing, and calibrating tracked tools, such as virtual pointers. This fast and easy-to-use procedure computes the required calibrations by using a sequence of viewpoint samples around the display. For each sample, the participant aligns their view to a "bullseye" type pattern projected on the display. Figure 5 illustrates the procedure for one sample: the participant starts in a position where the projected target pattern looks distorted (Figure 5, left with inset showing the participant's view) and then moves around the display until the target pattern appears to be undistorted (Figure 5, right). Once performed for multiple samples around the sphere, this visual alignment procedure provides us with a set of correspondences between the user's ideal viewpoint and the measured head position from which we can calibrate the display and head-tracker. The same calibration approach can be used when calibrating for monocular or binocular viewpoints by interleaving the intended viewpoints while collecting samples. Our viewpoint calibration approach can quickly compute a viewpoint registration with average angular error of less than one degree.

3.3 Rendering

We have developed a general-purpose rendering engine for multi-screen FVTR displays. The rendering system is built with the Unity game engine (Unity Technologies, San Francisco, CA). It supports a variety of display configurations, including multiple flat screens, multiple projectors on spherical or curved surfaces and virtual representations of FTVR displays. Stereoscopic rendering is supported, as well as streaming perspective-corrected views to auxiliary screens.

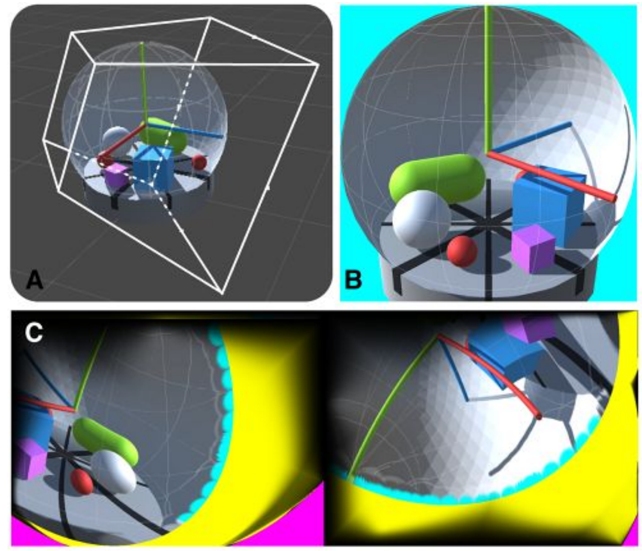

Figure 6: Overview of our rendering pipeline. We generate a view frustum for the user's viewpoint of the scene (A) and render to an off-screen texture (B). The mapping from projector-space to sphere-space is used to non-uniformly sample the rendered texture to generate the pre-warped image for each projector (C, two projectors shown). For illustration, the pre-warped image (C) is colored to show different projector regions, including: magenta regions that are not visible on the spherical surface (because they do not pass through the bottom hole of the sphere), yellow regions that are visible on the spherical surface, but not from the user's current viewpoint, and black regions that are alpha blended for a seamless transition in the overlap between projectors.

Rendering engine: Since FTVR displays mix perspective-corrected virtual imagery on a display with the real-world perspective cues of the background around the display, any mismatches between these perspective cues are disruptive to the 3D effect. For cubic displays, perspective mismatches are perceived as sharp kinks and discontinuities across the flat sides of the display. For spherical displays, mismatches result in a more subtle effect: an overall warping of the 3D scene. Empirically, the warping effect is more prominent in stereo viewing and can be disruptive to the 3D illusion.

Perspective mismatches have two main causes. Firstly, static perspective mismatches can occur because of the miscalibration of the viewer's viewpoint to the displays. These are mitigated by the viewpoint calibration methods described in Section 3.2. Secondly, dynamic mismatches occur due to latency between the tracking of a user's viewpoint and the rendering of the 3D scene. For this reason, low latency head-tracking and efficient rendering pathways are a critical aspect of FTVR displays.

As illustrated in Figure 6, for projector-based rendering, we use a render-to-texture pass and texture sampling to generate perspective-corrected imagery across multiple screens. A camera is placed in the scene at the tracked viewpoint of the user(s) and is rendered-to-texture with a normal rendering pass. For stereo or multi-viewer mono rendering, there is an additional pass for the additional viewpoint. The texture is then sampled in the mosaic projection shader to generate per-pixel color for each projector. For flat screens, e.g. on a cubic display, the homography for each screen is determined implicitly using an off-axis projection matrix for each screen and updated each frame based on the tracked viewpoint(s) of the user. For a spherical display, multiple mappings are used to compute the correct pixel color. Pixel locations in projector-space are transformed to the sphere surface location using the projector-to-sphere map computed during projector calibration (as described in Section 3.1). Sphere-space locations are transformed the viewpoint camera's clip space with the typical `ModelViewProjection` matrix. These locations are then used to sample the rendered texture saved from the last rendering pass. The per-screen texture sampling is efficient as it does not involve 3D rendering, just 2D texture-space sampling.

Mobile views: Additional views of the virtual environment can be seen on mobile screens. This is accomplished by streaming rendered frames to external devices over the network. Network messages are compressed using LZ4 and the frames are compressed as a JPEG image. Streaming in this fashion allows for lightweight mobile clients while limiting network traffic, rather than requiring the mobile devices to have a local copy of the 3D scene that has to be synchronized. Presently, we install a Unity application on the mobile device that features image streaming, remote menu controls, and using the device camera for pattern-based viewpoint calibration. The additional rendering on the workstation side increases the load on the graphics card, but we use an experimental feature of Unity called `AsyncGPUReadback`. This feature allows rendering to a mobile client at a lower priority than the main rendering loop. We are able to stream 720p frames at 30Hz to a wireless client.

We take advantage of the mobile display rendering by also tracking the mobile displays in the same way viewers are tracked. By doing so, the rendered image provides a window into the 3D scene so that the mobile display user has a view into the same 3D space as those directly looking at the spherical display. There are two different modes that are consistent with the multi-person experience using a mobile display. First, the 3D scene can be rendered as if from the viewpoint of the mobile display, providing a window into the scene. The user can then move the display around to look inside and from different angles. Second, the user's position can be tracked as well so that the image on the mobile display is rendered from the user's viewpoint through the display. The user can then move around the display to change the viewpoint and orient the screen between their eyes and the 3D scene. Both are useful depending upon the multi-person interaction appropriate for the application.

The mobile Unity application includes a menu for the system so that demos can be switched via a mobile screen and application parameters can be easily adjusted without using the workstation, which is convenient in an exhibition or user study setting. Using the mobile display as an input device provides a high fidelity interaction device that can be used to interact with the contents in the display.

Image Mosaicing: We use NVIDIA Mosaic technology to unify the image of the projectors into one screen. We use an Nx1 typology (where N is the number of projectors) to represent the circular arrangement of the projectors under the spherical display. Importantly, NVIDIA Mosaic synchronizes all screens in resolution and framerate for stereo rendering. Each projector has a resolution of 1024x768, making a total Mosaic resolution of 4096x768.

Figure 7: The player's viewpoint of the collaborative "find the fish" game when the players are pointing their magenta and cyan virtual lasers at different fish (left) and when they work together to target and capture the same fish (right).

Tracking data: We provide interfaces for a number of different tracking systems, including the Polhemus Fastrak, Liberty, G4, Patriot, optical tracking systems (OptiTrack Motive), Kinect head-tracking, and a virtual tracking system implemented directly in the Unity scene. The tracking system abstracts these trackers into a component that has a calibration for the display's position and orientation and manages a number of tracked objects in the virtual environment. These tracked objects have a position, orientation, and a viewpoint calibration. The calibrations are used to transform the object from tracking space to display space and to correct any translational or rotational offset between the tracked point and the viewpoint. The system supports both 3 DoF (position tracking only) and 6 DoF (position and orientation tracking) tracking systems.

4 EVALUATION & DISCUSSION

The system was demonstrated at the ACM SIGGRAPH Emerging Technologies exhibition in August 2018 over four days for over 1000 participants [27]. Typically, each pair of participants interacted with the system between 2 and 10 minutes depending upon how quickly they completed the tasks or spent time asking questions about the exhibition. We allowed each participant to first visualize a 3D fish bowl scene individually with stereo-glasses (Figure 1, left). We then had both participants view the fish bowl scene together using shutter-glasses set to binocular non-stereo mode (Figure 1, middle). The participants were given tracked wands that worked as virtual laser pointers. As FTVR supports real-world objects being situated in the same space as the 3D images, the laser beams appear as if they extend from the wand into the scene. Each laser beam had a distinct color (Figure 7). Participants were given a two-person collaborative 3D pointing task to hit a target that they agree upon at the same time. The target is presented as a white colored fish (the scene included three white fish and nine goldfish) which is animated to hide from the players' view by moving to regions in the scene with low visibility. We programmed the fish to hide to encourage players to move around the display while collaborating

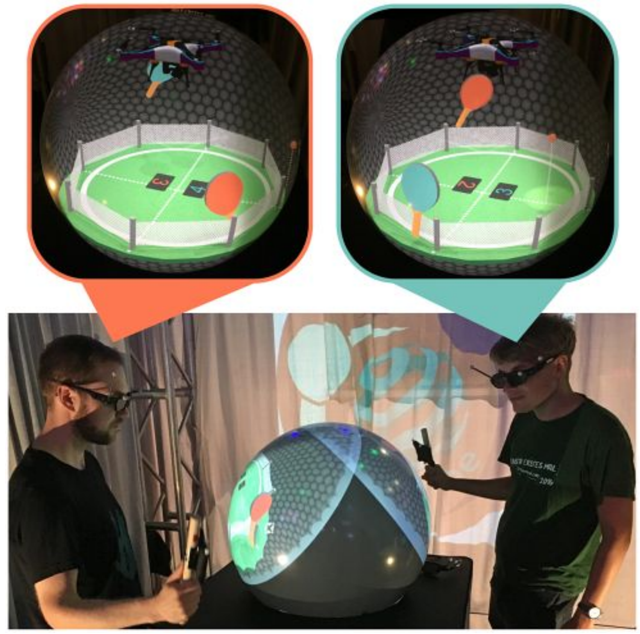

Figure 8: Two people playing the competitive ping-pong game on our system. The observers view shows distorted and overlapping rendering on the spherical display (bottom), but each participants gets their own perspective-corrected view into the game (top).

to point at the same target. When one person hit the target fish with their virtual laser beam it would slow down, but the target was not selected until both players' lasers were held on the target for 1.25 seconds at which time the fish turned into bubbles and disappeared (Figure 7, right). After finding all three target fish, a "game complete" message was displayed in the 3D scene.

Following the collaborative find-the-fish activity, approximately 10% of participant pairs (due to time constraints at the exhibit) were also invited to play a competitive ping-pong game (Figure 8). Again participants wore shutter-glasses in binocular non-stereo mode. In this game, the tracked wands were used as the handles of virtual ping-pong paddles that moved along the surface of the sphere. Participants were instructed to "block" the ping-pong ball as it bounced back-and-forth between the participants over a circular ping-pong table inside the spherical display. The orientation of the virtual paddles were constrained to face toward the center of the table to make the game easier to play, but at the same time the physics-based ping-pong ball bounces forced the players to move around the display in order to block the ball. A white target was rendered on the surface of the spherical display as a hint regarding where participants should place their paddle to correctly block the ball when it was moving toward them.

During busy times at the exhibit, when there was a line forming in front of our booth, we were able to show waiting participants the mobile display, allowing them to get a glimpse into the current experience while waiting for their turn. We offered several different views including static cameras, a stream of each user's current view as well as tracking the mobile display itself and introducing it as an additional viewpoint.

During the exhibit, three researchers helped to run the demonstration and observed participants' interactions and verbal comments while playing. While not a formal study, we summarize the general observations about participant experience with the collaborative spherical display at this exhibition.

General observations for 3D effect.

- The single-person 3D effect with perspective and stereo on a spherical surface was very strong. Many participants made an audible comment when first trying the display, e.g. "oh wow."
- The 3D effect was noticeably better for stereo as compared to binocular non-stereo mode. Approximately 20% of participants noted a preference for the stereo-mode. Some complained when transitioning to the two player mode that the scene looked more flat. We expect this is the case because the users are interacting at close proximity to the display, therefore the stereo disparity is strong.
- The vast majority of users got the concept of moving their viewpoint around without additional explanation from the experimenters either by observing previous users do the experience or by figuring it out themselves. The fish bowl scene had various objects for the fish to swim around that also encouraged people to move around the display to see where the different fish were.
- Participants reported that the motion parallax effect was strong. Many participants played with the motion parallax by intentionally moving unnaturally in order to test our system's reaction.
- The motion parallax effect was perceived differently by participants. Around 60% deemed the effect to be really good and natural. However around 25% actively criticized that they noticed a *floating* effect for virtual objects. Two potential causes for this are expected. First, we did not have time to run each participant through the personalized viewpoint calibration procedure, thus, there is some discrepancy between their actual viewpoint and the one estimated from the tracking markers attached to the shutter glasses. Second, when in stereo mode, static images appear to bend when a user moves. Thus, when tracking a participant, latency in updating the display when they move causes a similar bending phenomenon. With a 60 Hz update rate and an approximately 8 msec tracking latency, we expect our total latency is between 10-20 msec. Lower latency tracking and higher frame rate projectors would help to mitigate this effect.

Collaborative 3D pointing task (find-the-fish game).

- Participants immediately began shining their lasers on a target of their choice. Once they realized that they both needed to hit the same target, they began to negotiate the selection of the common target.
- To accomplish the task, participants would say directions out loud, such as, "Come look over here," and, "Look at the fish near the pink clam shell."

Figure 9: Two participants standing in close proximity while playing the "find the fish" game together. The experimenter (left side) is observing their communication and interactions.

- Non-verbal communication included participants pointing into or onto the sphere with their fingers, waving their virtual laser pointer and using gestures to indicate the direction of the movements.
- When participants attempted to point at a target with their finger, most (approximately 70-80%) pointed at the target shown on the surface of the display rather than pointing in the direction of its actual 3D position; instead of being a helpful hint, this behavior often confused the second user.
- Participants moved around the display throughout the game, often starting opposite their partner, but moving close together to gain similar perspectives for better pinpointing the target. A typical scene from the game is shown in Figure 9. When the fish were hard to find, players moved away from their partner at times in order to find a good secondary vantage point, so both lasers could intersect the target from different angles.
- Both the laser-pointer based interaction and head-tracking were intuitive; participants needed minimal instruction before they began moving around the display and changing their viewpoint, even crouching, to search for the virtual fish. Some participants (approximately 10-20%) figured out the interaction methods without any instruction from the researchers.
- The fish-finding task was an enjoyable experience, with many participants showing audible excitement, or even a high-five, upon catching a fish.
- Wand miscalibration was noticeable, with approximately 20% of participants complaining about the physical wand and virtual laser pointer not being aligned. Potential causes for misalignment include the monoscopic view used in the two-person mode, and misalignment during the device viewpoint calibration. New device calibration methods that account for participant viewpoint(s) are needed as future work.
- In monoscopic viewing, around five percent of participants were unable to judge the direction of the laser pointer. Although we added a shining effect to the end of the laser pointer, it was still unclear where the laser pointer started and ended in the display.

Competitive object tracking task (ping-pong game).

- Overall players found the ping-pong game less enjoyable than the find-the-fish game. Some potential explanations for this are discussed here.
- Binocular non-stereo cues were insufficient to provide depth cues for a small ball.
- Despite shadows and other cues, it was difficult to judge the ball's trajectory, particularly when it was moving toward the participant.
- Participants seemed to use the white target "hint" for where to place their paddle, rather than looking at the ping-pong ball.
- We had to slow down the ball velocity to make the game playable, but this resulted in slow gameplay that was not very exciting.
- We informally tested playing the game in one-person stereo mode against an AI opponent, which made the game better and able to be played with a faster ball speed, which would make the game more fun.
- Players moved around the display, but adjusted to a spot where they could reach most positions without a lot of additional body movement.

Mobile display.

- Participants were expecting the mobile view to be static and articulated surprise when they found out that it was viewpoint dependent.
- There were also options to directly stream a user's viewpoint to the mobile screen instead of using the screen's position. Participants generally said that they prefer the tracked display mode over streaming a user's view onto the mobile screen.

5 CONCLUSIONS

The resurgence of headset VR in recent years has renewed interest in collaborative VR experiences. Collocation within the physical world affords many tangible and intangible benefits for collaboration. Headsets, and VR versions in particular, impede feelings of real-world connectedness and situatedness between people as they share a virtual space. Volumetric FTVR displays, such as spherical displays, have better potential to merge with the real-world environment in order to create a true MR experience, but, to date, have been limited to a single person. Our new FTVR system brings together a number of technologies, including projectors, display surfaces, cameras, and 3D rendering, in order to realize a high-fidelity experience for two people interacting within the same virtual reality space, which is also situated within the real-world in order to better communicate and share the experience together. While each component of the system is relatively inexpensive, making the system easy to replicate, the tight integration of these components is required to make a believable and effective collocated MR experience. We have demonstrated the effectiveness of our design and system integration efforts by having many people use the system. We observed generally positive experiences, particularly for a collaborative game experience within a shared virtual fish bowl. Of all MR technologies under active research and development, we believe that spherical displays have the best potential to provide the visual illusion of a virtual scene merged indistinguishably with the real environment. Our present system takes a large step towards this realization. Future development, including ultra low latency and markerless viewpoint tracking, together with full stereo viewing for multiple participants, could bridge the perceptual barrier to provide a truly magical crystal ball experience.

ACKNOWLEDGMENTS

This research was undertaken thanks in part to funding from the Natural Sciences and Engineering Research Council of Canada. We also thank Brian Creber and the B-Con Engineering team for their contributions to the development of the system.

REFERENCES

[1] Kevin W Arthur, Kellogg S Booth, and Colin Ware. 1993. Evaluating 3d task performance for fish tank virtual worlds. *ACM Transactions on Information Systems (TOIS)* 11, 3 (1993), 239–265.

[2] OR Belloc, MR Nagamura, D Fonseca, André Rodrigues, DAR Souza, Celso Setsuo Kurashima, MM Almeida, Eduardo Zilles Borba, RD Lopes, and Marcelo Knörich Zuffo. 2017. OrbeVR: a handheld convex spherical virtual reality display. In *ACM SIGGRAPH 2017 Emerging Technologies*. ACM, 19.

[3] Hrvoje Benko, Andrew D Wilson, and Ravin Balakrishnan. 2008. Sphere: multi-touch interactions on a spherical display. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*. ACM, 77–86.

[4] Francois Berard and Thibault Louis. 2017. The object inside: Assessing 3d examination with a spherical handheld perspective-corrected display. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 4396–4404.

[5] Martin Böcker, Detlef Runde, and Lothar Mühlbach. 1995. On the reproduction of motion parallax in videocommunications. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 39. SAGE Publications Sage CA: Los Angeles, CA, 198–202.

[6] John Bolton, Kibum Kim, and Roel Vertegaal. 2011. Snowglobe: a spherical fish-tank vr display. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1159–1164.

[7] Jorge H dos S Chernicharo, Kazuki Takashima, and Yoshifumi Kitamura. 2013. Seamless interaction using a portable projector in perspective corrected multi display environments. In *Proceedings of the 1st symposium on Spatial user interaction*. ACM, 25–32.

[8] Christopher Codella, Reza Jalili, Lawrence Koved, J Bryan Lewis, Daniel T Ling, James S Lipscomb, David A Rabenhorst, Chu P Wang, Alan Norton, Paula Sweeney, et al. 1992. Interactive simulation in a multi-person virtual world. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 329–334.

[9] JP Djajadiningrat, Gerda JF Smets, and CJ Overbeeke. 1997. Cubby: a multiscreen movement parallax display for direct manual manipulation. *Displays* 17, 3-4 (1997), 191–197.

[10] Fernando Ferreira, Marcio Cabral, Olavo Belloc, Gregor Miller, Celso Kurashima, Roseli de Deus Lopes, Ian Stavness, Junia Anacleto, Marcelo Zuffo, and Sidney Fels. 2014. Spheree: a 3D perspective-corrected interactive spherical scalable display. In *ACM SIGGRAPH 2014 Posters*. ACM, 86.

[11] Daniel Gotsch, Xujing Zhang, Timothy 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*. ACM, 522.

[12] Dongdong Guan, Chenglei Yang, Weisi Sun, Yuan Wei, Wei Gai, Yulong Bian, Juan Liu, Qianhui Sun, Siwei Zhao, and Xiangxu Meng. 2018. Two Kinds of Novel Multi-user Immersive Display Systems. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 599, 9 pages.

[13] Naoki Kawakami, Masahiko INAMI, Taro MAEDA, and Susumu TACHI. 1997. Proposal for the object-oriented display: The design and implementation of the media3. In *Proceedings of ICAT*, Vol. 97. 57–62.

[14] 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*. ACM, 2531–2540.

[15] Juan Liu, Hanchao Li, Lu Zhao, Siwei Zhao, Guowen Qi, Yulong Bian, Xiangxu Meng, and Chenglei Yang. 2018. A Method of View-Dependent Stereoscopic Projection on Curved Screen. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 623–624.

[16] Oyewole Oyekoya, William Steptoe, and Anthony Steed. 2012. SphereAvatar: A Situated Display to Represent a Remote Collaborator. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 2551–2560.

[17] 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 (2016), 138–148.

[18] Ye Pan, William Steptoe, and Anthony Steed. 2014. Comparing flat and spherical displays in a trust scenario in avatar-mediated interaction. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 1397–1406.

[19] Ramesh Raskar, Jeroen van Baar, Srinivas Rao, and Thomas Willwacher. 2003. Multi-projector imagery on curved surfaces. *Mitsubishi Electric Research Labs* (2003), 1–8.

[20] Jun Rekimoto. 1995. A vision-based head tracker for fish tank virtual reality-VR without head gear. In *Virtual Reality Annual International Symposium, 1995. Proceedings.* IEEE, 94–100.

[21] Behzad Sajadi and Aditi Majumder. 2011. Automatic registration of multi-projector domes using a single uncalibrated camera. In *Computer Graphics Forum*, Vol. 30. Wiley Online Library, 1161–1170.

[22] Ian Stavness, Billy Lam, and Sidney Fels. 2010. pCubee: a perspective-corrected handheld cubic display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1381–1390.

[23] Ian Stavness, Florian Vogt, and Sidney Fels. 2006. Cubee: a cubic 3D display for physics-based interaction. In *ACM SIGGRAPH 2006 Sketches*. ACM, 165.

[24] Andrew John Wagemakers, Dylan Brodie Fafard, and Ian Stavness. 2017. Interactive visual calibration of volumetric head-tracked 3d displays. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3943–3953.

[25] Colin Ware, Kevin Arthur, and Kellogg S Booth. 1993. Fish tank virtual reality. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*. ACM, 37–42.

[26] Colin Ware and Glenn Franck. 1996. Evaluating stereo and motion cues for visualizing information nets in three dimensions. *ACM Transactions on Graphics (TOG)* 15, 2 (1996), 121–140.

[27] Qian Zhou, Georg Hagemann, Sidney Fels, Dylan Fafard, Andrew Wagemakers, Chris Chamberlain, and Ian Stavness. 2018. Coglobe: a co-located multi-person FTVR experience. In *ACM SIGGRAPH 2018 Emerging Technologies*. ACM, 5.

[28] Qian Zhou, Gregor Miller, Kai Wu, Daniela Correa, and Sidney Fels. 2017. Automatic calibration of a multiple-projector spherical fish tank vr display. In *Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on*. IEEE, 1072–1081.

[29] Qian Zhou, Kai Wu, Gregor Miller, Ian Stavness, and Sidney Fels. 2017. 3DPS: An auto-calibrated three-dimensional perspective-corrected spherical display. In *Virtual Reality (VR), 2017 IEEE*. IEEE, 455–456.