Multi-View Lenticular Display for Group Teleconferencing

Peter Lincoln
Dept. of Computer Science
UNC-Chapel Hill
Chapel Hill, NC 27599-3175
plincoln@cs.unc.edu

Andrew Nashel
Dept. of Computer Science
UNC-Chapel Hill
Chapel Hill, NC 27599-3175
nashel@cs.unc.edu

Adrian Ilie
Dept. of Computer Science
UNC-Chapel Hill
Chapel Hill, NC 27599-3175
adyilie@cs.unc.edu

Herman Towles
Dept. of Computer Science
UNC-Chapel Hill
Chapel Hill, NC 27599-3175
herman@cs.unc.edu

Gregory Welch
Dept. of Computer Science
UNC-Chapel Hill
Chapel Hill, NC 27599-3175
welch@cs.unc.edu

Henry Fuchs
Dept. of Computer Science
UNC-Chapel Hill
Chapel Hill, NC 27599-3175
fuchs@cs.unc.edu

(a) Local participant A's view.

(b) Local participant B's view.

Figure 1: Our prototype multi-view display offers two concurrent distinct views. Using Figure 2 as reference, the remote participants are pointing towards local participant A.

ABSTRACT

We present a prototype display system for group teleconferencing that delivers the proper views to multiple local viewers in different locations. While current state-of-the-art commercial teleconferencing systems can provide high-definition video, and proper placement and scaling of remote participants, they cannot support correct eye gaze for multiple users. A display capable of providing multiple simultaneous views to each local observer and multiple aligned cameras are required for generating a distinct and spatially-appropriate image for each local participant. If each local participant can observe the remote participants from an appropriate angle, then it becomes possible for viewers to properly identify where other participants are looking.

Our system provides view-appropriate imagery for multiple users by spatially multiplexing the output of the display. We achieve this by placing a lenticular sheet over the surface of the display, which directs light from different pixels in different directions. With knowledge about the subset of the display surface each participant can see, it is possible to combine each of the remote cameras' images into a single composite image. This image, when viewed though the multiplexing layer, appears to be the appropriate camera image for each local participant. The prototype system uses a camera-based display calibration technique capable of properly evaluating which pixels are visible from an arbitrary viewpoint without a physical model of the display.

1. INTRODUCTION

1.1 Traditional Teleconferencing

People often need to communicate in order to coordinate a multi-person project. While sometimes this communication can be handled through email, instant messenger, or phone calls, certain cases require face-to-face contact. If these people do not work or live near each other, a true face-to-face meeting could incur expensive transportation costs. Tradi-
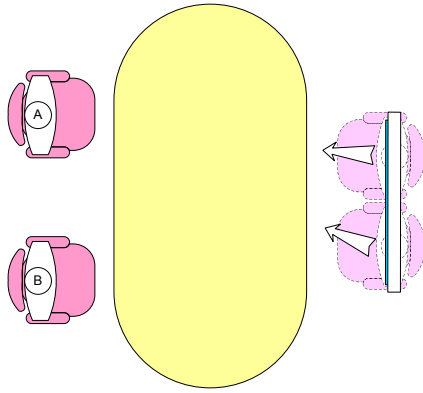
Figure 2: The prototype two-on-two teleconferencing scenario. Participants A & B are local. The other participants are remote and are illustrated by their virtual positions and gaze direction.

tional commercial video teleconferencing systems attempt to reduce costs by providing groups with remote cameras and local displays in an effort to simulate face-to-face contact without requiring users to spend time and money traveling. The image fidelity of such systems ranges quite widely; variable characteristics include display size, camera resolution, image compression, and frame rate. Low-end systems may use basic web-cams and video-instant messenger systems. An example high-end video teleconferencing system is the Tandberg Telepresence T3 system[1], which focuses on an attempt to maximize presence: the feeling that the remote and local participants are in the same room.

1.2 Spatial Awareness and Presence

Commercial video teleconferencing systems, from low- to high-end, typically have one camera per display. In purely one-on-one group situations, it is quite clear that each participant is directing his or her comments at the only other participant; as a result, this limitation in view count may not be much of an issue for clarity of intent. As the number of participants increases, however, spatial direction of comments and eye-contact can be adversely affected by the single camera per display setup, which reduces the sense of presence. Regardless of the number of participants, the level of presence can be affected by several issues: virtual size, image resolution, and gaze direction.

1.2.1 Virtual Size and Position

Virtual size correctness represents the degree to which a remote participant appears life-size. This is least well represented by instant messenger systems such as AIM, MSN, or Skype, in which the remote participant's image is projected into a small window on a user's desktop. Low-end commercial systems, although they may utilize a complete or large display, similarly may not accurately represent a remote user at life-size scales. High-end systems can use large displays and well positioned cameras to permit displaying a remote user at life-size. Additionally, these systems often

use local and remote scenery in order to provide visual continuity between the local and remote rooms. The Tandberg T3 Telepresence system, for example, maintains an illusion that the local and remote participants are sitting at a large conference table. By providing such context, these teleconferencing systems are able to maintain a degree of presence.

1.2.2 Image Resolution

Higher resolution displays are useful for reading remote text or observing virtually distant objects, but they can be especially useful for evaluating a remote participant's pupillary gaze direction. In a user study, Nguyen and Canny [4] evaluated gaze direction in a multi-view teleconferencing system, expecting that by staring directly into the remote cameras, local participants would believe that they were making direct eye contact. Instead, local participants found it difficult to make out the position of a remote participants pupil in the low-resolution displays; as a result, gaze directions were sometimes misevaluated. In reporting their next system [5], in which they improved both the video resolution of the cameras from CCTV to 1024×768 and the resolution of their projectors from 800×600 to 1024×768, they did not report any further resolution issues that affected gaze direction evaluation. This suggests that by meeting a minimum resolution, a sense of presence can also be met.

1.2.3 Gaze Direction

Identifying gaze direction is a significant issue in group situations. Consider a situation consisting of four people sitting around a conference table with two on each side. Each person can tell where a speaker is looking or pointing. Suppose instead, two participants are replaced by a standard video display, as in Figure 2, which presents a single camera view of these remote participants. If this camera is placed above the equivalent remote display showing A and B, then this situation of four people represents a standard one display per camera, two-on-two teleconferencing setup. Existing higher-end systems place the camera as vertically close to the virtual position of remote participants as possible; while this may cover part of the display, it would not cover any part of a remote participant's head. In this way, the vertical eye-gaze problem can be alleviated in a low impact manner. However, a horizontal eye-gaze problem persists.

Consider the horizontal gaze problem in the two-on-two scenario with a single camera per display: suppose a remote participant looks at position A. Participant B may believe they are looking at A since the remote user would appear to be looking to the B's left; however, participant A would also see them looking to their left. If a remote user looks directly at the camera, both A and B would think the remote user is looking directly at them. To resolve this issue, each local participant must receive a unique view of the remote participants; this introduces the need for a multi-camera per display configuration. If a collection of remote cameras represents each local participant, and each local participant can receive a spatially-correct view of the remote participants, then gaze direction can be correctly determined.

2. SYSTEM DESIGN

To provide improved spatial awareness, each local viewer should see a view-appropriate, life-size proxy of each remote participant. It is also important that participants not be encumbered or obscured by wearing special equipment. Partic-
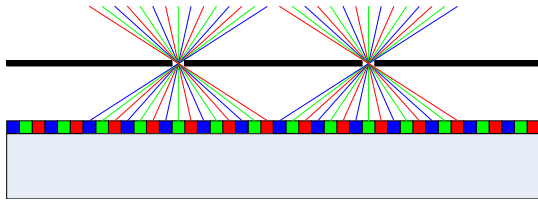

Figure 3: Parallax barriers spatially multiplex a display limiting light from the display to pass through particular holes, thus limiting the possible viewing direction.

ipants should also be able to sit comfortably at appropriate distances among local and remote users; this requires that the individual views be appropriately spaced apart and that the display surface be sufficiently close.

2.1 Multi-View Technique Selection

Since the design prohibits view-assistive wearable equipment, light polarization [1] or time-division multiplexing techniques, which would require users to wear polarizing lenses or shutter glasses, are ruled out. Autonomous multi-view technologies that do not require such encumbrances include retroreflective, parallax barrier, and lenticular lens displays.

2.1.1 Retroreflective Displays

Retroreflective displays consist of two parts: a front screen projector and a retroreflective projection screen. While the projector can be a standard projector, this type of screen behaves significantly different from a normal projection screen. A standard front screen projection screen diffuses incoming light in all directions; this diffusion reduces glare and permits the image to be viewed from any reasonable direction. An ideal retroreflective screen reflects incoming light back along the same direction it came; in ray tracing terms, the angle of incidence, θ_i, is equal to the angle of reflectance, θ_r. This technology has been used to implement research teleconferencing systems [4, 5]. To control view positions, the design must simply place a projector as close as possible to each local participant's head. Because each participant uses a personal projector, the full resolution of a projector can be solely utilized by that participant; no spatial multiplexing is required. The degree by which each participant's view conflicts with other participants is simply controlled by the quality of the projection screen; the more it behaves like an ideal retroreflector, the less potential for crosstalk between views. The main disadvantage of a system of this type is brightness; front-screen projection requires a moderately dim surrounding environment as compared to direct-view LCD or CRT displays.

2.1.2 Parallax Barrier Displays

Parallax barrier displays can make use of LCD or CRT monitors as direct-view displays. A parallax barrier consists of a optically opaque film with arranged holes or slits which permit light to pass [2]; this is illustrated in Figure 3. Alternatively, the film can be replaced by a LCD layer which can variably pass or block light across the surface; the Varrier™ system [6] is an example of such a display. A parallax barrier display often consists of a standard LCD display with one of these films placed in front at a particular distance. Each slit permits different subpixels, the individual red, green, and blue color components of a pixel, to be visible at different viewing angles. In this way, the display pixels are multiplexed among different views. If each local participant is assigned a particular view, then each participant can receive a distinct image.

While views with a retroreflective display can be somewhat arbitrarily placed, views with a parallax barrier display are controlled in a relatively uniform fashion through the design of the barrier. There are four basic parameters in a repeating barrier pattern: hole size, hole spacing, barrier distance, and slit orientation. Adjustments to the hole size control the degree of overlap between views; larger holes permit more light to pass, including light from adjacent subpixels. Increasing the hole size thus creates a brighter display at the expense of additional crosstalk; this crosstalk can take the form of ghosting, in which adjacent views become partially visible in the target view. Hole spacing controls the number of views a display can support; increasing the spacing increases the number of supported views. Since there is a limited number of pixels in the monitor, increasing the number of views decreases both the effective resolution of a view and the overall brightness of the display. Adjusting the barrier distance controls the spread of a set of non-repeating views; moving the barrier closer widens the total viewing angle. Adjusting the slit orientation, by rotating the sheet such that a single slit spreads over multiple columns, corrects for the structure of LCD panels. LCDs are usually composed of repeating columns of a single color; a single column of pixels is composed of three columns of subpixels: red, green, and blue. Patterns between correspondences in vertical slits and the subpixel columns can often result in visible color interference patterns in the form of beating. By instead rotating the barrier, a single view through a slit covers multiple subpixel columns and thus all three colors will be integrated in the view. Rotating the slits also has the benefit shifting some of the loss of horizontal resolution into a lesser loss of horizontal and vertical resolution [3].

It is useful to note that a parallax barrier display repeats views. It is possible, if a user is sufficiently left or right of the display, for a viewer to see through an adjacent slit to the primary slit and thus see the same subpixel through a different slit. As a result, a viewer sufficiently out of position may see an inappropriate view. Using Figure 2 as an example, a participant located to A's left may see B's view.

2.1.3 Lenticular Lens Displays

Lenticular lens displays, while structurally different from parallax barrier displays, operate in a similar fashion optically to parallax barrier displays. Lenticular lenses are a sheets composed of many adjacent lenticules, which are typically long narrow convex-planar lenses. Lenticular lens displays are typically composed of a standard LCD with a lenticular lens sheet placed over the display. Each lenticule directs the light from a single subpixel towards a particular direction; this is illustrated in Figure 4. Multiple subpixels under a single lenticule are directed in different directions; in this way, multiple views are formed.

Lenticular lens displays are tunable in a manner similar to parallax barrier displays. There are four basic parameters in a uniform lenticular lens sheet: lens width, lens radius, backing sheet thickness, and lens orientation. Adjusting the lens width controls the number of views the display can support; roughly the number of subpixels in a row under a sin-
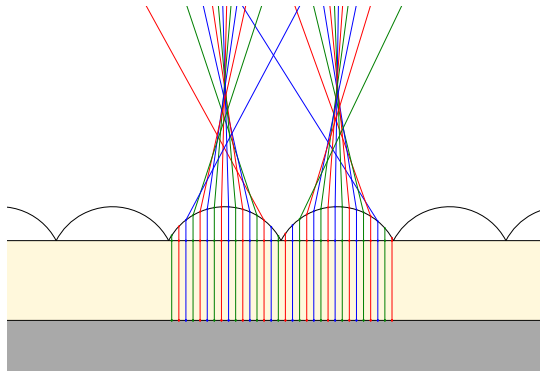
Figure 4: Lenticular lenses spatially multiplex a display by directing the light from different subpixels in different directions.

gle lenticule controls the maximal number of views before a repeat. Unlike a parallax barrier, increasing the number of views does not drastically reduce overall brightness; instead of allowing a subpixel to emanate light in all directions, a lenticule focuses the outgoing light into a subset of directions. However, like the parallax barrier, increasing the number of views decreases the spatial resolution of each view. Adjusting both the radius and the backing sheet thickness control both the total non-repeating view angle and the degree at which each view overlaps with adjacent views. Because lenses focus light optimally at a particular distance, one could achieve minimal crosstalk if the distance between the lens and the pixel surface is at the focal distance. However, due to manufacturing tolerances, namely the difficulty in producing the sharp valley between lenses associated with small radii and small width, it may become necessary to select a non-optimal thickness in order to achieve a desired total non-repeating view angle. Adjusting the radius, in order to control the total view angle, requires a corresponding change in thickness to account for the change in focal distance to avoid creating crosstalk. Decreasing the thickness, without adjusting the radius, can both increase crosstalk and increase the total non-repeating view angle. Rotating the lens corrects for color interference patterns in a similar manner to rotating a parallax barrier slit.

2.2 Display Configuration

Out current prototype system uses a lenticular lens multiview technique. Using a direct-view display avoids the brightness issues inherent to front screen projection systems, like the retroreflective technique. Furthermore, using lenticular lenses instead of a parallax barrier avoids both the black space between slits and darker image resulting from absorbing light from subpixels that a particular view does not include.

Given the selection of a lenticular lens display, it becomes necessary to define the parameters of the monitor and lenticular lens. The desired design plan for the prototype system should support at least two local participants and at two remote participants per display; this setup is illustrated in Figure 2. To achieve a comfortable and life-size experience, each display should be sufficiently large to support a virtual view of two people sitting side-by-side. Furthermore, the individual viewpoints should be spread out such that each local user can sit comfortably side-by-side. Rotating

the displays helps to ensure that the center of each display's set of views remains in the center of the local participants; it also follows the curve of table, which contributes towards the scenery aspect of presence. The local and remote participants should appear to be sitting at a distance equivalent to the width across a conference table; this has been defined to be about 1.8 meters.

2.2.1 Commercial Autostereoscopic Displays

Several companies, including NewSight Corporation[2] and Philips[3], produce and sell autostereoscopic displays. Each of these aforementioned companies produce displays sized in excess of 40" along the diagonal, which when viewed at the appropriate distance can appear to generate 3D stereo imagery. A stereo effect is generated through the use of a parallax barrier (NewSight) or lenticular lens (Philips WowVx). The effect results from a design which generates views that are spatially separated by the interpupillary distance (IPD), which is about 6.45 cm, at the designed optimal distance, which is typically in excess of 3m for large displays. Furthermore, to reduce sharp transitions between adjacent views, a non-trivial amount of crosstalk is introduced.

To evaluate a commercial system, a NewSight 40" display was used. Newsight supplies a calibration tool which allows a user to adjust the software masks in order to bring the viewing distance closer to the display. While it is possible to still achieve an autostereoscopic effect at the closer distance, teleconferencing results in severe limitations. At 1.8 meters, the eight views repeat over a distance of about two IPDs. As a result, in order to receive a single view, a user would be required to hold his or her head still and close one eye. It remains possible to separate users comfortably by using the repeat inherent to a parallax barrier display. Furthermore, due to the physical blending introduced by the barrier, a maximum of only two views was achieved using this system. Based on this result, a custom solution that provides wider views at a closer distance appears to be necessary.

2.2.2 Custom Lenticular Display

Given that a lenticular lens display is desired for the current prototype's design, it is necessary to select a base monitor of sufficient size and resolution to support the goals. We use a 47" 1080p-resolution HDTV display, which provides a 41"×23" viewing area. MicroLens Technology, Inc.[4], of Indian Trail, NC, USA, a manufacturer of lenticular lenses, has prior experience developing lenses for this displays of this scale. As shown in Figure 2, A pair of these LCD monitors display the remote users; at most two remote users can appear on a single monitor. Both the head and upper torso of a remote participant can be seen at life-size at the desired distance.

An optimal lens, designed to handle a single situation, would be shaped to contain only as many distinct views as local participants; maximizing each views resolution would require a lens width sufficient to cover the same number of subpixels as views. In order to minimize beating, which is a color fringing effect produced by slight differences in lenticule and pixel widths, the lens angle should be at least sufficient to present a unique subpixel color for each three rows

---

[2]http://www.newsight.com

[3]http://www.business-sites.philips.com/3dsolutions/about/Index.html
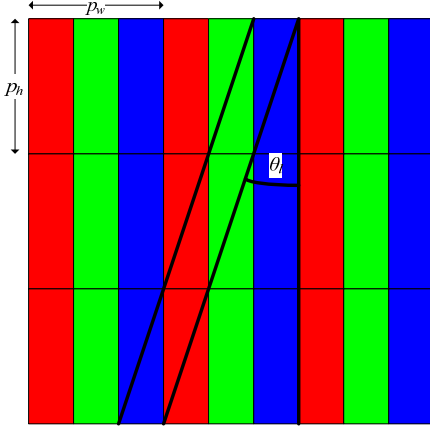
[4]http://www.microlens.com

Figure 5: Diagram illustrating the subpixel structure of LCD panels. Columns of pixels are composed of three columns of subpixels. The pair of parallel, diagonal lines represents a participant's view of the display.

of pixels. For an arbitrary pixel size, this angle can be computed as shown in Equation 1 and Figure 5; for a square pixel, this angle is about 18°. For comfort reasons, local participants would be spaced about 85 cm apart center-to-center; thus the radius and thickness of the lens would also need to reflect this condition.

$$\theta_l = \arctan \frac{p_w}{3 \cdot p_h} \qquad (1)$$

The specification of the design of the prototype lens favors flexibility over resolution. Dynamically permitting either two or three local participants at comfortable spacing requires that the local participants be in completely different locations between the two setups. Since individual views generated by a lenticular display tend to be equally spaced within the primary non-repeating view area, more than two, three, or five views are required. As designed, the prototype lenses are intended to support up to fifteen views, spread over two meters at 1.8 meters from the display; this implies that the lenticule width is the same width as fifteen subpixels. This results in a wider lens than would be specified in the aforementioned optimal case. The lens radius and backing sheet thickness would also need to be modified to reflect the different lens width. Lens angle would need not change from the optimal case as each view would still shift one pixel horizontally per three pixels vertically.

2.3 Rendering

Rendering the images for each display involves determining what composite image is appropriate such that each local participant receives the proper view. Because the surface of the display is multiplexed among multiple local participants, generating such an image accurately requires knowledge about which subpixels are visible to each participant. These sets of view-visible pixels can be composed as masks, with which the camera images are modulated and summed. As long as the masks resolve conflicts, which are subpixels on the display visible by multiple participants, the composite image, when viewed through the lenticular lens or paral-

lax barrier, should yield distinct views to each participant. Generating these masks, however, requires calibration.

2.4 Display Calibration

The primary goal of calibrating the display is to determine which subpixels are visible from a given view point. Preliminary calibration involves positioning the display, lenticular lens, and participant viewpoints. The camera-based calibration method used to calibrate the current prototype is designed to support arbitrary spatial multiplexing techniques from static local participant viewpoints. A completed calibration produces a set of viewpoint masks, which the aforementioned rendering phase can use.

2.4.1 Lens and Participant Alignment

It is first necessary to align the lenses on the display. While each of the 15 views spread their views widely, each view remains brightest in its center. It is therefore desirable to position each of the participants in the center of one of these views. Furthermore, to minimize chances of crosstalk, it is useful to maximally spread each of the participants among the 15 views; it is not desirable for any viewer to be able to see a significant portion of another viewer's subpixels (see 2.4.4).

Lens alignment involves both tweaking the image on the display and physically moving the lens over the display. The image on the display consists of a single diagonal line composed of a single subpixel color. By rotating and shifting the lens, it is possible to direct the center view towards a particular participant. In the two view system, placing the second participant involves illuminating a parallel line spaced about half of the repeat distance horizontally away from the original line; for fifteen views, this about 7 or 8 subpixels distant. Because these prototype displays are composed of two lenticular sheets, the process is repeated for the second sheet. In this way, selecting a position for one local participant directly influences the location for the second local participant.

2.4.2 Calibration Image Capture

The image capture stage of calibration process involves the collection of images of the display from each of the participants view points. To collect these images, a high resolution, digital, grayscale camera is placed at each participant's viewpoint. Multiple techniques exist for the composition of these images.

The slowest, and most naïve, algorithm consists of illuminating a single subpixel for all subpixels in the display. If that subpixel is visible in the camera image, then that subpixel can be considered visible at the camera's viewpoint. This would require, where v is view count, and h and w is display size in pixels, $3 \cdot v \cdot h \cdot w$ pictures. With the selected HDTV display, this is over six million images per view ans so is impractical.

A much faster procedure uses Gray code, a binary encoding of a mapping between camera space and screen space. By forming a known sequence of gray code images, each subpixel on the display can be uniquely represented by a binary number. If a pixel in the camera image can be found that duplicates the binary number of a subpixel, then that subpixel can be considered visible from the camera's viewpoint. While workable in theory, there is a major problem in practice; region edges, while sharp on the display, may not map

well to the camera space, which will degrade the output mask quality; some of these issues result from the property that large illuminated areas in contrast with dark surrounding areas can cause bleeding, which further degrades the edges.

A more camera-oriented approach is line-sweeping. Each image is composed of a single horizontal or vertical line of a single subpixel color. Like Gray Coding, a mapping between camera space and screen space can be computed. If a subpixel in screen space has a corresponding illuminated pixel in camera space, then that subpixel can be considered visible at the camera's viewpoint. Unlike Gray Coding, however, line-sweeping does not fall victim to area bleeding. This technique requires a total of $3 \cdot v \cdot (h + w)$ images. Because this technique requires a reasonable number of images and the quality of said images is significantly better than Gray Coding, the selected calibration capture method for the current prototype system is line-sweeping. At current imaging speeds, a single display can be fully imaged, all three colors and two viewpoints, in about five hours; the multiple viewpoints can be imaged in parallel.

2.4.3 Image Analysis

The image analysis stage uses the set of images generated by line-sweeping to determine which pixels are visible from the designated view point. Because line-sweeping is used to generate the set of images, a mapping between camera-space and screen-space is useful. For this operation, each color channel, and each view can be considered separately. The algorithm can be summarized as shown below. The input consists of a set of grayscale camera images of the horizontal and vertical lines from a single viewpoint and a single color. The output consists of a binary image of pixel visibilities.

1. Generate an array at camera resolution for Y coordinates, M_Y, initially zero.
2. For each horizontal line image,
   (a) For each pixel, (x, y), in the image,
      i. If the grayscale value is above threshold, assign the y-value of the line to $M_y(x, y)$, overwriting any previous value.
3. Generate an array at camera resolution for X coordinates, M_X, initially zero.
4. Repeat step 2 for each vertical line image using M_X.
5. Generate a binary array at display resolution, M_d, initially zero.
6. For each element, (i, j), in arrays M_X and M_Y
   (a) Assign 1 to $M_d(M_X(i, j), M_Y(i, j))$.

It can be noted that this algorithm makes no assumptions on the model of the display or multiplexing method. Furthermore, position and orientation of the calibration camera is not required to be known. As a result, it can be utilized for any spatially-multiplexed, pixel-based display. It is, however, limited by the relationship between camera resolution and screen resolution. Suppose the calibration camera resolution is much smaller than the portion of the screen resolution visible in the images. In this case, multiple horizontal or vertical lines could map to the same camera pixels. To combat this issue, it is necessary for the camera resolution to exceed the imaged display resolution; in this way, horizontal or vertical lines can map to multiple lines in the camera image, thus reducing the chances of conflicts in the mapping. This may involve performing multiple imaging runs over multiple subsets of the display.

Tuning the calibration becomes a matter of adjusting the aperture on the calibration cameras, in an effort to control light, and adjusting the threshold, in an effort to select truly visible subpixels. Furthermore, the camera's lenses must not overly distorted, as radial distortion effects can affect the uniformity of intensity across a single image. In experimentation, it has been observed that by ensuring that the aperture is adjusted such that central green lines are slightly below saturation, then the darker red and blue lines remain sufficiently visible to properly threshold.

In the current prototype system, this algorithm is implemented using Matlab, and requires a several hours to serially process a single display for one viewpoint and one color. Because each viewpoint and color is independent, the process can trivially parallelized among multiple computers to speed up computation. Instrumentation of the implemented algorithm currently shows it to be rate-limited by Matlab's implementation of loading images from disk. However, the benefit in being able to adjust the threshold without having to take additional pictures outweighs the disk access cost.

2.4.4 Mask Combination

Given a set of raw subpixel visibility collections, it is necessary to produce a set of colored view masks corresponding to each participant. Two methods are considered: black blending and equal-weight blending. Black blending operates on the principle that if two participants can see the same subpixel, then neither should see that subpixel lit. The resulting mask is composed of purely binary cells; a subpixel in the mask is either full on or full off. This technique thus minimizes crosstalk at the expense of reduced overall brightness. Equal-weight blending instead equally averages the blending among the views that can see the subpixel. If only one view can see the subpixel then full on, if two views can see it then half on, and if no views can see the subpixel full off. This technique maximizes potential brightness at the expense of crosstalk.

In observation of the described two-on-two prototype, despite the wide overlap each of the fifteen views cast, little difference could be observed between the black blended and equal-weight blended masks. This is likely due to the wider positioning of the two participants; the center view of each composite view is sufficiently distant from the edge-visible views of the other participant's composite view.

2.5 Participant Capture and Remote Camera Calibration

To generate the output image, both the masks and remote camera imagery are required. Remote camera imagery is generated by careful placement of cameras in the remote area. For a display with two views, two remote cameras are required. Each camera represents the local viewer in the remote location. The cameras are placed apart by the distance between the local participants, and the cameras are placed sufficiently high to represent the eye-heights of a seated participant.

For a half-duplex system, it is possible to place the cameras at eye-height without consideration for obscuring the display; however, in a full-duplex system, placing the cameras in such a position would cover the faces of the displayed participants. Fortunately, placing the cameras at the top of the display will result in a small difference in gaze perception; the small vertical offset is a significantly smaller effect than

(a)

(b)

(c)

(d)

Figure 6: In this scenario, the remote participants are pointing at the local user in position B. The top row presents single-view mode, and the bottom row presents multi-view mode. The left column is viewed from position A, and the right column is viewed from B.

a corresponding horizontal offset [4].

As it is a goal of the system to produce life-size images of the remote participants, camera calibration is required. This is achieved through selection of lens power and image warping. To maximize use of the camera pixels, the selected lens should maximize the size of the a virtual display, located in a position appropriately relative to the corresponding participant and real display. Image warping is achieved through the temporary use of a checkerboard pattern placed at the same position as the aforementioned virtual display. Once corners are located, a transform can be computed to map the camera image to the pre-masked display image. The transform can be saved for later use.

3. RESULTS

Completing a calibration run and testing the renderer with the new mask immediately reveals the quality of the generated masks. Glitches can appear as misplaced viewpoints and excessive crosstalk. Misplaced viewpoints often result from operator failure during the calibration camera placement step; it can be difficult to predetermine the exact direction for the center of a view. The crosstalk problem is typically corrected by shifting the calibration cameras into the center of the current composite views and rerunning the

capture, analysis, and combination stages in order to generate new masks. Errors resulting purely from crosstalk may require more drastic measures, including physically adjusting the lens alignment, reducing the local participant count, or redesigning the lens.

After executing a successful calibration run, correctness is also immediately apparent. The results of a successful run are presented in Figures 1(a), 1(b), 6(c), and 6(d). Examining the differences between the single-camera per display and multiple-camera per display modes, as shown in Figure 6, immediately reveals the benefits of using multi-view: proper gaze correction. In multi-view mode, both participants A and B can see that the remote participants are pointing at B.

Multiple manufacturing complications present a challenge for calibration. Lenticular lenses, when created using MicroLens' UV casting method, can only be produced at a width of about 26 inches; as a result, two lenticular sheets are required to cover a single display. This required a bar to be placed in the center of the display to hold the sheets close to the surface of the display. Furthermore, the pulling process used in UV casting to generate columns of lenticules can result in uneven forces across the sheet surface, resulting in non-uniform lens widths across the lenticules; fortunately, the prototype's calibration method does not depend on uni-

form lenses. Additionally, a focus issue prevented the resultant lenses from supporting fifteen non-overlapping views. For the lens shape selected by MicroLens, in order to spread the total non-repeating view angle sufficiently wide, a non-optimal backing sheet thickness is required. As a result, multiple adjacent subpixels are visible through a single lenticule from a single point 1.8 meters from the surface. The overlap between views was sufficiently large to limit a maximum of two composite views. Some of the effects of the significant overlap between views can be observed in Figure 6(c): a small amount of crosstalk between A's and B's views is apparent when there is high contrast between the views at the same point. In this case, participant 3's black shirt contrasts with the white room background.

4. EXTENSIONS

4.1 Sparse Calibration

The current prototype's calibration method involves sampling the entire display space; however, it may be possible to devise a more sparse sampling method. Sampling over the entire display has the advantage of not requiring any model of the multiplexing method, nor does it require that the multiplexing method behave in a uniform fashion. However, both commercial autostereoscopic displays and the lenticular lens prototype system use a regular pattern, and manufacturing processes for displays tend to produce consistent results. Furthermore, the behavior of parallax barriers and lenticular lenses is well known; it is therefore possible to generate a model of both the display and the multiplexing layer. A model of this type would only have a few parameters, which may include the lateral offsets between the lens or barrier and the display and the relative orientations of each. Using this model, it may be possible to sample a subset of points or lines on the display from multiple locations to numerically determine these parameters. Such a method could decrease calibration time or serve as inputs to a dynamic mask generating package.

4.2 Alternative Mask Combining Methods

The current prototype's mask combination methods make use of static blending and multitexturing; however, runtime conditions may permit or require additional control over the blending process. The current system requires participants to sit in relatively specific positions; if a user moves more than a foot to either side, significant ghosting may appear. In a very structured situation, this may be sufficient, but in a more casual situation, participants may desire more freedom of motion. To provide this freedom and determine viewpoints at runtime, knowledge about the position of local users is necessary. Fortunately, in a full-duplex system, cameras are already directed towards the local participants. It is therefore reasonable to introduce video-based head tracking as a means to determine local participant's viewpoints.

To generate masks at runtime, additional calibration information would be required. This could consist of a set of static masks within a reasonable working area. From these masks and a model of the multi-view display, one could devise a set of functions representing the visibility each pixel of the display as a function of viewer position. Given the video-tracked position of each user, dynamic masks could be generated. Furthermore, if the masks are generated at runtime, view conflicts could be better evaluated at runtime.

If there is no conflict, allocate the subpixel to the single view. If instead there is a conflict, evaluate the subpixel based on each view's camera image for that subpixel. If each view needs a widely different intensity, then it may be desirable to deactivate the subpixel; however, if each image proposes a similar value for the subpixel, then averaging may be more appropriate.

5. CONCLUSIONS

The current prototype system demonstrates an improvement in teleconferencing systems by using multiple cameras and a multi-view display. Each local participant can view a distinct and spatially appropriate view of the remote participants. As a result, participants can use non-verbal cues during conversation for indicating referred identities, which has the potential to reduce confusion, and participants can maintain eye contact, which has the potential to increase trust [4, 5].

The calibration method introduced to define the views presents another advantage: flexibility and arbitrary viewpoints. Because the calibration system requires no model of the display and multi-view technique, it can handle any spatial multiplexing method that discretely assigns subpixels to different views. This can be a significant benefit for calibrating future multi-view displays that require complex models. Simply placing the calibration camera in a particular location permits the evaluation of all visible pixels from that position. By developing this flexible calibration system and appropriately calibrated remote cameras, the prototype multi-view teleconferencing system achieved its goals: provide a strong sense of presence to a video teleconferencing system and provide a basis for calibrating future multi-view displays.

6. REFERENCES

[1] W.-C. Chen, H. Towles, L. Nyland, G. Welch, and H. Fuchs. Toward a compelling sensation of telepresence: demonstrating a portal to a distant (static) office. In *VIS '00: Proceedings of the conference on Visualization '00*, pages 327–333, Los Alamitos, CA, USA, 2000. IEEE Computer Society Press.

[2] M. Halle. Autostereoscopic displays and computer graphics. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Courses*, page 104, New York, NY, USA, 2005. ACM.

[3] N. Holliman. 3d display systems. Technical report, Department of Computer Science, University of Durham, 2005.

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

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

[6] D. J. Sandin, T. Margolis, J. Ge, J. Girado, T. Peterka, and T. A. DeFanti. The varrier™ autostereoscopic virtual reality display. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, pages 894–903, New York, NY, USA, 2005. ACM.