The Extended Window Metaphor for Large High-Resolution Displays

Malte Willert, Stephan Ohl, Anke Lehmann, Oliver Staadt

University of Rostock, Germany

Abstract

We present a novel tele-presence approach that extends the window metaphor by combining large high-resolution LCD walls with multi-camera 3D video. We propose to integrate an array of cameras into the bezels of the wall to support flexible camera placement for optimized video acquisition. The users's 3D video representation combined with the high-resolution LCD wall provides local and remote users with a shared virtual space in an extended life-size window metaphor. We discuss important system design aspects such as camera placement strategies, resolution, field of view, and dynamic camera selection for different 3D video reconstruction approaches, such as stereo and visual hulls. Finally, we describe our current prototype system based on the design guidelines described in this paper.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality

1. Introduction

In recent years, the demand for systems that support remote collaboration and tele-presence has increased steadily. Collaborating teams are often spread over large distances and ad-hoc meetings are difficult to organize. Video conferencing, online document sharing, and cloud computing cannot fully replace face-to-face meetings. In this paper, we present a new tele-presence approach that draws on recent developments in the areas of large high-resolution displays, 3D video, virtual reality, and user interaction. Our prototype system comprises off-the-shelf hardware components and can be replicated easily.

We explore the use of large high-resolution LCD walls for collaborative tele-presence. Tiled display walls have several advantages over traditional projection-based display systems due to their scalable pixel dimensions and display area [NSS*06]. Thus, we are able to display remote users at a natural scale at a sufficient resolution. Furthermore, this enables us to use the display as a shared application workspace for high-resolution content. As a result, frequent zoom-and-pan interaction using mouse and keyboard is reduced compared with standard desktop displays: If a user needs an overview of the workspace, she can step back. To obtain a detailed view, she can simply step closer to the area of interest on the screen.

One of our central ideas is the integration of cameras directly into the LCD wall by utilizing the "unused" bezel space between individual display panels. Integrating cameras directly into the display wall provides us with a number of interesting new options. We are able to capture free-viewpoint 3D video with multiple cameras that are facing the user directly. Previous approaches, such as the blue-c tele-presence system [GWN*03], achieved this by using shuttered display screens and placing the cameras behind the screen outside of the user's view. However, the blue-c prototype required numerous custom components and is difficult to replicate.

Our approach can be used as a virtual window into a remote space. We introduce a shared virtual space that is located between the local and remote users. We call this setting *Extended Window Metaphor*. We demonstrate how this shared virtual space enables us to explore new possibilities for interaction and collaboration and how it guides the design. We discuss how we can take advantage of the constraints of our system and present the current hardware implementation of our prototype.

The main contributions of our work are (i) introduction of the Extended Window Metaphor, (ii) integration of a camera array into a tiled LCD wall, (ii) easy-to-use estimations to guide camera placement, (iv) a discussion on how the shared virtual space helps to overcome some resolution problems and (v) a description of the chosen hardware for our prototype system.

2. Related Work

The Office of the Future by Raskar et al. [RWC*98], the National Tele-immersion Initiative [STL*01] and Lanier [Lan01] highlighted the importance of tele-immersion as advanced video-conferencing environments with support for important non-verbal communication cues such as gaze. The blue-c tele-presence environment by Gross et al. [GWN*03], introduced the first fully-functional bi-directional tele-presence system. These early systems, however, were proof-of-concept prototypes that are difficult to replicate.

Anstis et al. [AMM69] identified gaze direction as a crucial non-verbal cue in tele-communication applications. The influence of camera position, relation between camera position to image position and of the displayed image size on the perception of eye contact was studied in [GM03]. An evaluation of the influence of eye contact on trust in messaging was carried out in [BS06]. Nguyen and Canny [NC07] showed that non-verbal cues like gaze awareness help to improve trust in video conferencing. Recent studies by Roberts et al. [RWR*09] and Murray et al. [MRS*09] demonstrated the importance of realistic eye gaze in immersive collaborative environments. These studies conclude that (i) tele-presence system has to support high-resolution video avatars to enable the user to track eye gaze, (ii) realistic eye movement is necessary to judge gaze direction, (iii) a real image of the remote user is superior to synthetic avatars in perceived presence, and (iv) a 2D video stream without perspective correction and a correct parallax can lead to loss of trust and gaze judgement.

Lincoln et al. [LNI*09] presented a multi-user lenticular display with support for multiple perspectives for each viewer. The system supports non-verbal cues, but the users are required to remain in specific locations in front of the display. An early tele-conferencing system using a wall-sized display and enabling interaction in a virtual environment is presented in [GAB99]. A multi user approach described in [KS02] provides a perspective correct view on the conferees located around a virtual desk. The HoloPort system by Kuechler and Kunz [KK06] supports gaze-awareness as well as on-screen and free-space interaction using an infrared pen. However, shared collaboration data, which is overlaid on top of the remote video, occludes the remote user's hands.

Numerous multi-view video schemes have been presented in recent years. For an overview see Kubota et al. [KSM*07]

and Smolic [Smo09]. Many of these approaches are not suitable for online collaboration due to their lack of real-time support. Real-time systems frequently use silhouette-based visual hull approaches [GWN*03, KK09, MBR*00, MBM01, PLB*09, PLM*10]. Petit et al. [PLB*09] devised a parallel version of the polyhedral visual hull algorithm. Vasudevan et al. [VLK*10] presented a multi-camera system that covers a 360°-view of the user. They calculate a 3D mesh in real time from pairs of video images using a stereo approach. The resulting 3D user representation is merged with a shared virtual environment. The camera acquisition setup, however, cannot be integrated with the immersive display system in a straightforward manner to support bi-directional communication. Furthermore, the use of active-stereo glasses prohibits direct eye contact between local and remote users.

In contrast, autostereoscopic displays, such as the Varrier tiled LCD wall [SMG*05], enable the user to perceive 3D content without wearing glasses. Jones et al. [JLF*09] presented a one-to-many 3D Teleconferencing system that facilitates gaze cues and eye contact. The system captures a multi-view representation of the user's head and renders it on a volumetric autostereoscopic display. Currently, the display system supports only one 3D subject.

Recent work by Ebara et al. [EKLK07, ES09] placed video cameras at the center and the edges of a tiled display, respectively, to support gaze awareness and eye contact in a large display environment. Only the 2D video stream of a single camera is transmitted to the remote site and integration into a shared virtual space is not supported.

Several systems have used video cameras in large display environment for user interaction purposes. Strødle et al. [STLA09] placed video cameras along the floor to create an interaction space in front of their display wall where the 3D position of a user's finger can be triangulated without using markers. Luo and Kenyon [LK09] employed a single camera to control an ultra-high-resolution tiled display wall by using gestures.

Our goal is to integrate the idea of virtual see-through into a collaborative virtual environment with 3D video conferencing combined with the benefits of a large screen (e.g., life-size user, large interaction space, gaze awareness).

3. Extended Window Metaphor

The *Window Metaphor* uses a display as a real window into a virtual space. If a user is moving in front of the display, the virtual camera viewpoint, which can be obtained using a head tracker, is updated. Thus, depth perception is created via motion parallax.

Yuan [Yua09] called this setting "virtual 3D see-through experience" for high-resolution displays. The basic window metaphor creates a window that looks into a virtual space that contains a 3D scene. We suggest a simple and natural
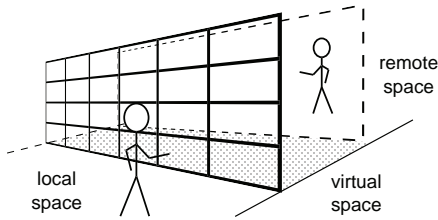
Figure 1: *Illustration of the Extended Window Metaphor*

extension to the basic window metaphor for collaborative environments. Firstly, we look through the window into a virtual space as before. Secondly, we place the remote physical space of our collaborator behind the virtual space. In other words, we are looking through the "second window" into our collaborator's physical space. This concept is illustrated in Figure 1.

The depth of the virtual space in between the two windows can be adjusted and can contain shared collaborative data. The collaborators are separated from this virtual space. We believe that this arrangement is suitable for very natural collaboration in distributed large display environments.

The benefits are:

1. **natural scale**: users (and virtual objects) appear in natural size.
2. **gaze awareness**: We can see where our remote collaborator is looking at, whether she is making eye contact or is looking at some object in the virtual space.
3. **collaboration space in between**: The virtual space for collaboration is in between of the two users.
4. **high resolution content**: Content in the virtual space may be displayed at ultra-high resolution.
5. **intuitive navigation**: The systems supports motion parallax based on tracking the users's head position and orientation.

To implement the Extended Window Metaphor for large high-resolution displays we need three main hardware components: (i) a display connected to a computation cluster, (ii) low-latency head tracking, and (iii) a camera array integrated into the display. All systems may be built from available off-the-shelf components.

Some consumer displays already contain integrated cameras. Generally, these cameras are positioned centered and above the display. One may use these displays directly for building a tiled display. To position the cameras in a more flexible fashion, we decided to integrate small micro-lens cameras with remote heads between the bezels of the display panels. See Section 5 for more details on the hardware setup of our prototype implementation.

4. System Design Considerations

In this section we discuss some general conditions and constraints associated with the proposed setup. Firstly, we consider some aspects of the user's action radius and visual acuity in front of a large high resolution display. We discuss camera positioning, camera tilt, and sensor dimensions. The impacts of all parameters are characterized by a set of approximation formulas. Then, the next subsection discusses in detail the screen-space resolution of objects on the tiled display using these estimations.

To create virtual viewpoints from camera arrays there exists a variety of algorithms. The more cameras are used, the more data needs to be processed. If the camera array is sparse, then more computational power is needed. We devise camera selection strategies to reduce bandwidth and computational requirements.

4.1. Visual Acuity

The position of the user in front of the large display is related directly to the placement of the virtual camera and the user's perceived resolution. In the following we will give some characterisations.

For ergonomic reasons, the minimal distance to a standard LCD panel in a desktop environment is normally 40 cm. Therefore, we assume that this is the *minimal user–display distance* for users in front of a tiled LCD wall. The *maximum user–display distance* is limited by the room size or the range of the tracking system.

The limit of human visual acuity is ≈ 0.5 min of arc (i.e., $0.0083°$) [Sta02]. Assuming a user-display distance, d_{ud}, and a display resolution, $r_{display}$, the resulting angular resolution ρ may be approximated as

$$\tan\rho \approx \frac{1}{r_{display}d_{ud}}. \tag{1}$$

Assuming $d_{ud} = 40$ cm and $r_{display} = 40$ px/cm (≈ 101 dpi), then, ρ is approximately $0.036°$/px. This value does not reach the limit for the visual acuity of the human eye. Nevertheless, it is a good landmark because this is the angular resolution in desktop environments.

4.2. Camera Placement

For practical reasons we restrict the placement of cameras to the bezels between pairs of tiles. As a consequence, some of the space in front of every display cannot be covered. However, cameras may be placed freely in locations outside of the display wall.

The exact location of the cameras on the grid does not only depend on the tile layout, but also on the number of cameras and the 3D video reconstruction algorithm. Therefore, in this section we will give some general rules and some
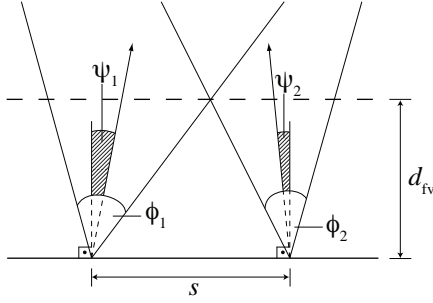
Figure 2: *Two cameras with spacing s. Both tilt angles, ψ_1 and ψ_2, are negative in this illustration.*

simple formulas to relate the field of view (FOV) and camera spacing on the grid.

We define the *full-view distance*, d_{fv}, as the distance from the display wall, from where a point in space is seen by at least one camera. The *full-overlap distance*, d_{fo}, is the distance where the point is seen by at least two cameras. These distances should be as small as possible and at best smaller than the minimal user–display distance d_{ud}.

The full-view distance, d_{fv}, may be calculated using two cameras where s is the spacing between the cameras (see Figure 2). The FOV and tilt angles for the i-th camera are ϕ_i and ψ_i respectively:

$$d_{\text{fv}} = s \left[\tan(\tfrac{1}{2}\phi_1 + \psi_1) + \tan(\tfrac{1}{2}\phi_2 + \psi_2) \right]^{-1}. \quad (2)$$

In the common case where the two cameras have the same FOV and the tilt is zero (the cameras principal rays are perpendicular to the display) the equation simplifies to

$$d_{\text{fv}} = \frac{1}{2} s \tan(\frac{180° - \phi}{2}). \quad (3)$$

For example, to get a full-view distance of at most the minimal user-display distance of 40 cm with a FOV $\phi = 80°$ the spacing must be $s < 67$ cm.

Obtaining bounds for full-overlap distance is more complex. Therefore, we just consider the most simple case, where all cameras have the same spacing, s, the same FOV, and the principal rays are all perpendicular. Then, the full-overlap distance, d_{fo}, is just two times the full-view distance, d_{fv},

$$d_{\text{fo}} = 2d_{\text{fv}} \quad \text{or} \quad d_{\text{fo}} = s \tan(\frac{180° - \phi}{2}). \quad (4)$$

Despite the overlap of the cameras FOV, another important factor is the object height in pixels, p_{obj}, seen by the camera sensor at different distances. Given an object of height h_{obj} at distance d_{obj} as well as vertical camera sen-

sor dimension in pixels, p_{cam}, and the FOV, ϕ, we estimate p_{obj} using

$$p_{\text{obj}} = \frac{p_{\text{cam}}}{2d_{\text{obj}} \tan(\frac{1}{2}\phi)} h_{\text{obj}}. \quad (5)$$

Often, only the horizontal FOV, ϕ_{h}, of camera lenses is available from the specification. If the aspect ratio of the chip h/w is known, the vertical FOV, ϕ_{v}, may be calculated from the following relation:

$$\tan \frac{\phi_{\text{v}}}{2} = \frac{h}{w} \tan \frac{\phi_{\text{h}}}{2}. \quad (6)$$

4.3. Camera Tilt

In typical camera layouts for 3D video acquisition, the cameras point at the center of the scene. This is not the best choice for large high-resolution displays, because the user is expected to move within a considerable distance to the left and to the right in front of the display. Therefore, the *horizontal camera tilt* should be $0°$. At the display's boundaries one may let the cameras point inwards, depending on whether the user can move the left or right boundary of the display.

However, the *vertical camera tilt* may be modified to increase camera overlap and reduce (vertically) the full-view distance. For example, using Equation 2, we set a moderate vertical tilt $\psi_1 = 0°$, $\psi_2 = 15°$, a vertical FOV $\phi_{1,2} = 70°$ and the vertical camera spacing, s_v to one display height $s_v = 40$ cm. The full-view distance then is $d_{\text{fv}} \approx 21$ cm. Without tilt ($\psi_1 = \psi_2 = 0°$) we obtain $d_{\text{fv}} \approx 28$ cm.

Now, using Equation 3 we can calculate the horizontal camera spacing that would be required to get the same full-view distance of $d_{\text{fv}} = 21$ cm. With $\phi_v = 70°$ the horizontal FOV is $\phi_h = 86°$ (with an aspect ratio h/w of 3/4). Inserting these values into Equation 3 results in a horizontal camera spacing of $s_h = 39$ cm. Note that this is approximately the vertical camera spacing s_v from above.

This example showed that introducing vertical camera tilt helps to reduce the full-view distance. Reducing the full-view distance also reduces the full-overlap distance. Overlap of neighbouring cameras is crucial for depth reconstruction that based on finding correspondences.

For this example we assume a fixed monitor height, because the camera sensor's aspect ratio is 4:3 whereas most modern LCD panels are 16:10. If the goal is to minimize the full-view distance, then placing the cameras on the horizontal gaps between displays is preferable.

4.4. Resolution

The depth of the virtual space, e, has effect on the projected size of the remote user with height d_{obj} on the tiled display as illustrated in Figure 3. The image height in pixels, p_{obj}, may be calculated as
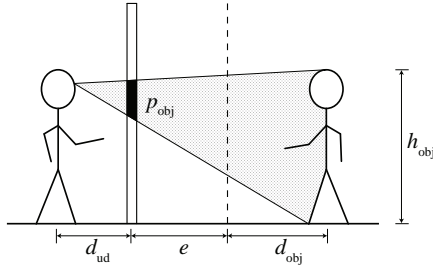
Figure 3: *Size of objects on the large high-resolution display.*

$$p_{\text{obj}} = r_{\text{display}} \frac{d_{\text{ud}} h_{\text{obj}}}{d_{\text{ud}} + e + d_{\text{obj}}}. \tag{7}$$

Combining Equations 7 and 5 leads to Equation 8. This means, that object resolution equals screen resolution. In other words, objects are acquired at the resolution that is needed for display on the LCD wall:

$$\frac{r_{\text{display}} \, d_{\text{ud}}}{d_{\text{ud}} + e + d_{\text{obj}}} = \frac{p_{\text{cam}}}{2 d_{\text{obj}} \tan \frac{\phi}{2}}. \tag{8}$$

The diagonal *size* of a typical LCD desktop monitor is in the range between 20" and 30". The display *resolution,* r_{display}, is between 30 to 45 px/cm (or 75 to 115 dpi). The *pixel dimension* of typical camera sensors varies between 640×480 (for VGA) and 1920×1080 (for HDTV 1080p) pixels.

Consider the case where both users are standing close to the wall (i.e., at minimal user–wall distance, $d_{\text{ud}} = d_{\text{obj}} = 40$cm), and we fix vertical field of view, ϕ_v, display resolution, and the depth of the virtual space to reasonable values (e.g., $\phi_v = 70°$, $r_{\text{display}} = 40$ px/cm, $e = 100$ cm), then we gain some insight into the required camera sensor's pixel dimensions. Using Equation 8, the result turns out to be $p_{\text{cam}} = 498$ px, which is just above VGA resolution.

If we set $e = 0$ cm, then $p_{\text{cam}} = 1120$px, which is larger than vertical HDTV 1080p. This means, that the added virtual space—if chosen sufficiently deep—will aid in resolving our resolution problem.

Note that in both cases more than one camera may be required to acquire the whole object. Equation 8 considers the number of required pixels without taking into account the object's height.

We provide a second example to illustrate what happens if we step back from the large display. We can see in Figure 3 that this increases the projected size of the remote user. Hence, to quantify the effect, we again set $d_{\text{ud}} = 200$ cm

and the extend $e = 100$ cm. Unlike in the above case, we are interested in the projected screen-space height of the remote user, p_{obj}. We assume a user who is $h_{\text{obj}} = 170$ cm tall. The values for r_{display} and d_{obj} remain the same and we obtain $p_{\text{obj}} = 4000$ px from Equation 7. Standing close to the display at a distance of $d_{\text{ud}} = 40$ cm we would obtain $p_{\text{obj}} = 1511$ px.

Now, standing farther from the LCD wall, we may not need the high resolution. To achieve the angular resolution of desktop displays, $\rho = 0.036°$, at a distance $d_{\text{ud}} = 200$ cm, we need a display resolution of $r_{\text{display}} = 8$ px/cm. Hence, the remote user's image size increases by a factor of 2.7, but the necessary display resolution decreases faster (by a factor of 0.2). That means angular resolution compensates well the increasing image size of the user.

At this point we have considered two extreme cases: (i) both users standing close to the display and (ii), the remote user remains close and the local user steps back. An exhaustive discussion would have to include the two remaining cases: (iii) the remote users steps back, whereas the local user stays close and (iv), both users step back. We will consider these cases briefly.

Using Equation 8 for case (iii) with $d_{\text{du}} = 40$ cm, $e = 100$ cm, $d_{\text{obj}} = 200$ cm, $\phi_v = 70°$, and $r_{\text{display}} = 40$ px/cm, we realize that the necessary camera resolution increases slightly to $p_{\text{cam}} = 659$ px. For the remaining case (iv) with $d_{\text{du}} = d_{\text{obj}} = 200$ cm, $e = 100$ cm, $\phi_v = 70°$, and $r_{\text{display}} = 40$ px/cm we obtain $p_{\text{cam}} = 4481$ px. If we multiply by 0.2 we get 897 px, which is the necessary vertical pixel dimension if we want to reach desktop angular resolution.

We conclude that resolution issues can be compensated by the added virtual space. There is a trade-off between wide FOV (to assure low minimal user-display distance) and high resolution (requiring smaller FOV and/or higher chip dimensions). The limiting case is where both users step back.

4.5. Dynamic Camera Selection

The user is typically not captured by all cameras simultaneously when standing close to the LCD wall. Hence, we may select a subset of cameras dynamically depending user's position.

Typical *camera chip resolutions* vary between 640x480 (VGA) and 1920x1080 (HDTV 1080p), which results in a bandwidth of about 1 - 6 MBytes for each RGB frame. A favorable *framerate* for realtime applications is 30 fps. Combined with a frame rate of 30 Hz, that means each camera generates 30 to 180 MBytes of (uncompressed RGB) data per second. For that reason bandwidth may become an issue if the camera array is dense.

Consequently, dynamic camera selection also saves processing time. Consider, for example, a visual hull based reconstruction scheme with a complexity of $O(n^2)$, where n is the number of silhouettes (see, e.g., [PLM*10]).

From the above discussion, we can estimate the set of cameras that can capture the user based on head tracking information. If the user is standing close to the display, only some columns of the camera grid are needed. If the user steps back she will be visible in more cameras. In this case the camera rows from the top and bottom may be left out, because the user is imaged completely in centre camera rows.

4.6. Standard Reconstruction Algorithms

Two standard algorithms which are frequently used in real-time 3D and depth reconstruction are visual hull based approaches and small baseline stereo.

Visual Hull based approaches reconstruct full geometry by carving the object from space using its silhouettes. For good results, different viewpoints around the objects are needed. Hence, we need additional cameras that are not integrated into the tiled display. Generally, cameras can be placed freely to the left and to the right of the wall. However, one has to keep in mind that any cameras capturing parts of the tiled display itself might not be advisable, because the dynamic content on the display will complicate background segmentation. Precise and fast segmentation is a crucial and very sensitive step in real-time visual hull reconstruction.

In small baseline stereo reconstruction settings, the cameras are arranged in pairs. We can treat a stereo pair as one camera when using Equation 2 to calculate the full-view distance. Within each stereo pair's cone we are able to acquire depth information, because we have overlap. That means, for small baseline stereo, the full-overlap distance is approximately equal to the full-view distance ($d_{fo} \approx d_{fv}$).

5. Prototype Setup and Discussion

Based on the design parameters outlined in the previous sections, we built our prototype system. We describe our prototype and its hardware components as outlined in Section 3. Furthermore, we discuss our results, possible drawbacks and propose further research.

(i) Displaywall and Computation Cluster: The basic construction of our tiled LCD wall follows the guidelines from the *OptIPortal* project [DLR*09]. It comprise 24 DELL 2709W displays. These displays were chosen due to their relative small bezels combined with a large viewing angle. A more detailed specification is presented in Table 1.

A set of four panels is connected to one compute node. Each compute node is equipped with two NVIDIA GTX 260 graphic card. The cluster is interconnected using Gigabit Ethernet.

(ii) Head Tracking System: We use a 12-camera *OptiTrack* infrared head-tracking system, which provides us with a large tracking volume in front of the LCD wall. The volume

Table 1: *Specification of our tiled LCD wall*

display wall properties	
# of 27" panels (6×4)	24
# of horizontal pixels	11,520
# of vertical pixels	4,800
(outer) height	(164 cm) 158 cm
(outer) width	(378 cm) 373 cm
bezel width between panels	5 cm
height above floor	69 cm
single panel heigth	41 cm
single panel width	63 cm

inside which a user can move while looking at the wall provides a maximal user-display distance of 3,5 m away from the screen and a minimal user-display distance of 40 cm. This reflects what we stated in section 4.1.

(iii) Camera Array: To evaluate the configurations for the cameras and as a proof of concept we integrated six remote-head PointGrey Dragonfly2 Firewire cameras inside the wall (see figure 4). The cameras are connected each to one compute node to use its full processing power for preprocessing the acquired images. To acquire additional views we can position eight PointGrey Flea2 cameras around the volume in front of the display wall. Our Dragonfly2 cameras are equipped with wide-angle high resolution 3.6 mm lenses with a field of view of 82.5° at a minimal working distance of 40 cm and an optical resolution of 150 lp/mm. We decided against lenses with a wider FOV due to the heavy loss of resolution outside their centre and the introduced distortion. The Dragonfly2 cameras use a 1/3" sensor with an aspect ratio of 4/3 and square pixels. Using equation 6 and our horizontal field of view of 82.5° this delivers a vertical field of view of 66.67°. Due to physical simplicity we consider the cameras mounted perpendicular inside the wall.

Now we are able to calculate required setup and number of cameras using the above specifications. Having a minimal working distance of 40 cm provided by the lenses, which is even the minimal ergonomic viewing distance according to section 4.1, we obtain a full-viewdistance with a horizontal camera spacing of 70 cm. For the vertical positioning we need a spacing of 52,6 cm. A full-overlap can be achieved by dividing the above full-view distance by two, resulting in 35 cm horizontal and 26.3 cm vertical distances between the cameras.

Our wall has a width of 378 cm and a height of 164 cm. Following the calculation above we would need 11 cameras horizontal and 7 cameras vertical positioned inside the wall to have full overlap in all directions. Due to the restriction of positioning the cameras on the bezels we are not able to position the cameras on a grid with the calculated spacing of 35 x 26.3 cm. Since full overlap is only important for
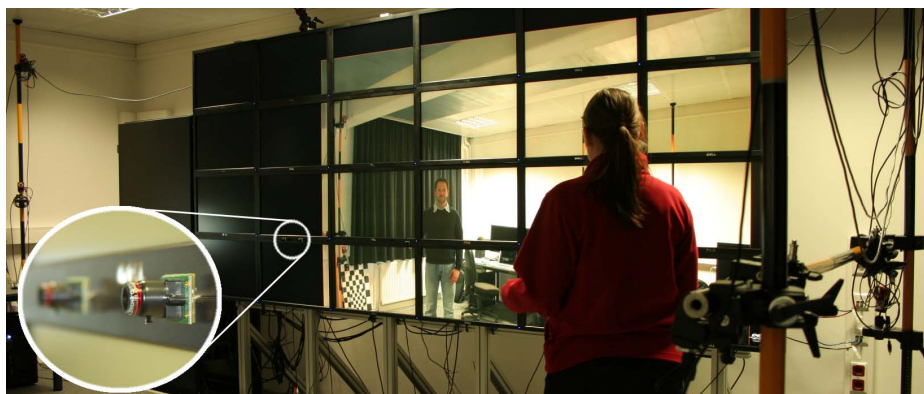
Figure 4: *Mockup of our prototype setup with perspectively corrected image to illustrate the Extended Window Metaphor. Note that the virtual space is not shown.*

wide baseline stereo depth extraction we can decide to just have full overlap in horizontal axis. In section 4.3 we also discussed why it is beneficial to place the cameras on the horizontal bezels. Positioning the cameras in equal distances without displacement from row to row, we can position 11 cameras horizontal per row in the required spacing of 35 cm and 5 cameras in each column in a distance of 41 cm due to the height of the panels. This results in a total number of 55 cameras if one wants to use wide baseline depth from stereo algorithms that will work up to the minimal distance of 40 cm. With the vertical camera distance of 41 cm we reach vertical full-overlap distance at 63 cm away from the display wall. Using the option of tilting the cameras vertically we could reduce this distance further (see 4.3 and Fig. 2).

Our approach has also some drawbacks. First, the display wall doesn't support stereo rendering. This would be possible with autostereoscopic displays (like in [SMG*05]) or glasses. Stereoscopic display that require the user to wear glasses are not suitable for tele-presence systems because they do not support eye contact. As a result, synthetic 3D objects can only be perceived as being inside the *virtual space* and in front of the display. However, by supporting motion parallax through 3D video, we believe that this will compensate for the lack of stereo parallax.

Currently, spatial audio recording is not considered, but to improve presence, we plan to investigate this in the future.

6. Conclusions

We introduced a novel tele-presence approach for large LCD walls, which we call the Extended Window Metaphor. We investigate the various constraints and option for camera setup and showed how the number of cameras can be reduced using vertical tilt. We further showed in 4.4 that due to the physical properties and the human visual acuity the use of cameras with a pixel resolution of 800x600 is sufficient for

most practical cases . We gave some principal ideas how to further reduce the acquired data due to a camera selection. Our results show that a considerable number of cameras is necessary to build a tele-presence system that supports 3D video acquisition at high resolution. Therefore, we plan to investigate alternative video acquisition schemes including the use of time-of-flight depth cameras.

7. Acknowledgments

References

[AMM69] ANSTIS S. M., MAYHEW J. W., MORLEY T.: The perception of where a face or television "portrait" is looking. *Am J Psychol 82*, 4 (Dec 1969), 474–489. 2

[BS06] BEKKERING E., SHIM J.: Trust in videoconferencing. *Commun. ACM 49*, 7 (2006), 103–107. 2

[DLR*09] DEFANTI T. A., LEIGH J., RENAMBOT L., JEONG B., VERLO A., LONG L., BROWN M., SANDIN D. J., VISHWANATH V., LIU Q., KATZ M. J., PAPADOPOULOS P., KEEFE J. P., HIDLEY G. R., DAWE G. L., KAUFMAN I., GLOGOWSKI B., DOERR K., SINGH R., GIRADO J., SCHULZE J. P., KUESTER F., SMARR L.: The OptIPortal, a scalable visualization, storage, and computing interface device for the OptiPuter. *Future Generation Computer Systems 25*, 2 (Feb. 2009), 114–123. 6

[EKLK07] EBARA Y., KUKIMOTO N., LEIGH J., KOYAMADA K.: Tele-immersive collaboration using high-resolution video in tiled displays environment. In *Advanced Information Networking and Applications Workshops, 2007, AINAW '07. 21st International Conference on* (2007), vol. 2, pp. 953 –958. 2

[ES09] EBARA Y., SHIBATA Y.: Study on realistic communication technology with tiled displays wall. In *Network-Based In-*

formation Systems, 2009. NBIS '09. International Conference on (2009), pp. 612 –617. 2

[GAB99] GIBBS S. J., ARAPIS C., BREITENEDER C. J.: Teleport— towards immersive copresence. *Multimedia Syst. 7*, 3 (1999), 214–221. 2

[GM03] GRAYSON D. M., MONK A. F.: Are you looking at me? eye contact and desktop video conferencing. *ACM Trans. Comput.-Hum. Interact. 10*, 3 (2003), 221–243. 2

[GWN*03] GROSS M., WÜRMLIN S., NAEF M., LAMBORAY E., SPAGNO C., KUNZ A., KOLLER-MEIER E., SVOBODA T., VAN GOOL L., LANG S., STREHLKE K., MOERE A. V., STAADT O.: blue-c: a spatially immersive display and 3d video portal for telepresence. *ACM Trans. Graph. 22*, 3 (2003), 819–827. 1, 2

[JLF*09] JONES A., LANG M., FYFFE G., YU X., BUSCH J., MCDOWALL I., BOLAS M., DEBEVEC P.: Achieving eye contact in a one-to-many 3d video teleconferencing system. In *SIGGRAPH '09: ACM SIGGRAPH 2009 papers* (New York, NY, USA, 2009), ACM, pp. 1–8. 2

[KK06] KUECHLER M., KUNZ A.: Holoport - a device for simultaneous video and data conferencing featuring gaze awareness. In *VR '06: Proc. of the IEEE conference on Virtual Reality* (Washington, DC, USA, 2006), IEEE Computer Society, pp. 81–88. 2

[KK09] KNOBLAUCH D., KUESTER F.: Focused volumetric visual hull with color extraction. In *ISVC '09: Proceedings of the 5th International Symposium on Advances in Visual Computing* (Berlin, Heidelberg, 2009), Springer-Verlag, pp. 208–217. 2

[KS02] KAUFF P., SCHREER O.: An immersive 3d videoconferencing system using shared virtual team user environments. 2

[KSM*07] KUBOTA A., SMOLIC A., MAGNOR M., TANIMOTO M., CHEN T., ZHANG C.: Multi-View imaging and 3DTV (Special issue overview and introduction). *IEEE Signal Processing Magazine, Special Issue on Multi-view Imaging and 3DTV 24*, 6 (2007). 2

[Lan01] LANIER J.: Virtually there. *Scientific American* (April 2001), 66–75. 2

[LK09] LUO X., KENYON R. V.: Scalable vision-based gesture interaction for cluster-driven high resolution display systems. *Virtual Reality Conference, IEEE 0* (2009), 231–232. 2

[LNI*09] LINCOLN P., NASHEL A., ILIE A., TOWLES H., WELCH G., FUCHS H.: Multi-view lenticular display for group teleconferencing. In *IMMERSCOM '09: Proc. of the 2nd International Conference on Immersive Telecommunications* (ICST, Brussels, Belgium, Belgium, 2009), pp. 1–8. 2

[MBM01] MATUSIK W., BUEHLER C., MCMILLAN L.: Polyhedral visual hulls for Real-Time rendering. In *Rendering Techniques 2001: 12th Eurographics Workshop On Rendering* (June 2001), Eurographics, pp. 115–126. ISBN 3-211-83709-4. 2

[MBR*00] MATUSIK W., BUEHLER C., RASKAR R., GORTLER S. J., MCMILLAN L.: Image-Based visual hulls. In *Proceedings of ACM SIGGRAPH 2000* (July 2000), Computer Graphics Proceedings, Annual Conference Series, ACM Press / ACM SIGGRAPH / Addison Wesley Longman, pp. 369–374. 2

[MRS*09] MURRAY N., ROBERTS D., STEED A., SHARKEY P., DICKERSON P., RAE J., WOLFF R.: Eye gaze in virtual environments: evaluating the need and initial work on implementation. *Concurr. Comput. : Pract. Exper. 21*, 11 (2009), 1437–1449. 2

[NC07] NGUYEN D. T., CANNY J.: Multiview: improving trust in group video conferencing through spatial faithfulness. In *CHI '07: Proc. of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 2007), ACM, pp. 1465–1474. 2

[NSS*06] NI T., SCHMIDT G., STAADT O., LIVINGSTON M., BALL R., MAY R.: A survey of large High-Resolution display technologies, techniques, and applications. In *Virtual Reality Conference, 2006* (2006), pp. 223–236. 1

[PLB*09] PETIT B., LESAGE J.-D., BOYER E., FRANCO J.-S., RAFFIN B.: Remote and Collaborative 3D Interactions. In *IEEE 3DTV Conference* (Potsdam Allemagne, 2009). 2

[PLM*10] PETIT B., LESAGE J.-D., MENIER C., ALLARD J., FRANCO J.-S., RAFFIN B., BOYER E., FAURE F.: Multicamera real-time 3d modeling for telepresence and remote collaboration. *International Journal of Digital Multimedia Broadcasting 2010* (2010), 12 pages. 2, 5

[RWC*98] RASKAR R., WELCH G., CUTTS M., LAKE A., STESIN L., FUCHS H.: The office of the future: a unified approach to image-based modeling and spatially immersive displays. In *SIGGRAPH '98: Proceedings of the 25th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1998), ACM, pp. 179–188. 2

[RWR*09] ROBERTS D., WOLFF R., RAE J., STEED A., ASPIN R., MCINTYRE M., PENA A., OYEKOYA O., STEPTOE W.: Communicating eye-gaze across a distance: Comparing an eye-gaze enabled immersive collaborative virtual environment, aligned video conferencing, and being together. In *VR '09: Proceedings of the 2009 IEEE Virtual Reality Conference* (Washington, DC, USA, 2009), IEEE Computer Society, pp. 135–142. 2

[SMG*05] SANDIN D. J., MARGOLIS T., GE J., GIRADO J., PETERKA T., DEFANTI T. A.: The Varrier autostereoscopic virtual reality display. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers* (New York, NY, USA, 2005), ACM, pp. 894–903. 2, 7

[Smo09] SMOLIC A.: An overview of 3d video and free viewpoint video. In *CAIP '09: Proceedings of the 13th International Conference on Computer Analysis of Images and Patterns* (Berlin, Heidelberg, 2009), Springer-Verlag, pp. 1–8. 2

[Sta02] STANNEY K. (Ed.): *Handbook of Virtual Environments: design, implementation, and applications.* Lawrence Erlbaum Associates, 2002. 3

[STL*01] SADAGIC A., TOWLES H., LANIER J., FUCHS H., VAN DAM A., DANIILIDIS K., MULLIGAN J., HOLDEN L., ZELEZNIK B.: National Tele-Immersion initiative: Towards compelling Tele-Immersive collaborative environments. Presentation given at Medicine meets Virtual Reality 2001 Conference, 2001. 2

[STLA09] STODLE D., TROYANSKAYA O., LI K., ANSHUS O. J.: Tech-note: Device-free interaction spaces. In *3D User Interfaces, 2009. 3DUI 2009. IEEE Symposium on* (march 2009), pp. 39 –42. 2

[VLK*10] VASUDEVAN R., LOBATON E., KURILLO G., BAJCSY R., BERNARDIN T., HAMANN B., NAHRSTEDT K.: A methodology for remote virtual interaction in teleimmersive environments. In *MMSys '10: Proceedings of the first annual ACM SIGMM conference on Multimedia systems* (New York, NY, USA, 2010), ACM, pp. 281–292. 2

[Yua09] YUAN C.: Creating virtual 3D See-Through experiences on large-size 2D displays. In *Virtual Reality Conference, IEEE* (Los Alamitos, CA, USA, 2009), IEEE Computer Society, pp. 237–238. 2