The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays

Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin,
and
Henry Fuchs

University of North Carolina at Chapel Hill and
the NSF Science and Technology Center for Computer Graphics and Scientific Visualization[†]

Abstract

We introduce ideas, proposed technologies, and initial results for an *office of the future* that is based on a unified application of computer vision and computer graphics in a system that combines and builds upon the notions of the CAVE™, tiled display systems, and image-based modeling. The basic idea is to use real-time computer vision techniques to dynamically extract per-pixel depth and reflectance information for the visible surfaces in the office including walls, furniture, objects, and people, and then to either project images *on* the surfaces, render images *of* the surfaces, or interpret changes *in* the surfaces. In the first case, one could designate every-day (potentially irregular) real surfaces in the office to be used as *spatially immersive display* surfaces, and then project high-resolution graphics and text onto those surfaces. In the second case, one could transmit the dynamic image-based models over a network for display at a remote site. Finally, one could interpret dynamic changes in the surfaces for the purposes of tracking, interaction, or augmented reality applications.

To accomplish the simultaneous capture and display we envision an office of the future where the ceiling lights are replaced by computer controlled cameras and "smart" projectors that are used to capture dynamic image-based models with *imperceptible structured light* techniques, and to display high-resolution images on designated display surfaces. By doing both simultaneously on the designated display surfaces, one can dynamically adjust or autocalibrate for geometric, intensity, and resolution variations resulting from irregular or changing display surfaces, or overlapped projector images.

Our current approach to dynamic image-based modeling is to use an optimized structured light scheme that can capture per-pixel depth and reflectance at interactive rates. Our system implementation is not yet imperceptible, but we can demonstrate the approach in the laboratory. Our approach to rendering on the designated (potentially irregular) display surfaces is to employ a two-pass projective texture scheme to generate images that when projected onto the surfaces appear correct to a moving head-tracked observer. We present here an initial implementation of the overall vision, in an office-like setting, and preliminary demonstrations of our dynamic modeling and display techniques.

CR Categories and Subject Descriptors: I.3.3 [Computer Graphics]: Picture/Image Generation—Digitizing and scanning; Display algorithms; Viewing algorithms; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality; I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture—Imaging geometry; Reflectance; Sampling; Scanning; I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Color; Range data; Shading; Shape; Surface fitting; Time-varying imagery; Tracking; I.4.9 [Image Processing and Computer Vision]: Applications; B.4.2 [Input/Output and Data Communications] Input/Output Devices—Image display

Additional Key Words and Phrases: display, spatially immersive display, intensity blending, image-based modeling, image-based rendering, range, depth, reflectance, projection, virtual environments, calibration, autocalibration.

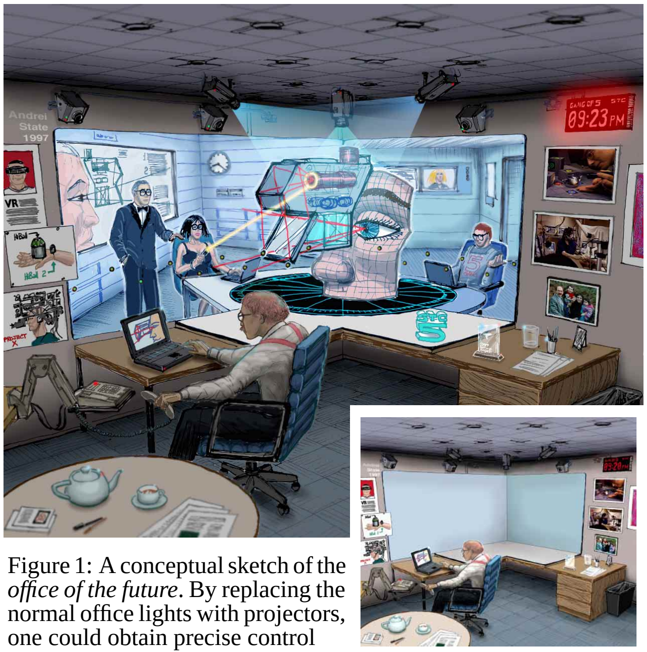

Figure 1: A conceptual sketch of the *office of the future*. By replacing the normal office lights with projectors, one could obtain precise control over all of the light in the office. With the help of synchronized cameras, the geometry and reflectance information can be captured for all of the visible surfaces in the office so that one can project images *on* the surfaces, render images *of* the surfaces, or interpret changes *in* the surfaces. The inset image is intended to help differentiate between the projected images and the real objects in the sketch.

1 INTRODUCTION

The impetus for this work is Henry Fuchs's long-time desire to build more compelling and useful systems for shared telepresence and telecollaboration between distant individuals. It was Fuchs who first inspired us with ideas for using a "sea of cameras" [39]

[†] CB 3175, Sitterson Hall, Chapel Hill, NC, 27599-3175
{raskar, welch, cutts, lake, stesin, fuchs}@cs.unc.edu
http://www.cs.unc.edu/~{raskar, welch, cutts, lake, stesin, fuchs}
http://www.cs.brown.edu/stc/

and imperceptible lights to extract a 3D scene and to reconstruct it at a remote location. These ideas have been refined over several years of collaboration with Ruzena Bajcsy of the University of Pennsylvania's GRASP Laboratory [18], and with our colleagues in the NSF Science and Technology Center for Computer Graphics and Scientific Visualization.

While we are making progress toward our vision for the *office of the future*, we do not yet have a complete working system—the ideas are by definition futuristic. As such, throughout the paper we present a mix of demonstrated results from new methods, and plausible ideas for future systems. We do our best to distinguish between the two lest the reader be led to believe that we have implemented something that we have not.

In the remainder of this section we present motivation for the idea in the form of a story that outlines the developments that led to the vision as presented in this paper. In section 2 we discuss the principal components of the envisioned system, without necessarily discussing a specific implementation. In section 3 we present our approach to actually implementing such an office of the future. In section 4 we discuss our current implementation, and in section 5 we discuss work to be done and future research topics. Note that rather than including a specific "previous work" section we have chosen to discuss related work throughout the paper.

Telecollaboration Interfaces

While telecollaboration systems using 2D "talking heads" and shared white boards have improved significantly over the years, we believe that the through-the-window paradigm itself often inhibits much of the interaction that would otherwise take place if the collaborators were actually together in the same room. In [5] Buxton identifies several tasks for which commercial televideo systems provide only limited support. Aside from limited resolution, part of the problem is that users are forced to maintain two separate *egocenters* (notions of where they are and who they are with): one in their local environment, and another egocenter with the remote collaborators. The participants must then consider and adjust normal behavior to "fit through the window."

One alternative is to implement a shared immersive virtual environment where a user dons a head-mounted display (HMD), disappears from the real world, and enters a shared virtual environment "inside the display" where for example they might see virtual objects along with 2D video avatars of their collaborators. Indeed, we have experimented with such paradigms, as have others. However this interface has several disadvantages. Most obvious are the typical ergonomic problems, for example size, weight, mobility, limited resolution, and limited field of view. Furthermore, for immersive HMD's, the resulting disassociation from a person's comfortable surroundings can be disconcerting and can limit their ability to interact with other people and real objects in the environment.

A more attractive alternative is to get the display off of the user's head, and to instead use a *spatially immersive display* (SID). A SID is a display that physically surrounds the viewer with a panorama of imagery [4]. SID's are typically room-sized, thus accommodating multiple viewers, and are usually implemented with multiple fixed front or rear-projection display units. Probably the most well-known examples of general-purpose SID's are the Cave Automated Virtual Environment (CAVE™) [12], the related tiled-display PowerWall and Infinity Wall™ systems, and Alternate Realities' VisionDome [2]. There are several good examples of telecollaboration applications where the users see and interact with their remote collaborators using a CAVE™ or CAVE™-like system, for example [34]. Such large systems typically require significant physical floor space, for example in a laboratory where there is room for both the screens and the projection units. But we would like to *avoid* going "down the hall" to use the system, instead we would like something as convenient as the telephone—a SID built into an office.

While such an endeavor would probably not be cost-effective solely for the purpose of telecollaboration, if it were of high enough quality one could use it for every-day 2D computer work, video, and 3D immersive visualization. However not only does the construction of such SID's require very careful engineering and assembly, but certain characteristics vary with time and environmental factors such as temperature or vibration. Such time-varying characteristics include the intrinsic and extrinsic projector parameters, intensity balance, color balance, edge alignment, and blending. These problems are most often addressed by periodic mechanical projector calibration, however this approach becomes increasingly difficult and less reliable as the number of projectors increases. Flight simulator developers have faced such problems for some time, and while they have developed digital calibration systems, the systems tend to be highly specialized, thus increasing development cost and overall complexity. A more general-purpose autocalibration scheme would be preferable so that one could modify the display surface or projector configuration as needed. If one could modify the display surface, one could spontaneously add a "drawing board" onto their desktop and the system would account for it. If one had some flexibility over projector placement, one could for example add projectors in an overlapping manner to increase the display resolution (a high-resolution region), or image intensity, or side-by-side to increase the display surface area.

Telecollaboration Infrastructure and Applications

There exists a relatively large body of work in telecollaboration infrastructures and applications, not to mention a large body of work in the area of Computer-Supported Cooperative Work (CSCW). Some representative examples are [34, 37, 6, 15, 28, 19, 27, 36, 3, 41, 10, 33, 37]. Our vision for the office of the future is one in which all of this and similar work can be applied, we hope in new and exciting ways. We envision our office as a particularly compelling interface to be used in support of these efforts, and every-day applications. We are aware of no other one system that attempts to achieve what we do in a similar unified approach.

Of the existing telecollaboration efforts that we know about, the only one that attempts to provide an office-like interface is TelePort [20]. The TelePort system uses wall-sized displays that each show a synthetic scene that is blended with video images of remote participants. As participants move, their locations are tracked so that the images are rendered from the proper perspective. The TelePort display is built into a room that is carefully designed to match the rendered room. The goal is for the virtual room to seem as an extension of the real room. They use carefully constructed geometric models for the office environment, and video-based human avatars obtained by separating the remote participants from the original background (via delta-keying). Rather than building a specialized telecollaboration system that resembles an office, we want to build capability for a life-like shared-room experience *into* existing offices.

"Every Millimeter at Every Millisecond"

One question in our minds was how should remote collaborators and their environment appear remotely? While acceptable for some tasks, we believe that 2D video-based avatars do not effectively engender the sense of being with another person that is necessary for effective interpersonal communication. We want to see and interact with collaborators in 3D, as naturally as we do when we are in the same physical room: gesturing, pointing, walking, waving, using all of the subtle nuances of both verbal and nonverbal communication. A visually attractive possibility would be to use a high-quality 3D image-based rendering or modeling system for each participant (see for example [30, 38, 40, 43]). However we dream of a room-sized working volume, not only because we want mobility, but also because we want to be able to see multiple participants, and to see everyone in their natural surroundings, i.e. their offices. In short, we envision a system similar to [20] where the local and remote offices appear to be physically joined together along some common junction such as a designated wall that is actually a SID. But unlike [20] which overlays 2D video of the remote participants onto a virtual adjoining office, we want to see an image-based 3D reconstruction of the remote office and all of its real contents including people and every-day clutter. That is, we have the ability to capture and remotely display a dynamic image-based model of an entire office.

At some point when considering all of these factors, we came to the realization that if we had access to a dynamic image-based model of the entire office, *including* the designated SID surfaces, we could automatically correct for changes in the time-varying geometric characteristics of the SID. Furthermore, if several cameras could see the display surfaces from a variety of angles, we should be able to observe view-dependent intensity and color variations in the designated display surfaces, thus inferring the surface reflectance properties. In other words, while obtaining an image-based model for the office we could autocalibrate all of the designated display surfaces. Thus the realization that *the SID could in effect be almost anything or anywhere in the office*! It wouldn't matter if the surfaces were irregularly shaped, or if the geometry was changing over time, the image-based model would indicate the variations. And if one was willing to sacrifice some dynamic range in the projected images, one might even be able to use the surface reflectance information to account for slight variations in view-dependent intensity. Note that a crucial advantage of this unified approach is that because the autocalibration and the projection are done by the same device, one eliminates the problems of calibration and drift of the calibration system itself.

Finally, we also note that if one has access to a dynamic image-based model of the entire office, including the occupants, one could potentially extract higher-level representations of the data, assign semantics to those higher-level objects, and then in real-time interpret and respond to object motion or collisions for the purpose of tracking, interaction, or augmented reality (AR). With such capability one could implement untethered interaction as in the Luminous Room, where cameras and projectors serve as "I/O Bulbs" [46, 47]. In this way for example one might be able to track a person's hands so that they could reach out and manipulate a floating 3D model, or perhaps one could detect collisions between real and virtual objects so that virtual objects could be placed on the desk.

Figure 1 depicts a conceptual sketch of our *office of the future,* replicated and in use at three different sites. Note the ceiling-mounted projectors and cameras, the use of lightly-colored material on the designated SID wall and desk area, and the mixed use of that SID for simultaneous image-based and geometric model visualization.

To achieve the above capabilities of acquisition, calibration, and display in a continuously changing office scene with both local and remote participants, we dream of being able to control light in the office over "every millimeter at every millisecond."

2 FUNDAMENTAL COMPONENTS

Our idea for the *office of the future* brings together several fundamental areas of computer science, components that can be enumerated independently from descriptions of their actual implementations. While one goal of this paper is to present the specific implementation of such a system, this does not preclude the use of any of the techniques others have developed in each of these areas. Quite to the contrary, we believe there are trade-offs with all of these techniques which warrant further investigation.

2.1 Dynamic Image-Based Modeling

One of the major components of the system is the module that will capture, continually and in real time, image-based models of the office environment including all of the designated display surfaces. A large body of literature exists from computer vision regarding the determination of depth from a scene. Some of the more common approaches include depth from motion, stereo, focus, and defocus. For our system we are interested not only in dynamic image-based modeling, but over a large volume also. With real-time in mind, many of the techniques traditionally used are difficult because of computational and bandwidth requirements. At CMU, a specialized hardware real-time depth from stereo architecture system has been developed [31]. It can take input from six cameras and produce, at 30 frames/second, a 256×240 depth map aligned with an intensity image. They also have the ability to produce an uncertainty estimation for each pixel. One advantage of this technique is the instantaneous sample-and-hold nature of depth from stereo. In contrast, using a laser scanner that cannot complete a scan of the image in a single frame may result in distorted shapes as objects in the scene move. Any technique which depends on computations made with several frames sampled at different times, including the structured light method described in section 3.1, will have this problem.

Another real-time depth system has been developed by the Columbia University Automated Visual Environment Group [38]. They have demonstrated the ability to produce 512×480 depth estimates at 30 Hz with an accuracy of 0.3%. Their technique relies on a precise physical model of all the optical sensing and computational elements in the system: the optical transfer function, defocus, image sensing and sampling, and focus measure operators. They project a high frequency texture onto the scene and, via the same optical path, image the scene. An advantage of this system over the depth from stereo is that they do not have to worry about the correspondence problem faced by depth from stereo. One concern is the distracting high frequency textures which much be projected onto the scene. These patterns could prove unacceptable if the user wants to be in the environment while the scene is being captured.

2.2 Rendering

Our vision for the *office of the future* requires the ability to generate images that when projected onto the display surfaces appear correct to a moving head-tracked observer. This is true also for systems such as the CAVE™, but our situation is somewhat unusual in that we want to be able to project onto general surfaces whereas the CAVE™ system is tailored to planar surfaces. Future capabilities in image generation will allow the increased burden of display on arbitrary surfaces to be realized.

An interesting technique is presented in [16] for the use of computer graphics systems in theater design, where she models the appearance of backdrops from the audience's perspective. If left uncorrected, the backdrops would appear distorted. Essentially, we are faced with the same problem in our system, except with multiple projectors. We need to determine how to predistort the images such that, when projected from the projector's viewpoint, it will look correct from the user's viewpoint. Dorsey *et al.* also extend this technique to model the projector optics and demonstrates an extended radiosity method to simulate directional lighting characteristics.

2.3 Spatially Immersive Displays

The most well known spatially immersive display in the graphics community is probably the CAVE™ [12]. The CAVE™ exists in many forms, typically it is configured as a left, right, and rear wall rear projection system. In some implementations they use a mirror above the CAVE™ that projects an image onto the floor. While the

CAVE™ does provide head-tracked stereo views surrounding the user of the system, current implementations are limited to 1 projector displayed on each wall. The CAVE™ does not deal with intensity blending and has no method of capturing the geometry of the environment, which is reasonable since this was not an intended goal of their system.

The military simulation/flight simulator industry is full of numerous examples of spatially immersive displays [9, 23, 29, 32, 35]. These systems typical use CRT projectors which need frequent calibration. Also, they usually (but not always) restrict themselves to matching the seams of the display instead of considering the whole display area as something that needs to be blended seamlessly. Another technique of the flight simulator industry is to place a high resolution display in the center of view of the user and project a low resolution image on the surrounding screen, or to only project an image in the view frustum of the user. While this is effective, it cannot easily be repositioned and may show a seam where the high resolution image meets the low resolution image. The seam is a problem because it is disconcerting and severely disrupts the goal of achieving a feeling of being somewhere else—the user is always reminded they are looking at an imperfect display surface. The attempts at creating a seamless display are discussed in the previously cited flight simulator papers.

Domed displays are another example [2]. Such systems are often limited to only one high resolution projector and have rarely employed a mechanism to capture depth or projection surface information from the scene. A method is presented in [29] that corrects the warping of the dome by modeling a dome with a 5-degree polygon mesh and a GUI for manipulating the coordinates of this mesh, but this is not done in real-time or automatically: direct user intervention is required. This method is meant to only be used when the system is moved or for some infrequent reason falls out of alignment, it is not meant to be a method that can update the projection in real-time as the display surface changes shape or occlusion properties.

A final important point about all of these systems is that they rely on special geometric configurations and they present no general solution, which is a completely reasonable design decision on their part: typically they had an unchanging environment with uniform, ideal display surfaces. Also, they had control of every issue of the display system, from the lighting to the precise calibration of the projectors. Instead, we propose a general solution to the problem of projecting onto arbitrary display surfaces with real-time, automatic calibration procedures. Understand that we do not necessarily believe people will want to project on every type of surface or object, but we feel that thinking about the problem in this way is useful.

3 METHODS

In this section we describe the idea of using cameras and projectors that can be operated in either a capture or display mode. (See Figure 2 and Figure 8.) When in capture mode, the projectors and cameras can be used together to obtain per-pixel depth and reflectance information from the designated display surfaces. When in display mode, the image-based models may be used to render and then project geometrically and photometrically correct images onto the (potentially irregular) display surfaces.

3.1 Dynamic Image-Based Modeling

Image-based modeling is a difficult problem, one that has occupied the computer vision community for many years.

Depth Extraction

The depth extraction method for office scenes should work in large working volumes populated with areas with high frequency texture as well as surfaces that lack texture. To model display surfaces, we typically need high accuracy so that the projections are correct. To model dynamically changing scenes we need higher update rates to represent motion with potentially decreased resolution or accuracy. The method should be non-intrusive so that people can work in the environment. This prevents the use of lasers or other invasive methods.

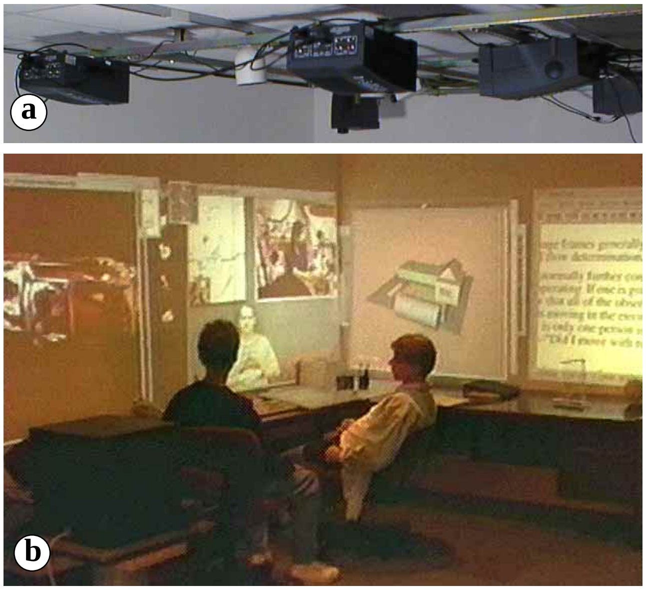

Figure 2: Our current implementation of the *office of the future*. (a) The cameras and digital light projectors are mounted near the ceiling using Unistrut material. (b) The walls consist of white foam-core board mounted on vertical posts. As shown in (b), they are used as display surfaces for a spatially-immersive display (SID).

Our system currently uses one video camera and one projector in a pair, although multiple cameras could work with one projector [43]. The correspondence problem is solved by projecting binary coded vertical bars [14]. The camera looks at a set of *n* successive images, creates binary images using adaptive thresholding on a per-pixel basis, and generates an *n*-bit code for every pixel corresponding to the code for the vertical bar that was imaged at this pixel. This allows us to distinguish between 2^n projected vertical bars. A pre-computed sparse triangulation lookup table based on calibration data allows trilinear interpolation to compute intersection of pseudo-camera rays and projected vertical planes. The 3D coordinates of the surface points imaged at every pixel are later used with color information to complete the image-based model. The vertical bars can be projected repeatedly to compute depth for dynamic scenes. The same method can be used for scanning walls, people or moving objects with different levels of sophistication for sub-pixel image thresholding. See Figure 3 for some still-frame sample results.

A choice of a camera and projector to actively scan 3D environments is necessitated by the fact that display surfaces may lack texture to provide enough correspondence cues. Use of the same projector for scanning and display allows unification of the two tasks and the only additional component is a camera. Speed versus accuracy trade-offs led us to two types of camera-projector pairs. Relatively static display surfaces such as walls and furniture in office scenes are modeled more accurately and slowly by the outward looking pairs of camera-projector than people and moving objects, which are scanned by inward looking pairs of camera-projectors.

The difficult part of using two separate devices for depth extraction is calibration. We use [45] first to find intrinsic and extrinsic parameters of the camera using a checkerboard pattern on

Figure 3: Some example results (still frames from live updates) of depth extraction using binary coded structured light.

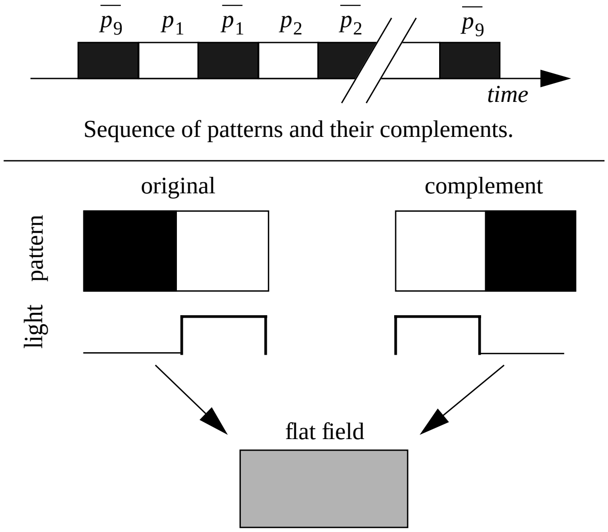

Figure 4: Pattern and complement are visually integrated over time, the result is the appearance of a flat field, or "white" light.

a flat surface. Then the same method is used to calibrate the projector with respect to the same flat surface. Combining the two gives the relationship between the camera and the projector. To find the relationship between two camera-projector pairs, the transformation between the two cameras is first determined by viewing a common checkerboard pattern on a flat surface. Then, using the method described above, the two pairs are calibrated with respect to the working volume. The procedure is easier if the frustums of the two cameras overlap considerably.

Detection of changes in scene geometry by camera image differencing is not robust when display surfaces lack texture. However, changes in a projected random texture can be imaged The random texture itself will be imperceptible to the human eye as described in section 3.2. Detected changes in scene geometry over a period of time could be the result of either actual changes in the surfaces or drift in the calibrated system.

The use of cameras allow the possibility to self-calibrate the system periodically to compensate for errors due to environmental factors such as temperature or vibrations in the setup.

Color and Reflectance

The projector is used in two modes, scene extraction and display. To get color information about the surfaces the projector is used as a bright light source along with a synchronized camera. However, the modes can be interleaved by inserting completely white frames in between the display frames. In the binary pulse-coded modulation (PCM) coded light projectors, only a few bits are used to project white frames while other bits can be used to project the display frames at reduced color resolution.

Currently we illuminate the scene with a black followed by white pattern and observe the resultant dark image and bright image from one view point to estimate the per-pixel reflectance function. The reflectance function is primarily used to threshold images of projected binary coded structured light patterns assuming the camera response is linear to intensity. Camera response curves can be estimated by illuminating the scene with different levels of intensity. To complete the image based model, surfaces in the scene can be sampled from multiple view points to estimate a bidirectional reflectance distribution (BRDF) function.

The camera is used for per-pixel depth extraction as well as color extraction. Since the two procedures share the same optical axis, there is no drift. Similarly, the same projector is used for projecting structured light patterns, for depth extraction, and for display on the depth extracted surface. This eliminating problems due to drift or misalignment.

3.2 Imperceptible Structured Light

With respect to the structured light described in section 3.1, our goal is to make it appear to the casual observer as nothing more than incandescent white light, not a succession of flashing binary patterns. Our method for doing this is to use *imperceptible structured light*. The approach is a combination of time-division multiplexing and light cancellation techniques to hide the patterns in a rapid series of white-light projections. Figure 4 depicts a sequence of patterns projected in time. A binary structured light approach such as ours uses n patterns to resolve 2^n projected vertical bars. The period of pattern repetition would be, for example, 1/60 second for a 60 Hz depth extraction rate. Figure 4 depicts how a given pattern p_i and complement $\overline{p_i}$ are integrated by the visual system in such a way that the sequence appears to be the projection of a flat field or white light. The same approach can be applied to project imperceptible structured light along with video or graphics images, facilitating imperceptible autocalibration of designated display surfaces. We use a digital light projector [25, 26] which uses PCM to project the pattern and its complement. A synchronized camera can measure the structured light by integrating light during the pattern projection.

While limited access to digital light projector specifications currently limits our ability to implement completely imperceptible image-based modeling, we are able to separately demonstrate real-time capture and imperceptible structured light. Figure 5 shows the effect in a laboratory experiment. Figure 6 shows the use of a similar approach to embed imperceptible structured light in a still image as opposed to white light. We are working together with the developers of the projector technology to obtain lower-level access to the technology, which will introduce a whole new realm of possibilities for dynamic structured light and display.

3.3 Geometric Registration

The depth data acquired from different cameras needs to be zipped together to complete the display surface geometry. This is required for generating correct images from a single projector for multisurfaces [42] and also for intensity blending if the projections from multiple projectors overlap. The depth images are taken from distinct cameras and are in different coordinate systems. Thus in order to tile the extracted surfaces together corresponding points

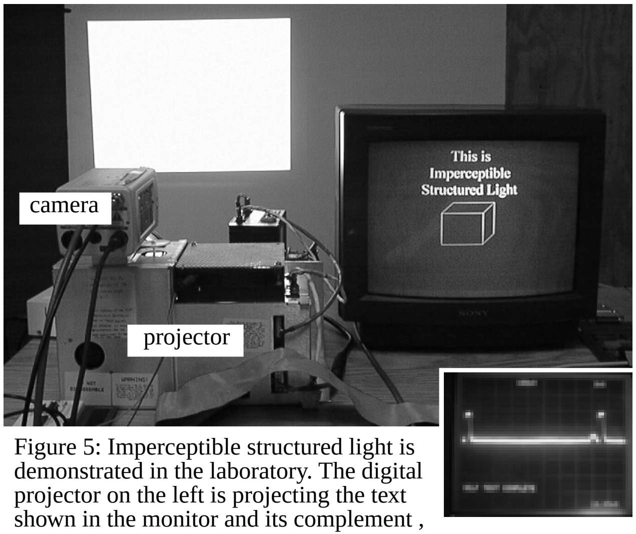

Figure 5: Imperceptible structured light is demonstrated in the laboratory. The digital projector on the left is projecting the text shown in the monitor and its complement , however the text can only be seen with a synchronized camera such as that sitting on the projector above. The inset snapshot of an oscilloscope shows the pulses that correspond to the brief time when the pattern (text in this case) is being projected.

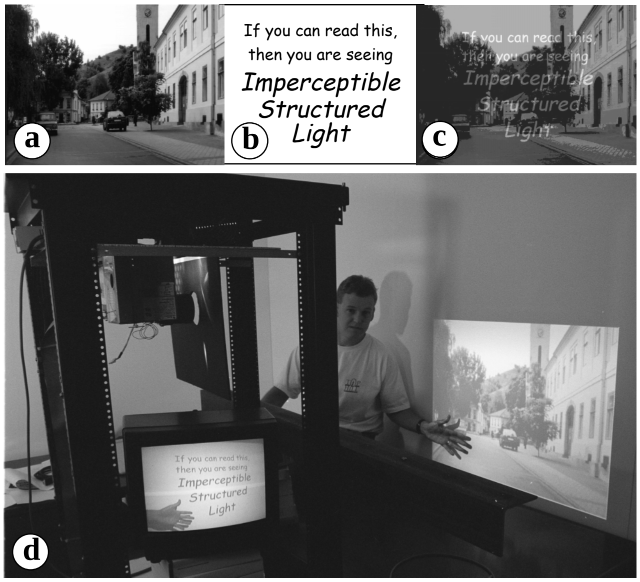

Figure 6: Imperceptible structured light embedded in images. (a) An initial image (Tokaj, Hungary). (b) The binary image that is to be imperceptible. (c) The two images combined and mapped to the proper time-division bit sequence. (d) The final result, showing the initial image (with reduced dynamic range) being projected on the wall, while the embedded imperceptible image is captured and displayed on the monitor (lower left).

from overlap regions are used. The corresponding points are generated using the binary coded structured light approach for rows and columns of projector pixels. The binary code of an imaged point uniquely identifies the corresponding projector pixel. Pairs of pixels in two cameras that share the binary code are used to compute transformation between depth data sets. Note that this transformation between two cameras can also be pre-computed during calibration stage but is usually not sufficiently accurate to register two depth data sets. Otherwise, for blending purposes, the geometric correspondence between pixels of two different projectors is established by observing the projection overlap region from a single camera. We assume that every pair of images has a substantial overlap (about one-third of its total area).

3.4 Rendering and Display

Our goal is to generate images that appear correct to an observer when projected onto (potentially irregular) display surfaces. Since the observer can move around in the office, we currently use magnetic head-tracking to determine the viewer's location. The inputs to the algorithm are a model of the surface, the projector's intrinsic and extrinsic parameters, the viewer's location, and a "desired image," the image which we want the viewer to see. The desired image will typically be the result of conventional 3-D rendering.

Our algorithm can work with any general type of surface representation (e.g. NURBs), as long as the model of the real-world display surface is accurate. Likewise, rendering could be done with many different methods (e.g. ray-tracing); our current implementation uses projective textures with OpenGL primitives to achieve hardware acceleration. The underlying projective textures [44] technique is an extension of perspectively-correct texture mapping that can be used to do arbitrary projection of two dimensional images onto geometry in real-time.

We describe a two pass approach for rendering and displaying images of 3D scenes on potentially irregular surfaces. In the first pass, we compute the "desired image" for the viewer by rendering the 3D scene from the observer's viewpoint. This desired image is stored as a texture map. In the second pass the texture is effectively projected from the user's viewpoint onto the polygonal model of the display surface. The display surface (with the desired image texture mapped onto it) is then rendered from the projector's viewpoint. The resulting image, when displayed by the projector, will produce the desired image for the viewer. As the user moves, the desired image changes and it is also projected from the user's new location.

Multiple projectors can be used to increase the display surface area. To ensure complete coverage of the display surface, every part of the display surface must be visible to at least one projector. To ensure complete coverage from a given viewpoint, at least one projector must be able to image on every surface visible from that viewpoint. The projectors' viewpoints in the second pass typically remains fixed. Neither projector overlap nor self-occlusions of display surfaces from observer's viewpoint need hinder the effective image from the user's viewpoint. See Figure 7.

To specify the viewing direction for projecting textures with monocular viewing, we only need the position of the user and not orientation. A field of view that contains all the polygons of the synthetic object is sufficient for the frustum of texture projection. This frustum may be trimmed if it exceeds the frustum of the display surface model.The frustum is oriented from the viewer's location toward the polygonal model of display surfaces. The user's frustum parameters during the first pass and texture projection in the second pass are identical.

We assume that the projectors have no radial distortion and hence the projectors can be modeled with a pinhole camera. For the optics of digital micromirror device (DMD) projectors, this assumption is valid. However, if the projector has radial distortion, we must pre-distort the rendered image before it is sent to the projector framebuffer. This pre-distortion can be done using non-linear 3D warp of display surface geometry or using screen space 2D warp with texture mapping.

Challenges

Speed.

The two pass rendering method consists of normal 3D rendering in the first pass, followed by a second pass that maps the desired image to a display surface model. The additional cost of the algorithm comes from transferring the framebuffer from the first pass into texture memory and rendering the display surface model with texture mapping applied in the second pass. Thus it is crucial to simplify the display surface geometries.

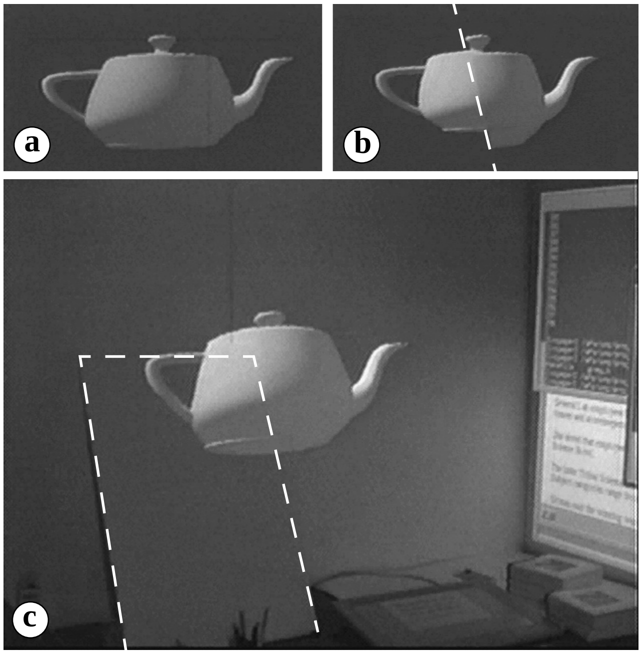

Figure 7: Multisurface rendering. (a) A teapot projected onto a single planar surface. (b) A distorted projection resulting from the introduction of a second planar display surface. (c) The final corrected projection obtained by extracting the display surface geometry and then employing our two-pass rendering scheme. The edges of the second planar surface—shown leaning against the wall on the desk in (c)—are highlighted with a dashed line.

Parallelization

If more than one projector is used, each projector can have a separate display engine. Rendering can be done in parallel, and each display engine need only load those parts of the display surface model that are visible from the corresponding projector. We have some initial evidence [42] that our method is faster versus conventional rendering techniques with multiple projectors, complex displays, or complex graphics models. The reason is that the first pass of the algorithm (conventional rendering) only needs to be done once, and then the second pass can be performed for each projector. Conventional techniques require the graphics model to be re-rendered for each projector or even for each polygon in the display surface.

If the first pass is also parallelized, all the corresponding graphics pipelines need to access the synthetic model being displayed simultaneously. This could be a problem if the model is dynamically changing and the graphics pipeline must read the model during every display iteration. Other parallelization issues include if the display surface model is dynamic, as well as network issues if the display engines are on different machines.

Latency

There is inherent latency in the system in addition to traditional tracking and rendering latency due to two pass method for drawing models. For large models, rendering times could be different for different projectors so that there is inter-projector delay during rendering. If all the projectors are driven from a single machine, then setting up viewports within a single rendering program for each projector and synchronously updating the framebuffers will eliminate this problem.

3.5 Generating Blending Functions

We use the traditional computer graphics phrase “alpha values” to describe the blending functions. When building some sort of a tiled multi-projector SID, one is faced with two approaches for handling the transitions between the projected images: one can either design the system such that the images do not overlap but can be adjusted so that they are barely touching and thus “seamless,” or one can allow projected images to overlap and employ some means of blending. The second approach typically uses a roll-off function such as a linear ramp or a cosine curve to smooth the transition between projectors.

Designers of the CAVE™ exercise the first option by limiting the system to a well-defined, relatively simple screen arrangement whereby no projectors overlap. However we want to be able to project images onto arbitrary potentially irregular display surfaces in the office, which means that we cannot use the first approach as we assume no control over the surfaces. Furthermore, we envision a more flexible setup whereby multiple projectors can be used to project into the same space in order to achieve higher resolution (e.g., via a “high-resolution insert”) or increased light.

We implemented a weighting function by assigning alpha values between 0 and 1 to every pixel in every projector, and as described in section 3.3 ensure that every illuminated world point corresponding to a single camera pixel has an alpha sum equal to one. This assumes that the projectors have similar intensity response. There are two cases: a point resulting from a projection of only one projector and a point resulting from a number of projectors. In the first case the solution is trivial. However, in the second case, the case where overlap occurs, we make alpha values a function of the distance to the beginning/end of overlapping region, with the constraint that alpha values of points from different projectors, corresponding to the same point in space, must sum up to one. To assign different weights to projector pixels, we actually create an alpha image for each projector. This image contains (1 - *desired_alpha*) at each pixel. This alpha image is rendered last. In our OpenGL implementation this is achieved using transparent textures. A camera in the closed loop system allows one to photometrically correct images even when the projectors have different brightness or when the display surface has non-uniform reflectance properties. Although the digital light projectors have linear intensity response, they use a de-gamma correction [25]. Use of alpha image allows us to compensate the de-gamma correction.

3.6 Simplification of Depth Data

Dynamic image-based modeling of an entire office will result in tremendously large data sets, given that the data would be per-pixel for multiple cameras, occurring at video rates. However it is reasonable to expect that the majority of the data is highly correlated both temporally and spatially. In fact most of the office is unlikely to change dramatically over short periods of time, in particular this is likely to be true for the office walls and most designated display surfaces. It makes sense to attempt to simplify the data so that the system does not have to deal with such a horrendous volume of data. For example, Radim Sara and Ruzena Bajcsy at the University of Pennsylvania have created a depth data set of an office that has approximately half a million vertices. The simplification method must be careful not to simplify in regions of rapid change or high curvature where information might be lost. The automatic reconstruction of surfaces from range data is explored in [13, 1, 8, 11, 17, 22].

Unfortunately, the dynamic nature and the presence of noise in our system, disallow the use of well-established simplification algorithms. The method we currently employ is not a well-defined mathematical approach, but rather a heuristic-based method that produced qualitatively pleasing results based on the characteristics of our data sets. We first apply a curvature approach to the data set using a tangent method similar to [24], and we then use a Euclidean distance approach on the remaining points. We chose this particular sequence of steps because the curvature method is usually much more successful in eliminating points than the second one. This approach produces elimination rates of 80% to 90% without any visible loss of information. This is because most of the objects in the office environment are locally planar.

3.7 Tracking

While our office of the future could certainly and possibly very effectively be used in a 2D-only mode, we believe that it is more compelling to consider its additional use as a 3D visualization environment. We need the ability to track viewers' heads in order to render perspectively correct images. Interestingly enough, for monoscopic viewing one does not need the orientation of the eye because the display image is uniquely determined by the eye position. For stereoscopic viewing one needs to be able to either track one eye and the user's head orientation, or two eyes, each with position only. The system involves projecting synthetic images onto real surfaces for which the extracted surface model is assumed to be correct in world space. Any small error in tracker reading (after transformation) in world space will result in visibly incorrect registration between projected images and the display surfaces. This situation is similar to augmented reality systems where traditionally a vision based registration system is used to correct the tracker readings. A similar closed loop system may be necessary for accurate rendering for our system.

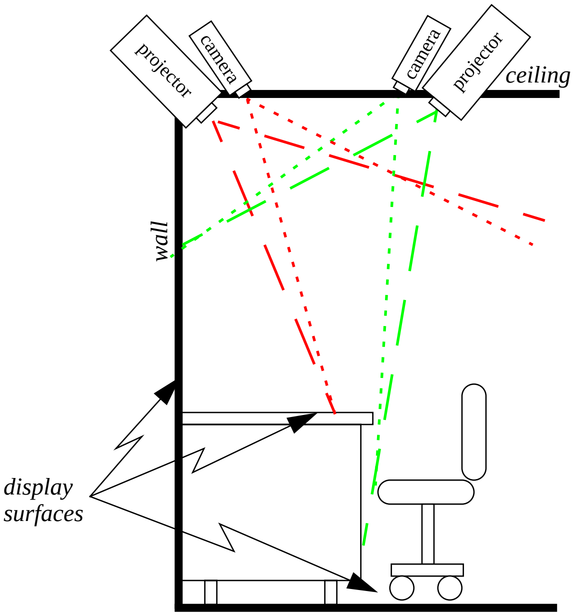

Figure 8: Digital projectors and cameras work together to capture depth, color, and surface reflectance information for objects and participants in the scene. A subset of the projectors is also used for display purposes; captured display surface depth, color, and reflectance information can be used to autocalibrate the display surfaces so that projected images are geometrically and photometrically correct from the viewer's viewpoint, and so overlapping projections are properly blended

4 CURRENT IMPLEMENTATION

While a complete realization of such an office of the future is, by definition, years away, we are making steady progress and present here some promising results and demonstrations. We have implemented a working system using projective textures. We have also independently demonstrated 1) a depth extraction system running at 3 Hz, 2) imperceptible structured light, and 3) initial experiments in intensity blending.

The office size is $10 \times 10 \times 10$ feet, and is populated with five 800×600 resolution *digital light projectors* and two 640×480 resolution video cameras (Figure 2). The outward looking five projectors are driven simultaneously from an SGI Infinite Reality. Four of these project on relatively vertical surfaces and one projects down on the table and the floor. The binary PCM coded light projectors project 24 bit color at a 60Hz update rate. The projectors allow off-axis projection with a small offset without significant focus problems. The camera framegrabbers are SGI O2s and currently we use a video switcher to switch between the video cameras while capturing images of the environment. We expect that in the very near future the rendering will be done in different viewports of a single window and hence can be used to maintain synchronous updates of all projector framebuffers. The system also includes an Ascension magnetic tracker for tracking user position.

The walls of the office are made up of relatively inexpensive foam-core board, and does not need solid support because the system supports non-planar surfaces and off-axis projection. A separate projector-camera pair setup is used to create dynamic image based model in a small working volume of $3 \times 3 \times 2$ feet, and can be used for dynamic image-based modeling of human faces (see Figure 3) [43]. The system creates a 640×240 depth map at three updates per second, which is texture mapped with live video. Using a direct digital interface to the digital light projectors from a PC, we have been able to project patterns that are imperceptible to human eye but visible from a synchronized camera (see Figure 5). We have capability to change the binary PCM coding of light for the projectors allowing us to use different bits for different purposes; in particular we have the ability to burn an equal-bit-time PROM for the projectors which allows us to demonstrate compelling imperceptible structured light (see Figure 5). With the equal-bit-time PROM, a binary pattern can be displayed 24 times per 60 Hz frame, i.e. every 694 microseconds. Thus a pattern and its complement can be displayed sequentially in approximately 1.4 milliseconds. The synchronized camera with exposure time less than 700 microsecond was used in Figure 5. Using this PROM we could in theory project and capture $60 \times 24 = 1440$ binary patterns per second. Our current framegrabbers, however, can process only 60 images per second. A better digital interface to DLP's will also allow us to render stereo images at 60 Hz.

Although we expect the participants to be seating in a chair most of the time (Figure 2), the current setup allows participants of average height (under 6 feet) to stand and move around without blocking the projection on the walls if they are at least 4 feet away from the walls.

The office of the future setup allows scalability in terms of more pairs of camera and projector to either increase resolution of extracted surfaces, or resolution of display on surfaces. The system other than computer hardware costs approximately $35,000. We expect minimal maintenance of projector, cameras or display surfaces because the system employs self-calibration methods.

5 FUTURE WORK

Much work remains to be done, some of which we have concrete plans to attack, some we are attacking with collaborations, and some we hope others will pursue.

We plan to integrate scene acquisition and display in such a way that the acquisition is imperceptible, or at least unobtrusive. This will involve some combination of light control and cameras, possibly wide-field-of-view high-resolution clusters as described in [7]. Together with our collaborators in the GRASP Laboratory at the University of Pennsylvania, we are exploring the continuum of options between strict control of all of the lights in the environment (as outlined herein) and little or no control of the lights but using multiple cameras and passive correlation-based techniques. We expect to have within the coming year a new multibaseline correlation system on hand for experiments with our structured light acquisition and display environment.

As part of scene acquisition, one can detect display surface changes and adapt the rendering accordingly. Currently it can be done at non-interactive rates. Eventually we also want to explore methods as in [21] to detect surface changes for purposes such as tracking and gestural input.

We also want to improve image generation with better blending, by exploiting image-based rendering methods to construct a target image from multiple reference images. There is a good possibility that a distributed rendering scheme employing the multiprojector and multisurface display algorithms that we present and analyze in [42] will prove to be effective. In addition, we want to correct for surface reflectance discontinuities *dynamically*, and to make use of the information during run-time to adjust the rendered images.

We are planning to use our system in an on-going telecollaboration involving multi-disciplinary mechanical design and manufacturing with our collaborators in the NSF Science and Technology Center for Graphics and Visualization. In addition as part of The Tele-Immersion Initiative we are planning to make use of the CAVE™ library or similar framework to connect several laboratories over high speed networks with novel immersive display environments.

6 SUMMARY AND CONCLUSIONS

We have shown initial results for a novel semi-immersive display in an office-like environment, one that combines acquisition and display. We have developed techniques to acquire the geometry of an irregular surface and then modify rendering to allow projection onto that irregular surface so that it looks correct to an observer at a known location. We have described a method of injecting structured light into a scene that is imperceptible to the participants but measurable to synchronized cameras. These techniques can be applied to other display environments which use multiple projectors or that involve complex display geometries.

In conclusion, we note that a major trend in computer science over the past few decades has been from one to many, from being restricted by resources' proximity to employing resources irrespective of their locations. One field unaffected by this global development has been the computer display or the area where the results of our work are being presented. Our system pushes this envelope, thus enabling any object, or a collection of such, located anywhere to be used as a display surface. From now on, one does not have to cramp the information into a relatively small monitor, but to have as much space as possible and to be limited only by the amount of space around. Anything can be a display surface - a wall or a table, and anywhere - be it an office or a conference hall. Of course, the system faces many challenges, but they can be overcome by the increasing power of graphics hardware and general purpose computing.

7 ACKNOWLEDGMENTS

This work was supported by (1) the National Science Foundation Cooperative Agreement no. ASC-8920219: “Science and Technology Center for Computer Graphics and Scientific Visualization”, Center Director Andy van Dam (Brown University). Principal Investigators Andy van Dam, Al Barr (California Institute of Technology), Don Greenberg (Cornell University), Henry Fuchs (University of North Carolina at Chapel Hill), Rich Riesenfeld (University of Utah); (2) the “National Tele-Immersion Initiative” which is sponsored by Advanced Networks and Services, President and Chief Executive Officer Al Weis, Chief Scientist Jaron Lanier; and (3) DARPA grant no. N66001-97-1-8919: “Three Applications of Digital Light Projection for Tiled Display and 3-D Scene Capture.”

We thank Nick England for sharing his significant knowledge about wide-area tiled display systems, both in terms of past work and fundamental issues; Mary Whitton, David Harrison, John Thomas, Kurtis Keller, and Jim Mahaney for their help in designing, arranging, and constructing the prototype office of the future; Todd Gaul for help with our video taping; Jean-Yves Bouguet of Caltech for the camera calibration code and useful discussions; Nick Vallidis for his work on creating new DLP binary PCM coding; Andy Ade and Jai Glasgow for their administrative help; our department’s Computer Services staff members for keeping our networks and machines humming in spite of our experimental modifications; and Hans Weber and Rui Bastos for help photographing the imperceptible structured light demonstrations. Finally, we gratefully acknowledge Andrei State for his illustrations of the office of the future (Figure 1).

References

[1] Bajaj, C.L., F. Bernardini, and G. Xu. “Automatic reconstruction of surfaces and scalar fields from 3D scans,” SIGGRAPH 95 Conference Proceedings, Annual Conference Series, ACM SIGGRAPH, Addison-Wesley, pp. 109-118, August 1995.

[2] Bennett, David T. Chairman and Co-Founder of Alternate Realities Corporation. Internet: http://www.virtual-reality.com. 215 Southport Drive Suite 1300 Morrisville, NC 27560.

[3] Bowen, Loftin, R. “Hands Across the Atlantic,” IEEE Computer Graphics and Applications, Vol. 17, No. 2, pp. 78-79, March-April 1997.

[4] Bryson, Steve, David Zeltzer, Mark T. Bolas, Bertrand de La Chapelle, and David Bennett. “The Future of Virtual Reality: Head Mounted Displays Versus Spatially Immersive Displays,” SIGGRAPH 97 Conference Proceedings, Annual Conference Series, ACM SIGGRAPH, Addison-Wesley, pp. 485-486, August 1997.

[5] Buxton, W. 1992. “Telepresence: integrating shared task and person spaces,” Proceedings of Graphics Interface '92, 123-129. Earlier version appears in Proceedings of Groupware '91, Amsterdam, Oct. 29, 1991, 27-36..

[6] Capin, Tolga K., Hansrudi Noser, Daniel Thalmann, Igor Sunday Pandzic and Nadia Magnenat Thalman. “Virtual Human Representation and Communication in VLNet,” IEEE Computer Graphics and Applications, Vol. 17, No. 2, pp. 42-53, March-April 1997.

[7] Chi, Vern, Matt Cutts, Henry Fuchs, Kurtis Keller, Greg Welch, Mark Bloomenthal, Elaine Cohen, Sam Drake, Russ Fish, Rich Riesenfeld. 1998. “A Wide Field-of-View Camera Cluster”, University of North Carolina at Chapel Hill, Dept of Computer Science, Technical Report TR98-018.

[8] Chien, C.H., Y.B. Sim, and J.K. Aggarwal. “Generation of volume/surface octree from range data,” In The Computer Graphics Society Conference on Computer Vision and Pattern Recognition, pp. 254-260, June 1988.

[9] Clodfelter, Robert M. “Predicting Display System Performance,” Presented at the 1996 IMAGE Conference, Scottsdale, AZ, pp. 1-5, June 23-28, 1996.

[10] Conner, D.B, Cutts, M., Fish, R., Fuchs, H., Holden, L., Jacobs, M., Loss, B., Markosian, L., Riesenfeld, R., and Turk, G. “An Immersive Tool for Wide-Area Collaborative Design,” TeamCAD, the First Graphics Visualization, and Usability (GVU) Workshop on Collaborative Design. Atlanta, Georgia, May 12-13, 1997.

[11] Connolly, C.I. “Cumulative generation of octree models from range data,” Proceedings, Int’l. Conference Robotics, pp. 25-32, March 1984.

[12] Cruz-Neira, Carolina, Daniel J. Sandin, and Thomas A. DeFanti. “Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE,” Computer Graphics, SIGGRAPH Annual Conference Proceedings, 1993.

[13] Curless, Brian, and Marc Levoy. “A Volumetric Method for Building Complex Models from Range Images,” SIGGRAPH 96 Conference Proceedings, Annual Conference Series, ACM SIGGRAPH, Addison-Wesley. pp. 303-312, 1996.

[14] DePiero, F.W., and Trivedi, M.M., “3-D Computer Vision using Structured Light: Design, Calibration, and Implementation Issues,” Advances in Computers(43), 1996, Academic Press, pp.243-278

[15] Dias, José Miguel Salles, Ricardo Galli, António Carlos Almeida, Carlos A. C. Belo, and José Manuel Rebordã. “mWorld: A Multiuser 3D Virtual Environment,” IEEE Computer Graphics and Applications, Vol. 17, No. 2., pp. 55-64, March-April 1997.

[16] Dorsey, Julie O’B., Fransco X. Sillion, Donald Greenberg. “Design and Simulation of Opera Lighting and Projection Effects,” SIGGRAPH 91 Conference Proceedings, Annual Conference Series, ACM SIGGRAPH, Addison-Wesley, pp. 41-50, 1991.

[17] Edelsbrunner, H. and E.P. Mucke. “Three-dimensional alpha shapes,” In Workshop on Volume Visualization, pp. 75-105, October 1992.

[18] Fuchs, Henry, Gary Bishop, Kevin Arthur, Leonard McMillan, Ruzena Bajcsy, Sang Lee, Hany Farid, and Takeo Kanade. “Virtual Space Teleconferencing Using a Sea of Cameras,” Proceedings of the First International Symposium on Medical Robotics and Computer Assisted Surgery, (Pittsburgh, PA.) Sept 22-24, 1994.

[19] Gajewska , Hania , Jay Kistler, Mark S. Manasse, and David D. Redell. “Argo: A System for Distributed Collaboration,” (DEC, Multimedia '94)

[20] Gibbs, Simon, Constantin Arapis and Christian J. Breiteneder. “TELEPORT-Towards Immersive Copresence,” accepted for publication in ACM Multimedia Systems Journal, 1998.

[21] Goncalves, Luis , Enrico Di Bernardo, Enrico Ursella, Pietro Perona. “Monocular Tracking of the Human Arm in 3D,” Proc. of the 5th Inter. Conf. on Computer Vision, ICCV 95.

[22] Hilton, A., A.J. Toddart, J. Illingworth, and T. Windeatt. “Reliable surface reconstruction from multiple range images,” In Fouth European Conference on Computer Vision, Volume 1, pp. 117-126. April 1996.

[23] Holmes, Richard E. “Common Projector and Display Modules for Aircraft Simulator Visual Systems,” Presented at the IMAGE V Conference, Phoenix, AZ, June 19-22, pp. 81-88, 1990.

[24] Hoppe, Hugues, Tony DeRose, Tom Duchamp, John McDonald, Werner Stuetzle. “Surface Reconstruction from Unorganized Points,” SIGGRAPH 92 Conference Proceedings, Annual Conference Series, ACM SIGGRAPH, Addison-Wesley, pp. 71-76, 1992.

[25] Hornbeck, Larry J., “Deformable-Mirror Spatial Light Modulators,”Proceedings SPIE, Vol. 1150, Aug 1989.

[26] Hornbeck, Larry J., “Digital Light Processing for High-Brightness High-Resolution Applications,” [cited 21 April 1998]. Available from http://www.ti.com/dlp/docs/business/resources/white/hornbeck.pdf, 1995.

[27] Ichikawa, Y., Okada, K., Jeong, G., Tanaka, S. and Matsushita, Y.: “MAJIC Videoconferencing System: Experiments, Evaluation and Improvement',” In Proceedings of ECSCW'95, pp. 279-292, Sept. 1995.

[28] Ishii, Hiroshi, Minoru Kobayashi, Kazuho Arita. “Iterative Design of Seamless Collaboration Media,” CACM, Volume 37, Number 8, pp. 83-97, August 1994.

[29] Jarvis, Kevin. “Real-time 60Hz Distortion Correction on a Silicon Graphics IG,” Real-time Graphics, Vol. 5, No. 7, pp. 6-7. February 1997.

[30] Kanade, Takeo and Haruhiko Asada. “Noncontact Visual Three-Dimensional Ranging Devices,” Proceedings of SPIE: 3D Machine Perception. Volume 283, Pages 48-54. April 23-24, 1981.

[31] Kanade, Takeo, Hiroshi Kano, Shigeru Kimura, Atsushi Yoshida, Kazuo Oda. “Development of a Video-Rate Stereo Machine,” Proceedings of International Robotics and Systems Conference (IROS ’95). pp. 95-100, Pittsburgh, PA., August 5-9, 1995.

[32] Lacroix, Michel. “A HDTV Projector For Wide Field of View Flight Simulators,” pp. 493-500. Presented at the IMAGE VI Conference, Scottsdale, AZ, July 14-17, 1992.

[33] Lamotte, Wim, Eddy Flerackers, Frank Van Reeth, Rae Earnshaw, Joao Mena De Matos. Visinet: Collaborative 3D Visualization and VR over ATM Networks. IEEE Computer Graphics and Applications, Vol. 17, No. 2, pp. 66-75, March-April 1997.

[34] Lehner, Valerie D., and Thomas A. DeFanti. “Distributed Virtual Reality: Supporting Remote Collaboration in Vehicle Design,” IEEE Computer Graphics and Applications, Vol. 17, No. 2, pp. 13-17, March-April 1997.

[35] Lyon, Paul. “Edge-blending Multiple Projection Displays On A Dome Surface To Form Continuous Wide Angle Fields-of-View,” pp. 203-209. Proceedings of 7th I/ITEC, 1985.

[36] Macedonia, Michale R. and Stefan Noll. “A Transatlantic Research and Development Environment,” IEEE Computer Graphics and Applications, Vol. 17, No. 2, pp. 76-82, March-April 1997.

[37] Mandeville, J., T. Furness, M. Kawahata, D. Campbell, P. Danset, A. Dahl, J. Dauner, J. Davidson, K. Kandie, and P. Schwartz. “GreenSpace: Creating a Distributed Virtual Environment for Global Applications,” Proceedings of IEEE Networked Virtual Reality Workshop, 1995.

[38] Nayar, Shree, Masahiro Watanabe, Minori Noguchi. “Real-time Focus Range Sensor,” Columbia University, CUCS-028-94.

[39] Neumann, Ulrich and Henry Fuchs, "A Vision of Telepresence for Medical Consultation and Other Applications," Proceedings of the Sixth International Symposium on Robotics Research, Hidden Valley, PA, Oct. 1-5, 1993, pp. 565-571.

[40] Ohya, Jun, Kitamura, Yasuichi, Takemura, Haruo, et al. “Real-time Reproduction of 3D Human Images in Virtual Space Teleconferencing,” IEEE Virtual Reality International Symposium. Sep 1993.

[41] Panorama Project, WWW description. [cited 13 January 1998]. Available from http://www.tnt.uni-hannover.de/project/eu/panorama/overview.html

[42] Raskar, Ramesh, Matt Cutts, Greg Welch, Wolfgang Stüerzlinger. “Efficient Image Generation for Multiprojector and Multisurface Displays,” University of North Carolina at Chapel Hill, Dept of Computer Science, Technical Report TR98-016, 1998.

[43] Raskar, Ramesh, Henry Fuchs, Greg Welch, Adam Lake, Matt Cutts. “3D Talking Heads : Image Based Modeling at Interactive rate using Structured Light Projection,”University of North Carolina at Chapel Hill, Dept of Computer Science, Technical Report TR98-017, 1998

[44] Segal, Mark, Carl Korobkin, Rolf van Widenfelt, Jim Foran and Paul Haeberli, “Fast Shadows and Lighting Effects using Texture Mapping,” Computer Graphics (SIGGRAPH 92 Proceedings), pp. 249-252, July, 1992.

[45] Tsai , Roger Y. “An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision,” Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, pp. 364-374, 1986.

[46] Underkoffler, J. “A View From the Luminous Room,” Personal Technologies, Vol. 1, No. 2, pp. 49-59, June 1997.

[47] Underkoffler, J., and Hiroshi Ishii. “Illuminating Light: An Optical Design Tool with a Luminous-Tangible Interface,” Proceedings of CHI '98, ACM, April 1998.