COMPAQ

Shadow Elimination and Occluder Light Suppression for Multi-Projector Displays

Tat-Jen Cham   Rahul Sukthankar   James M. Rehg   Gita Sukthankar

Cambridge Research Laboratory

Cambridge Research Laboratory

Technical Report Series

CRL 2002/03

March 2002

Cambridge Research Laboratory

The Cambridge Research Laboratory was founded in 1987 to advance the state of the art in both core computing and human-computer interaction, and to use the knowledge so gained to support the Company's corporate objectives. We believe this is best accomplished through interconnected pursuits in technology creation, advanced systems engineering, and business development. We are actively investigating scalable computing; mobile computing; vision-based human and scene sensing; speech interaction; computer-animated synthetic persona; intelligent information appliances; and the capture, coding, storage, indexing, retrieval, decoding, and rendering of multimedia data.

Robert A. Iannucci, Ph.D.
Vice-President of Research

Shadow Elimination and Occluder Light Suppression for Multi-Projector Displays

Tat-Jen Cham[1] Rahul Sukthankar[2,3] James M. Rehg[4] Gita Sukthankar[2,3]

[1]School of Computer Engineering, Nanyang Technological University, Singapore 639798

[2]Compaq Research (CRL), One Cambridge Center, Cambridge, MA 02142

[3]The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213

[4]College of Computing, Georgia Institute of Technology, Atlanta, GA 30332

March 2002

Abstract

Two related problems of front projection displays which occur when users obscure a projector are: (i) undesirable shadows cast on the display by the users, and (ii) projected light falling on and distracting the users. This paper provides a computational framework for solving these two problems using multiple overlapping projectors and cameras. The overlapping projectors are automatically aligned to display the same dekeystoned image. The system detects when and where shadows are cast by occluders and is able to determine the pixels which are occluded in different projectors. Through a feedback control loop, the contributions of unoccluded pixels from other projectors are boosted in the shadowed regions, thereby eliminating the shadows. In addition, pixels which are being occluded are blanked, thereby preventing the projected light from falling on a user when they occlude the display. This can be accomplished even when the occluders are not visible to the camera. The paper presents results from a number of experiments demonstrating that the system converges with low steady-state errors.

CRL Technical reports are available on the CRL's web page at
http://crl.research.compaq.com.

Compaq Computer Corporation
Cambridge Research Laboratory
One Cambridge Center
Cambridge, Massachusetts 02142 USA

1 Introduction

The increasing affordability and portability of high quality projectors has generated a surge of interest in projector-camera systems. Recent examples include the construction of seamless multi-projector video walls [14, 4, 6, 13], real-time range scanning [3] and immersive 3-D virtual environment generation [9]. In most of these previous systems, cameras are used to coordinate the aggregation of multiple projectors into a single, large projected display. In constructing a video wall, for example, the geometric alignment and photometric blending of overlapping projector outputs can be accomplished by using a camera to measure the keystone distortions in projected test patterns and then appropriately pre-warping the projected images. The result is a highly scalable display system, in contrast to fixed format displays such as plasma screens.

While scalable, projector-based displays are quite attractive for visualization applications, interacting with a projected light display currently involves some unpleasant trade-offs. Rear-projection systems support seamless interaction (e.g. the Xerox LiveBoard), but can be prohibitively expensive and difficult to deploy because of the need for significant space behind the screen. In comparison, forward projection systems are inexpensive and flexible. However, any interaction with the display results in shadows being cast onto the screen while the projected light falls on the user.

Recent display systems have demonstrated the ability to compensate for shadows by fusing the outputs of multiple verged (i.e. redundant) projectors [11, 5]. Unfortunately, these previous systems did not have the ability to determine *which* projected rays were being blocked. As a result, they could not prevent projected light from falling on and distracting the user. In this paper, we describe a new algorithm for multi-projector fusion which not only eliminates shadows but also removes projected light from the user, making it possible to *completely* simulate the user experience of a rear-projection system using forward-projection technology.

In addition to the development of interactive wall-sized displays, we also want to incorporate forward projection systems into standard physical environments such as classrooms, living rooms, or kitchens. In these settings, a projector-camera system could create ubiquitous, interactive displays using the ordinary visible surfaces in a person's environment. Displays could be conveniently located on tabletops, nearby walls, etc. Users could reposition or resize them using simple hand gestures. Displays could even be "attached" to objects in the environment or be made to follow a user around as desired. This could be particularly compelling as means to augment the output capabilities of handheld devices such as PDAs. In order to realize this vision, two challenging sensing problems must be solved: (1) Create stable displays in the presence of environmental disturbances such as changing ambient light and occlusions by the users. (2) Determine where and when to create displays based on user activities.

In this paper we examine the novel visual sensing challenges that arise in creating stable occlusion-free displays using projected light in real-world environments. In particular, we address the two problems that arise in front-projection systems when a user passes between the projectors and the display surface: (1) Shadows cast on the display surface due to the occlusion of one or more projectors by the user, (2) Bright light projected on the user, which is often a source of distraction and discomfort. We demonstrate that these problems can be solved without accurate 3-D localization of projectors, cameras, or occluders, and without accurate photometric calibration of the display surface. The key is a display-centric camera feedback loop that rejects disturbances and unmodeled effects.

Our system uses multiple, conventional projectors which are positioned so that their projections overlap on the selected display surface.[1] It produces shadow-free displays even in the presence of multiple, moving occluders. Furthermore, projector light cast on the occluders is suppressed without affecting the quality of the display. The result is shown in Figure 5.

2 System Overview

Our system comprises a number of projectors which are aimed at a screen such that their projection regions overlap and a camera which is positioned such that can view the entire screen. During normal functioning, the system displays a high quality, dekeystoned image on the screen. When users walk between the projectors and the screen, shadows are cast on the screen. These shadows can be classified as *umbral* when all projectors are simultaneously occluded, or *penumbral* when at least one projector remains unoccluded. The system eliminates all penumbral shadows cast on the screen,[2] as well as suppressing projector light falling on the occluders. This enables the system to continue presenting a high

[1] In the event that not enough fixed projectors are available for a given environment, projector-mirror systems such as [8] could be used.

[2] By definition, pixels in an umbral shadow are blocked from every projector and cannot be removed. Umbral shadows can be minimized by increasing the number of projectors and by mounting the projectors at highly-oblique angles.

quality image without projecting distracting light on users. See Figure 1 for the setup.

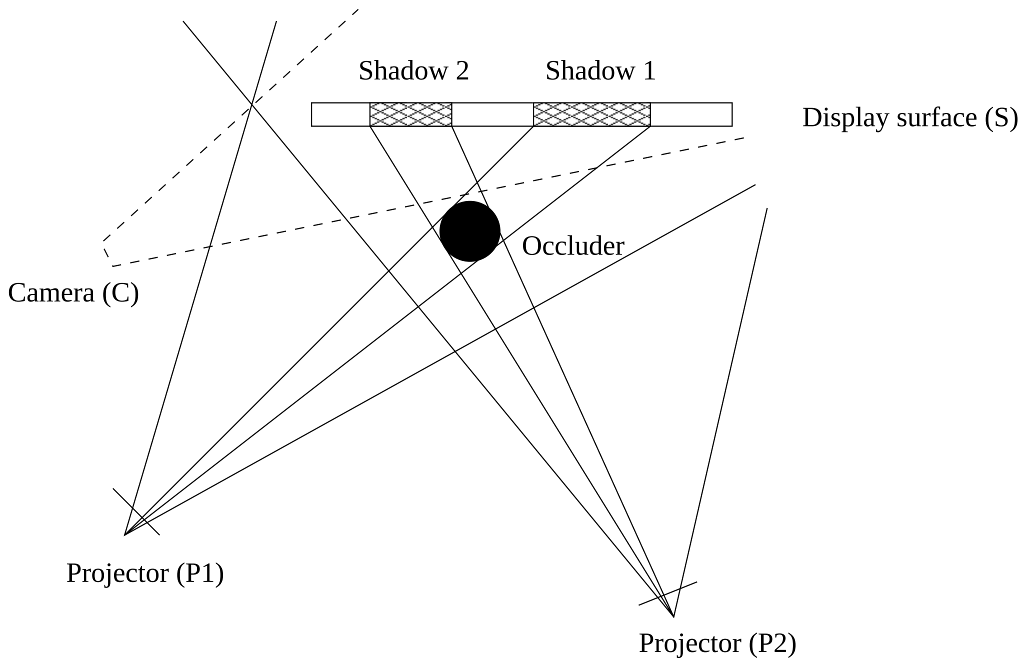


Figure 1: An overhead view of the multi-projector display system. Several projectors (P_1, P_2) are placed such that their projection areas converge onto the display surface (S). A camera (C) is positioned so that S is clearly visible in its field of view. Once the homographies relating P_i, C, S are automatically determined, the projectors combine their pre-warped outputs to create a single high-quality image on S. Each occluder in the environment may cast up to one shadow per projector on the display surface. The system locates these penumbral occlusions in the camera image and dynamically adjusts the projected images to make the shadows disappear. Simultaneously, the undesirable light projected on the occluders is suppressed. This can be achieved even though none of the occluders are directly visible in the camera.

A classical approach to designing such a system would be to localize the projectors, camera, display, and occluders in 3-D. This would make it possible to predict the shadowed region and the projector pixels which are being occluded. Given further information about the reflectance properties of the surface, the optimal compensation could be determined and applied. This approach has a strong disadvantage that any errors in localization and sensing have a direct impact on the quality of the display. Particularly in the case of tracking occluders, significant errors are likely. This paper explores the alternative approach of defining the desired screen output and using geometric and photometric compensation, inside a tight visual feedback loop, to correct for errors and disturbances.

The two classes of problems to be addressed in our system are: (i) geometric, and (ii) photometric. The geometric problems relate to computation of the spatial correspondences between pixels in the projectors and the projected display on the screen. The projectors should be accurately and automatically calibrated to the screen, to the camera and to each other. The calibration should enable the images in each projector to be pre-warped so as to create a desired projected display that is aligned with the screen. The photometric issues are the accurate and fast computation of the desired pixel intensities in each projector so as to eliminate shadows and suppress occluder illumination. This involves occlusion detection based on camera input and correctly adapting the projector output to achieve the necessary goals. These two classes of problems are addressed in sections 3 and 4 respectively.

3 Automatic Multi-Projector Alignment

As shown in Figure 1, several projectors are placed so that their projection areas all converge onto a display surface S. The goal is to combine the light from the projectors to create a single, sharp image on S. Clearly, one cannot simply project the same raw image simultaneously through the different projectors; not only does a given point on S correspond to very different pixel locations in each projector, but the image produced on S from any single projector will be distorted (since the projectors are off-center to S).

We assume that: the positions, orientations and optical parameters of the camera and projectors are unknown; camera and projector optics can be modeled by perspective transforms; the projection screen is flat. Therefore, the various transforms between camera, screen and projectors can all be modeled as 2-D planar homographies:

$$\begin{pmatrix} xw \\ yw \\ w \end{pmatrix} = \begin{pmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \\ p_7 & p_8 & p_9 \end{pmatrix} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}, \qquad (1)$$

where (x, y) and (X, Y) are corresponding points in two frames of reference, and $\vec{p} = (p_1 \ldots\ p_9)^T$, constrained by $|\vec{p}| = 1$, are the parameters specifying the homography. These parameters can be obtained from as few as four point correspondences, using the camera-projector calibration technique described in [12].

The homography for each camera-projector pair T_{c,P_i} can be determined by projecting a rectangle from the given projector into the environment. The coordinates of the rectangle's corners in projector coordinates (x_i, y_i) are known *a priori*, and the coordinates of the corners in the camera frame (X_i, Y_i) are located using standard image processing techniques.[3]

The display area is either automatically determined using the camera, or interactively specified by the user. The former case requires the display surface to be a white, rectangular projection screen against a contrasting background. Such a screen shows up clearly in the camera image as a bright quadrilateral, and can be unambiguously identified by the automatic calibration process. The corners of this quadrilateral fully specify the homography between camera and screen $T_{c,s}$, and the homography between each projector and the screen $T_{P_i,s}$ can be recovered using the equation:

$$T_{P_i,s} = T_{c,P_i}^{-1} T_{c,s}, \tag{2}$$

where the homographies on the right hand side of the equation are all known.

Alternatively, the user may interactively specify the display area by manipulating the outline of a projected quadrilateral in any projector until it appears as a rectangle of the desired size on the display surface. This directly specifies the homography between the selected projector and the screen $T_{P_i,s}$; the outline of the selected rectangle is then be detected in the camera image to determine the camera to screen homography $T_{c,s}$ and to automatically align all of the other projectors.

The projector-screen homographies $T_{P_i,s}$ model the geometric distortion (keystone warping) that is induced when an image is projected from an off-center projector P_i. This distortion can be corrected by projecting a *pre-warped* image, generated by applying the inverse transform $T_{P_i,s}^{-1}$ to the original image.[4] Since $T_{P_i,s}^{-1} T_{P_i,s} = I$, one can see that the pre-warping also aligns the images from different projectors so that all are precisely projected onto S. Applying the homographies derived from camera images, a multi-projector array can thus be efficiently configured to eliminate keystoning distortions and double images on the display surface.

[3] Hough-transform line-fitting [1] locates the edges of the quadrilateral, and its corner coordinates are given by intersecting these lines.

[4] In our current system, this pre-warp is efficiently implemented using the texture-mapping operations available in standard 3-D graphics hardware.

4 Photometric Framework for Multi-Projector Display

After the projectors have been geometrically aligned, we can easily determine which source pixels from the projectors contribute to the intensity of an arbitrary screen pixel. In the following analysis, we assume that the contributions are at some level additive. Given N projectors, the observed intensity Z_t of a particular screen pixel at time t may be expressed by

$$Z_t = C\Big(k_{1t}S_1(I_{1t}) + k_{2t}S_2(I_{2t}) + \cdots + k_{Nt}S_N(I_{Nt}) + A\Big), \tag{3}$$

where I_{jt} is the corresponding source pixel intensity set in projector j at time t, $S_j(\cdot)$ is the projector to screen intensity transfer function, A is the ambient light contribution which is assumed to be time invariant, $C(\cdot)$ is the screen to camera intensity transfer function and k_{jt} is the *visibility ratio* of the source pixel in projector j at time t. Note that all the variables and functions also depend on the spatial position of the screen pixel, but this is omitted from the notation since we will consider each pixel in isolation. See Figure 2.

When occluders obstruct the paths of the light rays from some of the projectors to the screen, Z_t diminishes and shadows occur. This situation is quantitatively modeled via the visibility ratios, which represent the proportion of light rays from corresponding source pixels in the projectors that remain unobstructed. If the projectors were modeled as point-light sources, occluders would block either none or all of the light falling on a given pixel from any particular projector; therefore, k_{jt} would be a binary variable. However, this assumption is not valid in real-world conditions. Our system must cope with partial occluders (created by objects near the projector) that cast fuzzy-edged shadows on the screen. In these cases k_{jt} denotes the degree of occlusion of projector j for the given pixel.

4.1 Occlusion detection

Rather than locating occluders by tracking objects in the environment, the system focuses exclusively on detecting deviation of the observed intensities on the screen from the desired intensities when occluders are not present. The major cause of deviation is occlusion, although deviation can also occur because of changes in ambient lighting, projector failure, etc. Our system can handle all of these problems (as discussed in the next section). No assumptions are made about the locations, sizes or shapes of occluders.

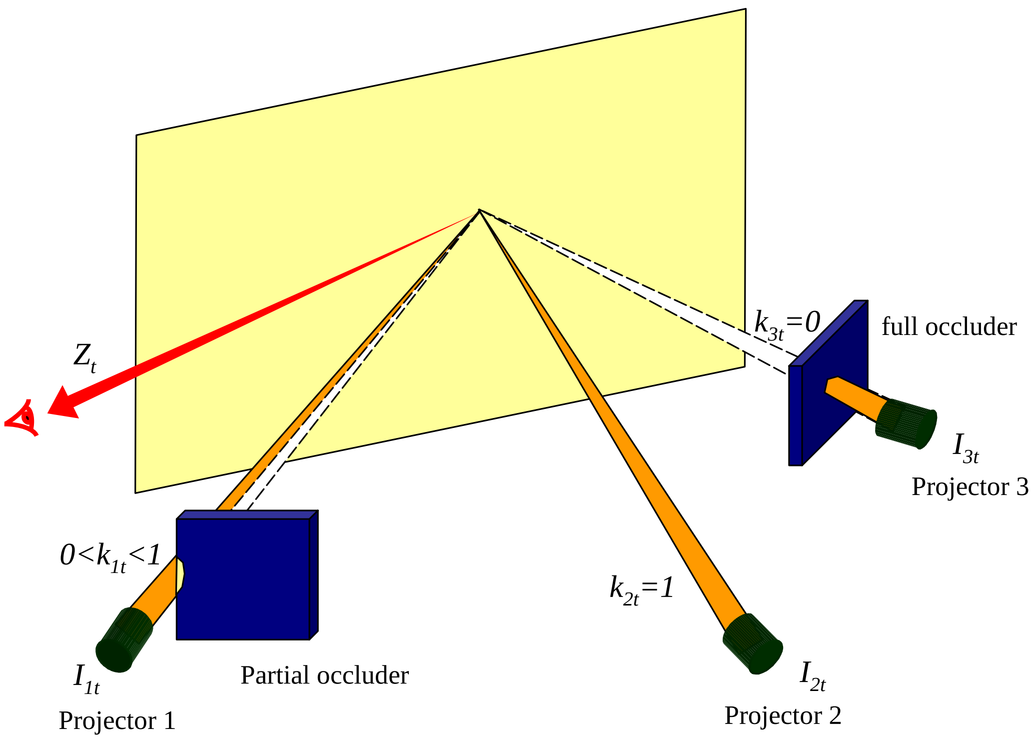


Figure 2: Photometric framework. This diagram illustrates equation (3), in which the observed display intensity Z_t is related to the combination of projector source pixels I_{jt} and the corresponding visibility ratios k_{jt}. The visibility ratios vary accordingly with non-occlusion, partial and full occlusion.

Mathematically, the desired intensity of a particular screen pixel may be represented by Z_0. This may be obtained in the initialization phase when the system projects each presentation slide and captures several camera images of the projected display while occluders are absent. As an occluder is introduced in front of projector k to create penumbral shadows, the visibility ratio k_{jt} decreases, such that $k_{jt} < 1$. Hence $Z_t < Z_0$. These deviations in the screen can be detected via a pixel-wise image difference between current and reference camera images to locate shadow artifacts.

4.2 Iterative Photometric Compensation

Our system handles occluders by

1. compensating for shadows on the screen by boosting the intensities of unoccluded source pixels; and

2. removing projector light falling on the occluder by blanking the intensities of occluded source pixels.

The degrees-of-freedom available to us are the source pixels I_{jt} which may be changed. Hence for a shadowed screen pixel where $Z_t < Z_0$, we ideally want to compensate for the shadow (i.e. setting $Z_{t+1} = Z_0$) by (i) increasing $I_{j(t+1)}$ to be larger than I_{jt} if $k_{jt} = 1$, and (ii) reducing $I_{j(t+1)}$ to zero if $k_{jt} < 1$.

However, it is very difficult to accurately model $C(\cdot)$ and $S_j(\cdot)$. Even if we know the exact values for the ambient lighting and visibility ratios, it is almost impossible to update the source pixels such that in one time step the shadows are eliminated. Fortunately, we expect $C(\cdot)$ and $S_j(\cdot)$ to be positive monotonic, and an iterative negative feedback loop can be used to compute

$$\arg \min_{I_{1t},\ldots,I_{Nt}} (Z_t - Z_0). \tag{4}$$

The advantages of such a system are:

- it does not require explicit modeling of $C(\cdot)$ and $S_j(\cdot)$,

- it does not require explicit measurement of the visibility ratios k_{jt},

- it is able to handle slowly varying ambient light.

As in [11], the change in the intensity of each source pixel in each projector is controlled by the alpha value associated with the pixel:

$$I_{jt} = \alpha_{jt} I_0, \tag{5}$$

where I_0 is the original value of the source pixel (i.e. pixel value in the presentation slide) and is the same across all projectors, while $\alpha_{jt}, 0 < \alpha_{jt} < 1$ is the time-varying, projector-dependent alpha value. The alpha values for the source pixels in one projector is collectively termed the alpha mask for the projector.

The earlier system described in [11], as well as the related approach of [5], can compensate for shadows but is incapable of suppressing projected light falling on the occluder. Shadow compensation does not require the ability to distinguish between the contributions of individual projectors. Instead, all projectors can boost their pixel intensities for each occluded region. This simple approach has two undesirable consequences: (1) bright “haloes” may appear around eliminated shadows, particularly when occluders are in motion; and (2) the amount of distracting light projected on users is *increased* rather than reduced by the system.

In contrast to shadow elimination, occluder light suppression is much more challenging, as it requires the ability to determine *which* projected rays are being blocked at any given point in time. This in turn leads to a more complex implementation in which each projector has a potentially different alpha mask. The approach adopted here is to design components which separately handle the problems of shadow elimination and occluder light suppression, and integrate them into a complete system.

4.3 Shadow Elimination

Eliminating shadows involve increasing values for corresponding source pixels. The shadow elimination (SE) component of the system is based on

$$(\Delta\alpha_{jt})_{\mathrm{SE}} = -\gamma(Z_t - Z_0), \tag{6}$$

where $\Delta\alpha_{jt} = \alpha_{j(t+1)} - \alpha_{jt}$ is change of α_{jt} in the next time-frame, and γ is a proportional constant. This component is a simple, linear feedback system.

4.4 Occluder Light Suppression

Suppressing projector light falling on occluders involve diminishing the source pixels corresponding to the occluded light rays. We determine whether a source pixel is occluded by determining if any changes in the source pixel result in changes in the screen pixel. However, since there are N possible changes of source pixel intensities from N projectors but only one observable screen intensity, we need to probe by varying the source pixels in different projectors separately. This cyclical probing results in a serial variation of the projector intensities.

The light suppression (LS) component of the system is based on

$$(\Delta\alpha_{jt})_{\text{LS}} = -\beta \frac{\Delta\alpha^2_{j(t-N)}}{\Delta Z_t^2 + \epsilon}, \tag{7}$$

where $\Delta Z_t = Z_t - Z_{t-N}$ is the change in the screen pixel intensity caused by the change of alpha value $\Delta\alpha_{j(t-N)}$ in the previous time frame when projector j is active, and β is a small proportional constant and ϵ is a small positive constant to prevent a null denominator.

The rationale for (7) is that if the change in α_{jt} results in a corresponding-sized change in Z_t, the subsequent change in α_{jt} will be relatively minor (based on a small β). However if a change in α_{jt} does not result in a change in Z_t, this implies that the source pixel is occluded. The denominator of (7) approaches zero and α_{jt} is strongly reduced in the next time frame. Hence occluded source pixels are forced to black.

Note that the probe technique must be employed during shadow elimination as well. In particular, the system must be able to discover when a pixel which was turned off due to the presence of an occluder is available again, due to the occluder's dissapearance. This constraint is smoothly incorporated into our algorithm.

4.5 Integrated System for Shadow Elimination and Occluder Light Suppression

The integrated iterative feedback system combines (6) and (7) to get

$$\Delta\alpha_{jt} = (\Delta\alpha_{jt})_{\text{SE}} + (\Delta\alpha_{jt})_{\text{LS}}\,. \tag{8}$$

The alpha values are updated within limits such that

$$\alpha_{jt} = \begin{cases} 1, & \text{if } \alpha_{jt} + \Delta\alpha_{jt} > 1, \\ 0, & \text{if } \alpha_{jt} + \Delta\alpha_{jt} < 0, \\ \alpha_{jt} + \Delta\alpha_{jt}, & \text{otherwise.} \end{cases} \tag{9}$$

The following synthetic example illustrates the system. For a particular screen pixel at a typical steady state when shadows have been eliminated, suppose the corresponding source pixel at projector 1 has $\alpha_{1t} = 0$ and the source pixel at projector 2 has $\alpha_{2t} = 1$. At this state, $Z_t = Z_0$ and $\Delta\alpha_{jt} = 0$ via (8). If source pixel 2 is suddenly occluded and source pixel 1 is unoccluded, then $Z_t < Z_0$ because source pixel 1 is still black. However, $\Delta\alpha_{1t}$ becomes dominated by $(\Delta\alpha_{1t})_{\mathrm{SE}}$ which forces source pixel 1 to be bright. On the other hand, $\Delta\alpha_{2t}$ becomes dominated by $(\Delta\alpha_{2t})_{\mathrm{LS}}$ since the screen pixel does not change when α_{jt} is changed. This forces source pixel 2 to be dark. Note that even when source pixel 2 becomes unoccluded, nothing changes if source pixel 1 remains unoccluded since the shadows have already been satisfactorily eliminated. See Figure 3. This particularly illustrates the *hysteresis effect* in which source pixels are not boosted or blanked until new shadows are created – the system does not automatically return to an original state, nor change as a result of deocclusion.

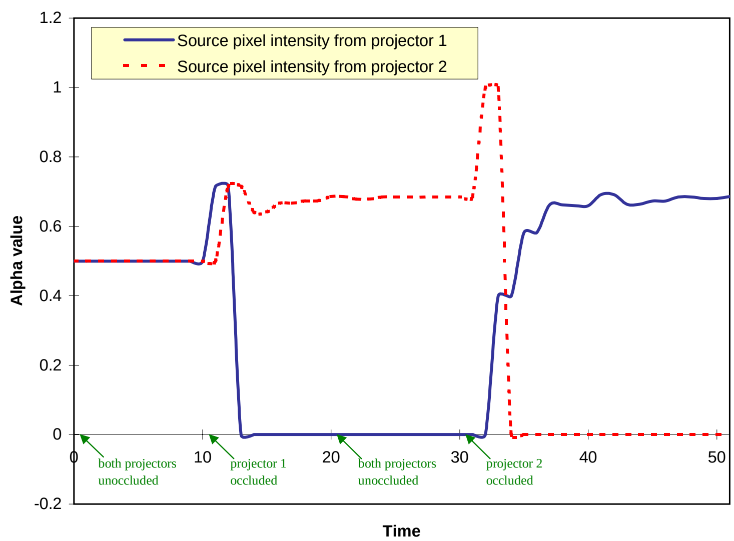


Figure 3: Synthetic example of transitions in projector source pixel intensities. This graph shows the intensity transition of two corresponding projector source pixels over time, subject to four events of occlusions and deocclusions. Note the hysteresis effect in which the source pixels are not boosted or blanked until new occlusion events occur – the system does not change as a result of the deocclusion event at time 20.

4.6 System Details

Most of the computation is carried out in the camera frame of reference. The differences between the observed and desired intensities $Z_t - Z_0$ for all screen pixels are obtained simply by subtracting the current camera image from the reference camera image. Similarly, the alpha values of the source pixels within the same projector is collectively the alpha mask for the projector. The alpha masks are also computed in the camera frame of reference and must be transformed into the screen frame of reference before they can be applied; this is done using the camera-screen homography $T_{c,s}$ discussed in section 3.

Applying the alpha masks to the current slide is straightforward. For each projector, the corresponding alpha mask is warped and replaces the alpha channel of the slide image. The slide is then pre-warped for the projector (using its particular screen-to-projector homography) and displayed. This is done separately for every projector.

Figure 4 illustrates the algorithm. During its initialization phase (when the scene is occluder-free) the system projects each presentation slide and captures several camera images of the projected display. These images are pixel-wise averaged to create a reference image for that slide, and this image represents the desired state of the display (Figure 4, *reference image*). The goal of occlusion detection is to identify regions in the current image that deviate from this ideal state. During operation, the system camera continuously acquires images of the projected display which may contain uncorrected shadows. The comparison between the observed images and the reference image facilitates the computation of the alpha masks for individual projectors through (8). These are merged with the presentation slide in the screen frame of reference, followed by further warping into the projector frame of reference. These projected images from all projectors optically blend to form the actual screen display.

Since we do not have good photometric models of the environment, projectors or camera, we cannot predict precisely how much light is needed to remove a shadow. However, the iterative feedback loop used to update the alpha mask allows us to avoid this problem: the system will continue adding light to shadowed regions until the region appears as it did in the reference image. Similarly, the system will blank projector source pixels which are occluded and do not affect the observed images. This approach has additional benefits. For instance, the system does not require an accurate photometric model of the shadow formation process to correct for occlusions with non-binary visibility ratios, e.g. the diffuse shadow created by a hand moving near the projector. The drawback to such an it-

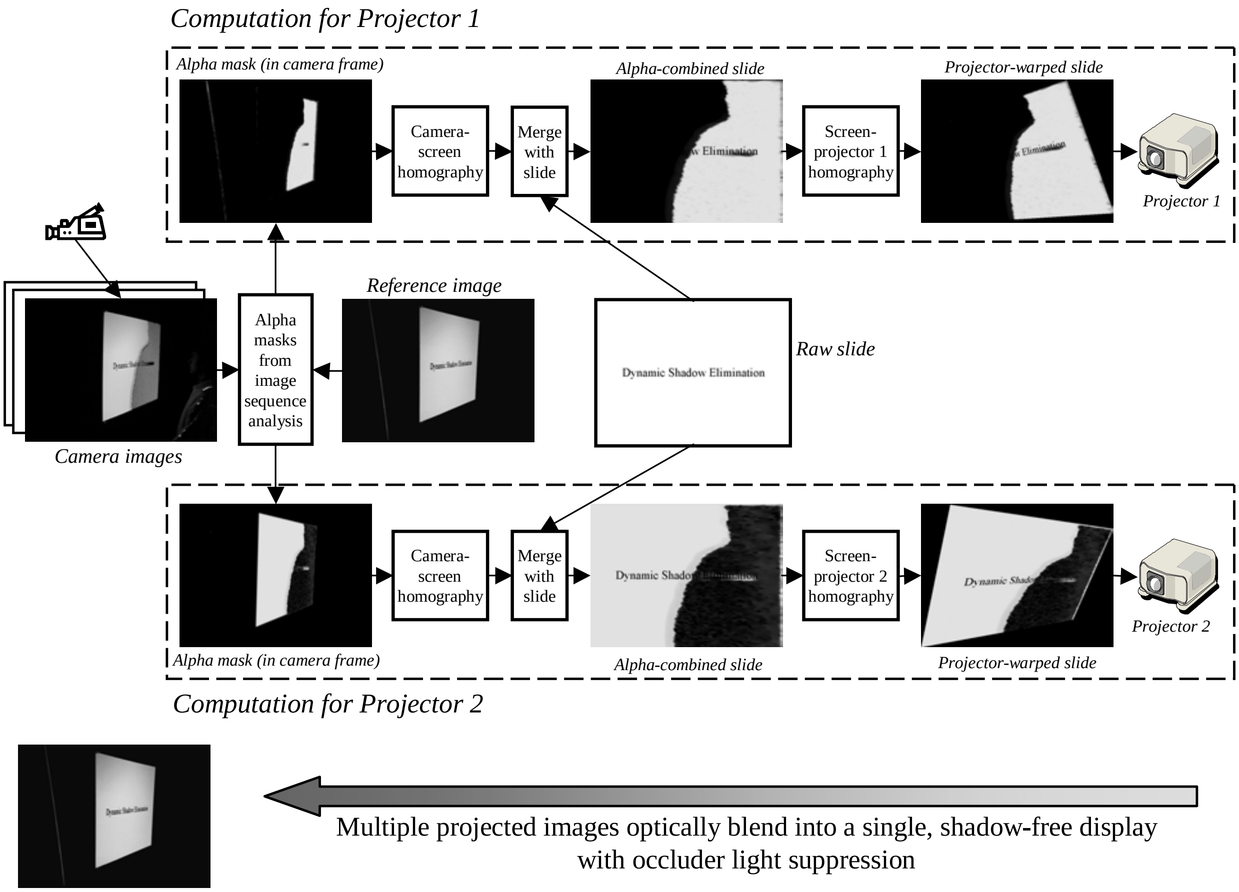


Figure 4: System summary diagram. This diagram summarizes the shadow elimination and occluder light suppression system. See text for details.

erative technique is that the alpha mask can require several iterations to converge; in practice, shadows are eliminated in approximately 5–7 iterations.

To reduce the effects of camera noise and minor calibration errors, we apply a 5×5 spatial median filter to the difference image. A negative value in a difference image pixel means that the corresponding patch on the screen was under-illuminated in the current image.

5 Results

The experiments described in this section are based on the following implementation. Images are acquired using an NTSC camera (640×480) attached to a PCI digitizer; the output subsystem consists of two Compaq MP-2800 XGA-resolution (1024×768) DLP microportable projectors driven by a dual-headed graphics card; the software runs on a standard workstation, and image warping exploits OpenGL texture-mapping primitives. The projectors are positioned on either side of a whiteboard and the camera is mounted at an extreme angle to ensure occlusion-free coverage of the display surface The location, orientation and optical parameters of the camera and projectors are unknown.

Note that the cost of shadow elimination is the use of redundant projectors. This means that at any point in time there are pixels on one or more projectors that are not being utilized because they fall outside the display surface or are occluded. We feel this is a small price to pay, particularly in comparison to the large costs, in either expense and required space, for other display technologies such as rear projection or plasma. Fortunately, portable projectors are becoming increasingly affordable as their image quality improves and their weight decreases.

Some images of the system in action are shown in Figure 5. The images demonstrate the difference between shadow elimination alone and in combination with occluder light suppression.

5.1 Steady State Performance

We measured the steady state performance of our system for the shadow elimination task and compared it to passive one- and two-projector configurations. We projected a series of slides using the three systems, first without occluders and then with static occluders in the scene. Images of the display region were captured using a tripod-mounted camera. We computed a sum-squared-difference (SSD) of the grey-level intensities (over the pixels corresponding to the display region in

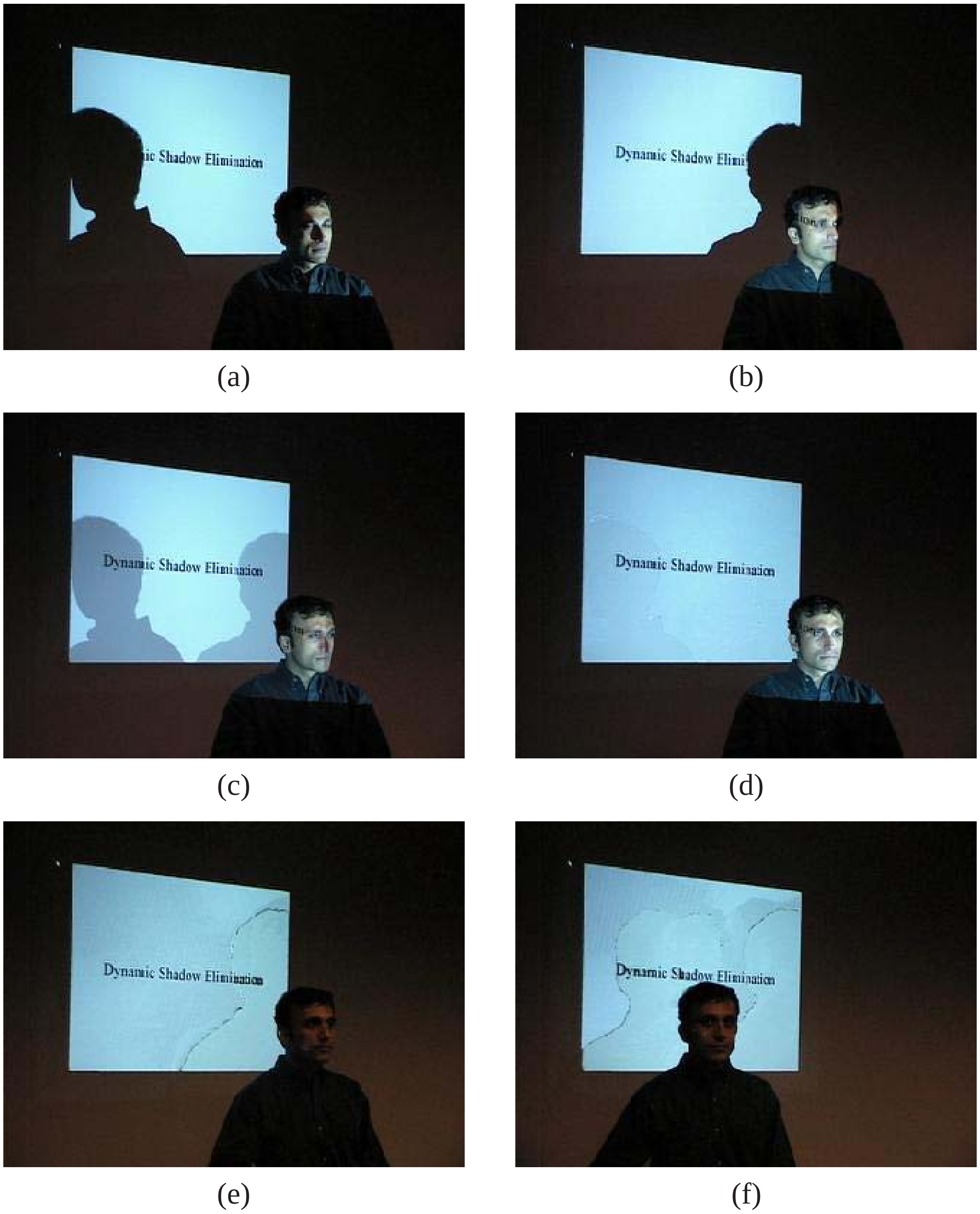


Figure 5: Comparison between different projection systems. These images were taken from an audience member's viewpoint: (a) Right projector only; (b) Left projector only; (c) Two aligned projectors, passive; (d) Two aligned projectors with shadow elimination only; (e,f) Two aligned projectors with shadow elimination and occluder light suppression. Note that the user's face is harshly illuminated in (a)-(d), and completely dark in (e) and (f).

the camera image) for each slide pair (with and without occluders). Table 1 summarizes these results. As expected, the (umbra) shadows in the single-projector display are a major source of error. We see a significant improvement with two passive projectors since the majority of occlusions become penumbral. The best results are achieved when these penumbral occlusions are actively eliminated using our system.

Table 1: Average SSD error between camera images of occluded and unoccluded display. Fusing a passive second projector reduces the errors. Our system gives an additional order of magnitude of improvement.

	SSD error per pixel over display (normalized intensities)
One projector (passive)	148.0×10^{-3}
Two projectors (passive)	12.8×10^{-3}
Shadow elim. + occluder light supp.	2.5×10^{-3}

We also measured the steady state performance on the occluder light suppression task. We captured multiple images of a person's face from an audience viewpoint, while it occluded both projectors (see Figure 5 for the basic set-up). We computed the average pixel intensity over the illuminated region. Table 2 shows the results. The passive single-projector system brightly illuminates the user's face from one direction. In the passive two-projector system, each projector operates at a reduced brightness. The average intensity is lower because the user's face is illuminated from two directions and some of the light from the second projector illuminates areas of the face that are not visible to the camera. The shadow-elimination system from [11] also uses two projectors, but each projector increases illumination for pixels in the shadowed region, creating a very bright patch of light on the occluder. In contrast, the system described in this paper dramatically suppresses the occluder illumination.

5.2 System Response

Since our algorithm uses an iterative technique, we wanted to confirm that the method converges quickly to a stable value without excessive oscillation or overshoot. We evaluate the system response to occluders by measuring the time-

Table 2: Normalized intensity of light on the occluder, averaged over the visible pixels. Shadow elimination alone [11] is worse than passive projection because the projectors emit additional light.

	Average normalized intensity measured on occluder's face
One projector (passive)	0.568
Two projectors (passive)	0.455
Shadow elimination only	0.561
Shadow elim. + occluder light supp.	0.113

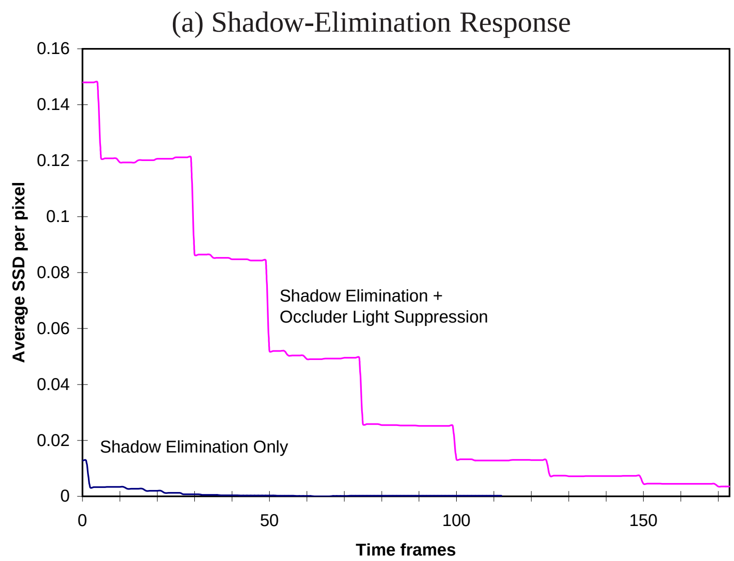
variation of (i) the average SSD per pixel between the observed and reference screen images, and (ii) the average pixel intensity on the occluder's face. The response is measured by suddenly introducing an occluder into the scene and recording the observed state of the display from that instant.

Figures 6 (a) and (b) present system traces for both experiments. In each graph, one trace shows the system from [11] while the other plots the response of the combined shadow elimination and occluder light suppression system. The combined system takes much longer to converge for two reasons. First, the system is more complex and runs more slowly. Each iteration of the algorithm takes approximately 3–4 times longer than a comparable iteration for the former system; this is easily seen from the length of the steps in the graphs. Second, the combined system requires cyclical projector probing to determine the occluded projector. Consequently, an N projector system requires N times longer to converge.

The initial SSD error for the two systems differ because the former starts with both projectors on (equivalent to a 2-projector passive system) while the combined system starts with one projector on, and the other projector off. Thus, when the occluder is initially introduced, the shadows for the combined system are as dark as umbral occlusions. The SSD errors for both systems drop quickly with each iteration and the final errors are comparable.

Figure 6(b) demonstrates the clear benefits of the combined system. The introduction of the occluder causes an *increase* in brightness on the user's face for the former system while the combined system quickly removes the undesirable light.

Another advantage of our current system is that it is significantly more re-



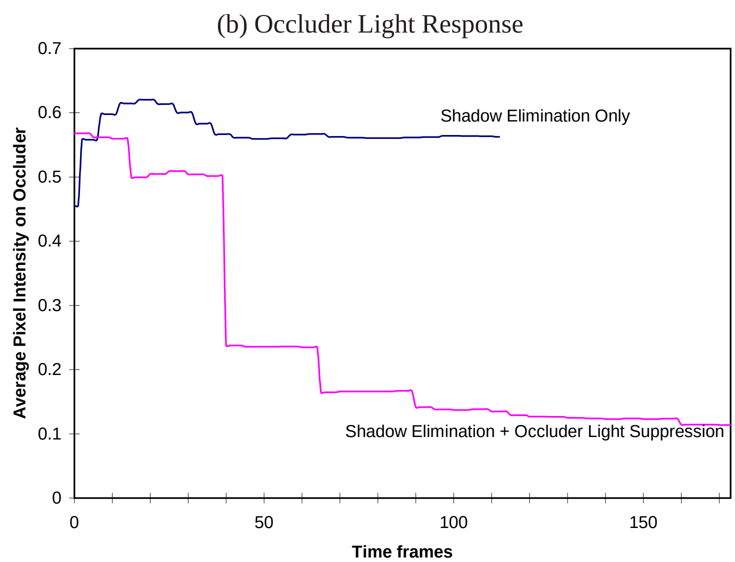


Figure 6: (a) System responses for shadow elimination, based on average SSD error per pixel over time. (b) System responses for occluder illumination, based on average pixel intensity of region containing occluder's face. Although the combined system is slower, there is substantial occluder light suppression and the steady-state shadow elimination error is comparable.

sistant to a moving occluder. In [11], the slight swaying motion of a standing occluder creates very noticeable "halo" effects due to the occluder's leading edge creating a shadow while the trailing edge creating an inverse-shadow (when occluded source pixels in a projector are unoccluded). In our system, the occluded source pixels are blanked; hence no inverse-shadows are created by deocclusion. Additionally, the hysteresis effect ensures that source pixel states are not changed until new shadows are created. This means that source pixels cyclically occluded by the span of the occluder swaying motion eventually become blanked and no new shadows (or artifacts) are caused.

A benefit of using a feedback system (as opposed to a photometric model approach) is that the system is surprisingly robust. For instance, if one of the projectors in the multi-projector array were to fail, the remaining projectors would automatically brighten their images to compensate. Furthermore, the overall brightness of the entire multi-projector array can be changed simply by adjusting the camera aperture.

6 Related Work

Research in the area of camera-assisted multi-projector displays is becoming more popular, particularly in the context of seamless video walls [14, 7, 4, 6, 13, 9]. Two previous papers [11, 5] presented solutions to the shadow elimination problem for forward-projection systems. However, we are not aware of any work on the problem of occluder light suppression. This problem is technically much more challenging because it requires the ability to determine *which* rays of projected light are being occluded. We believe our results in occluder light suppression are unique. A simple camera feedback system, related to the one presented here, was used by [13] to adjust projector illumination for uniform blending in the overlap region of a video wall. The Shader Lamps system [10] uses multiple projectors and a known 3-D model to synthesize interesting visual effects on 3-D objects. The self-calibration techniques used in this paper were adopted from [12], where they were applied to the task of automatic keystone correction for single projector systems.

7 Conclusions and Future Work

This paper describes a practical system for creating a stable, occlusion-free display using multiple overlapping projectors and a camera. The system automatically aligns multiple projected images to the display surface, compensates for shadows cast on the display due to the obstruction of one or more projectors, and prevents the obstructed projectors from casting light onto any of the occluding objects. The system simulates the user experience of a rear-projection display using front-projection technology. It supports the creation of immersive, interactive displays in any general indoor location, as an alternative to monitors, plasma displays and rear-projection systems.

In the future, we plan to extend the system in several ways. In addition to increasing the frame rate at which the system operates, we will incorporate multiple cameras into the visual feedback loop. This will increase the system's reliability in cases where a single camera's view of the screen is also partially occluded. We aim to further develop the connections between our system and the visual servoing literature [2]. Our current feedback loop can be viewed as a proportional controller and we hope to explore extensions to PID control. We are also developing user-interface techniques for controlling and adjusting virtual displays using hand gestures. In particular, we are exploring shadow detection as a means to support touch-based interaction with the virtual display.

References

[1] D. Ballard and C. Brown. *Computer Vision*. Prentice-Hall, 1982.

[2] P. Corke. Visual control of robot manipulators — a review. In K. Hashimoto, editor, *Visual Servoing*. World Scientific, 1994.

[3] O. Hall-Holt and S. Rusinkiewicz. Stripe boundary codes for real-time structured light range scanning of moving objects. In *Proceedings of International Conference on Computer Vision*, 2001.

[4] M. Hereld, I. Judson, and R. Stevens. Introduction to building projection-based tiled displays. *Computer Graphics and Applications*, 20(4), 2000.

[5] C. Jaynes, S. Webb, R. M. Steele, M. Brown, and W. B. Seales. Dynamic shadow removal from front projection displays. In *Proc. IEEE Visualization*, 2001.

[6] K. Li, H. Chen, Y. Chen, D. Clark, P. Cook, S. Daminakis, G. Essl, A. Finkelstein, T. Funkhouser, A. Klein, Z. Liu, E. Praun, R. Samanta, B. Shedd, J. Singh, G. Tzanetakis, and J. Zheng. Building and using a scalable display wall system. *Computer Graphics and Applications*, 20(4), 2000.

[7] A. Majumder and G. Welch. Computer graphics optique: Optical superposition of projected computer graphics. In *Proceedings of Eurographics Workshop on Rendering*, 2001.

[8] C. Pinhanez. The Everywhere display. In *Proceedings of Ubiquitous Computing*, 2001.

[9] R. Raskar, M. Brown, R. Yang, W. Chen, G. Welch, H. Towles, B. Seales, and H. Fuchs. Multi-projector displays using camera-based registration. In *Proceedings of IEEE Visualization*, 1999.

[10] R. Raskar, G. Welch, and K.-L. Low. Shader Lamps: Animating real objects with image-based illumination. In *Proceedings of Eurographics Workshop on Rendering*, 2001.

[11] R. Sukthankar, T.-J. Cham, and G. Sukthankar. Dynamic shadow elimination for multi-projector displays. In *Proceedings of Computer Vision and Pattern Recognition*, 2001.

[12] R. Sukthankar, R. Stockton, and M. Mullin. Smarter presentations: Exploiting homography in camera-projector systems. In *Proceedings of International Conference on Computer Vision*, 2001.

[13] R. Surati. *A Scalable Self-Calibrating Technology for Seamless Large-Scale Displays*. PhD thesis, Department of Electrical Engineering and Computer Science, Massachussetts Institute of Technology, 1999.

[14] R. Yang, D. Gotz, J. Hensley, H. Towles, and M. Brown. PixelFlex: A reconfigurable multi-projector display system. In *Proceedings of IEEE Visualization*, 2001.

CRL 2002/03

March 2002

Shadow Elimination and Occluder Light Suppression for Multi-Projector Displays

Tat-Jen Cham Rahul Sukthankar James M. Rehg
Gita Sukthankar