Realizing a Stereo Tiled Display Using Commodity Components

D.M. Germans\textsuperscript{1}, H.J.W. Spoelder\textsuperscript{1},
T. v.d. Schaaf\textsuperscript{2} and H.E. Bal\textsuperscript{1,2}

\textsuperscript{1}Division of Physics and Astronomy
\textsuperscript{2}Division of Mathematics and Computer Science
Faculty of Sciences
Vrije Universiteit
De Boelelaan 1081
1081 HV Amsterdam
The Netherlands

May 2, 2003

Keywords: Applied Imaging, Computer Graphics, Depth Vision, Multi-projector Displays

Abstract

This paper shows how, with low-cost commodity components, one can realize a high-resolution scalable stereo display with a good stereo image. We show that geometric, photometric and viewpoint calibration are very important, and we describe how these calibrations can be performed accurately with help of computer vision techniques.

1 Introduction

In the field of visualization, PC cluster-driven tiled displays are becoming more and more popular due to their excellent price-performance ratio [4]. Tiled displays provide a scalable and high resolution display by tiling together projector outputs on a large screen. Each projector is driven by a graphics-oriented off-the-shelf PC, interconnected via a fast network (Fast Ethernet or Myrinet[1]).

For a number of applications, (semi-)immersiveness - providing the user with depth vision - greatly enhances the usability of these displays. In order to realize a stereo tiled display, two main topics need to be addressed. First, the process of generating and coding different images for the audience’s left and right eyes. For this, and because low-cost equipment is used, geometric and photometric calibration are very important. Second, the implementation of proper depth vision. Here, the frustum, inter-ocular distance and the viewer’s fusion point are important. Opposed to existing VR systems (like the CAVE [2]), a stereo tiled display has multiple viewers that are not tracked. Because of this, the image can not be calculated correctly for every viewer and average solutions have to be considered.

At the ICWall site [6] of the Faculty of Sciences at the Vrije Universiteit in Amsterdam, we converted the 4x2 tiled display array as a high resolution stereo display.

The contributions of this paper are:

\begin{itemize}
  \item setup considerations in order to realize a stereo tiled display,
  \item a sub-millimeter-accurate and deterministic method to geometrically calibrate the display for both eyes,
  \item a phenomenological approach to address photometric calibration issues,
\end{itemize}
• a description on viewpoint calibration for convincing stereo images.

The rest of the paper is structured as follows. Section 2 introduces issues in calibration and architecture for tiled displays as well as stereo-graphic projection. Section 3 explains our setup. Section 4 shows how geometric, photometric and viewpoint calibration is done. Section 5 gives a summary and some conclusions.

2 Related Work

2.1 Stereo Displays

A lot has happened since Sir Charles Wheatstone in 1833 demonstrated that depth vision is in fact the difference between what the left and the right eye see. The stereo separation techniques are refined and applied to interactive computer displays. Nowadays, there are several ways to produce a stereo image, and a lot is known on how humans perceive depth [13, 11].

Next to the more passive viewing experiences (images, animations, amusement parks), interactive stereo graphics have been refined in the realm of virtual reality. Displays support stereo in a more classical sense [2, 9], and recently, commodity component setups support stereo at lower costs [7, 10].

2.2 Tiled Displays

Even though the mentioned stereo displays produce very acceptable stereo images, cost-effective solutions are usually tailored for a smaller audience. In order to reach a large audience scalably, research has been done to combine several projector outputs and effectively tile up a larger display. However, to our knowledge, all tiled display initiatives are mono-sopic.

Important issues for tiled display realization are the geometric and photometric calibration of each projector output in order to produce a large and seamless display with equal luminance and chrominance over the entire surface. Recent developments in geometric calibration have shown that a single-pass deterministic homographic calibration method is most accurate [12]. This method puts very little constraints to the physical positioning and orientation of each projector. More specific, as long as each projector is aimed at the same surface and overlaps with the neighboring projectors, a seamless display can be constructed.

Roughly, the method works as follows. After projecting a known stimulus on each tile, a computer connected to a CCD camera determines the orientation of each projector by fitting a homographic transformation to stimulus features that are recognized from the image captured by the camera. From this transformation, the configuration of the entire screen can be inferred. Currently, this method results in sub-millimeter accurate geometric calibration on the screen, a very acceptable accuracy.

Notable in the field of photometric calibration is work by [8]. Here they give a thorough analysis of various issues in photometric calibration and give a phenomenological method to equalize luminance (the most important component) of the picture, using what they call LAMs (Luminance Attenuation Maps). This mapping is based on a per-pixel analysis and does not attempt to functionally model the phenomena that cause the luminance variations.

3 Setup

This section will explain how our setup is constructed, and how separation is achieved between left- and right-eye images. For this separation, there are commonly two techniques:

Active Stereo Here, a single array of projectors would project images for left and right eye interleaved. Shutter glasses that alternatingly block the left and right eye are synchronized via infrared to the projectors in such a way that when the projector projects an image for the left eye, the right eye is blocked and vice versa. For a reasonable output, the alternation of the images needs to be 40Hz or faster for each eye, and, each projector needs to be perfectly
synchronized. This requires the projectors to be able to project at 80Hz or higher and the graphics cards to be fast and synchronized, effectively ruling out commodity LCD or DLP projectors and commodity cards. In order to reach larger audiences, this solution is not suitable. It does not scale well in display size, as well as cost (CRT projectors and shutter glasses are relatively expensive).

**Passive Stereo** Now, two arrays of projectors project the image for the left and right eye simultaneously to the screen. The light from each projector is polarized with a polarization filter in such a way that the image on the screen for the left eye consists only of linearly polarized light at -45 degrees, whereas the image for the right eye is polarized at 45 degrees. Inexpensive glasses using the same polarization filters allows each eye to see only light that is correctly polarized. The -45 degree light arrives at the left eye and is blocked at the right eye and vice versa. Now, demands on the projector refresh speeds and synchronization are not crucial for the eye separation, so commodity LCD and DLP projectors can be used. LCD projectors, however, tend to have a preferred polarization angle, which can interfere with the desired -45/45 separation. This technique allows for a large audience due to the low cost of the glasses.

Our implementation uses passive stereo on a back-projection screen. We selected a back-projection screen from StewartFilm for it’s polarization properties. Light passes through and is distributed with the polarization angle left unchanged. To avoid the preferred polarization of commodity LCD-projectors, we use DLP-projectors, two for each tile. Each set of projectors is connected to a PC with a dualhead (two outputs) GeForce 4 graphics card. The PCs as well as one extra host machine are interconnected via a Myrinet network, allowing high throughput and low network latency. Figure 1 shows our setup.

Figure 1: The frame, projectors and driving PCs at the ICWall site.

Figure 2: Each projector pair is stacked with the top projector slightly tilted.
We choose a projector configuration that allows easy and low-cost maintenance of the projectors. For each tile, both projectors are mounted so that they can shift and rotate with approximately 1 degree of freedom. The top projector is slightly tilted to approximately project onto the same area as the bottom one (Figure 2). Between the tiles there is a region where adjacent projector outputs overlap. This is done to allow the geometric calibration process to generate a smooth transition from one tile to the other.

4 Calibration

This section describes the geometric, photometric and viewpoint calibration applied to the display. Geometric calibration achieves a correspondence between coordinates of each individual projector and coordinates on the screen, photometric calibration achieves a correspondence between intended image luminance and chrominance (sent to the projector), and actual luminance and chrominance. Viewpoint calibration achieves a correspondence between the viewer in front of the screen and the utilized 3D frustum in the virtual world that is being displayed on the screen. We will look in detail at all three calibration processes and show how they become important in producing a convincing stereo image.

4.1 Geometric Calibration

To characterize the position and orientation of the individual projectors, we detect features of a synthetic stimulus generated by each projector. In our case we use a checkerboard pattern of known size (Figure 3). The features that are being looked for, are the crossings in the checkerboard. These crossings can be detected by performing a convolution of the captured image with a matched filter that has an extreme response where the filter matches the crossing [15]. Figure 4 shows a graphical representation of such a response. The white and black dots indicate full match and full anti-match (a checkerboard crossing where white and black are swapped). Additionally, [14] has
shown that one can obtain sub-pixel accurate localization of the checkerboard crossings by calculating the center-of-mass of the responses around each extremum.

Via this technique, we arrive at a list of landmarks for each tile/eye combination. We know that the list should represent the stimulus under a homographic transformation:

\[
\begin{align*}
x_{\text{out}} &= \frac{ax_{\text{in}} + by_{\text{in}} + c}{gx_{\text{in}} + hy_{\text{in}} + 1} \quad (1) \\
y_{\text{out}} &= \frac{dx_{\text{in}} + ey_{\text{in}} + f}{gx_{\text{in}} + hy_{\text{in}} + 1} \quad (2)
\end{align*}
\]

Here, \((x_{\text{in}}, y_{\text{in}})\) are the input coordinates (the known crossing locations in normalized projector space) and \((x_{\text{out}}, y_{\text{out}})\) are the output coordinates (the recognized crossings in CCD camera space). \(a,b,c,d,e,f,g\) and \(h\) are parameters of the transformation. We can now fit this transformation to the list of landmarks using a non-linear least squares fitting algorithm.

Where \(r\) is the distance of the point \((x_{\text{out}}, y_{\text{out}})\) with a parametric center \((x_c, y_c)\), \((x_{\text{ldm}}, y_{\text{ldm}})\) is the new output point and \(\alpha_2\) and \(\beta_2\) are distortion parameters. Implementing this as well, we arrive at sub-millimeter accuracy on the display surface.

With a transformation for each tile/eye combination, we can now setup quadrilaterals in screen space that represent the projectable region of each tile/eye combination. From this information, we configure the contents of the display. First the largest rectangle is found that fits on the screen for both eyes. When this rectangle is found, it is intersected with each quadrilateral. The resulting polygons represent the displayable area polygon of each tile/eye combination. Intersections between adjacent displayable area polygons give the overlap polygons.

![Figure 6: The gradual falloff of the overlap polygons for two adjacent tiles add together to form a smooth overlap.](image)

The transformation itself is now used to transform the pixels of the projector for geometric calibration. OpenGL, which is used to generate the stereo images, allows for a flexible homogeneous 3D transformation at no additional performance cost. However, the OpenGL transformation can only approximate the homographic transformation [12]. For mono displays, this approximation is sufficient and left-over distortions are not noticeable. For stereo displays however, the left-over distortions interfere with the eye separation, so the approximation is not enough and a per-pixel transformation process is required. This can be done most accurately by using the texture mapping logic of the 3D hardware in a second rendering pass. The texture mapping hardware is designed to take perspective plane transformations, and this is precisely what we require.
4.2 Photometric Calibration

Figure 7: Luminance response due to non-Lambertian characteristics of the back-projection screen. The hot spots indicate a direct line of sight with the heart of the projectors. Note that for the histogram, the overlap areas are not taken into account.

Photometric calibration comes down to equalizing the luminance and the chrominance response of each tile. For effective stereo output, it is important that images for both eyes are similar in luminance and chrominance. [8] shows that chrominance-differences are much less noticeable, so we concentrate on equalizing luminance only. Commodity DLP projectors add white to the image in order to boost the contrast, which complicates modeling of the chrominance considerably.

We measure luminance with the same CCD camera as used for geometric calibration. What needs to be said here is that the mapping between resulting camera output values and actual luminance is unknown. In order to calibrate this, one could follow [3]. However, we choose to model the entire process from projector input to camera output, including all calibration issues related to projectors or cameras.

A large contribution to luminance differences is the fact that the back-projection screen is not a perfect Lambertian source. Light that passes through and is diffused along a large angle is much less bright than light that passes through and is not bent at all (Figure 7). Because this effect is viewer-dependent, a perfect compensation does not exist and we have to choose a preferred location for the viewer. To allow scalability in the audience, one viewer for which the image is compensated is not enough, and so we have to blend luminance compensations of a whole range of viewers.

For each tile/eye-combination, the luminance difference can be modeled by a second-order polynomial in the distance to the center of a hot spot:

\[ r(x, y) = \sqrt{(\theta_1 - x)^2 + (\theta_2 - y)^2} \] (6)

\[ g(x, y) = \theta_3 + \theta_4 r(x, y) + \theta_5 r(x, y)^2 \] (7)

\( g(x, y) \) is the resulting luminance, \((x, y)\) is a point on the tile, \(\theta_1\) and \(\theta_2\) are the center of the hot-spot, \(\theta_3\), \(\theta_4\) and \(\theta_5\) are parameters of the quadratic model. To estimate these parameters, all projectors are set to output a specific level of gray on all pixels. A capture is made of the outputs of the projectors and the model is fitted to the projector response using a non-linear least squares fitting method. Figure 8 shows the luminance response for the display after this correction. Note that, contrary to Figure 7, the distribution is now very narrow.

Figure 8: Luminance response after applying the modeled correction. Note that the spread in the accompanying histogram is now minimal. Note again that for the histogram, the black overlap areas are not taken into account.

inter-ocular To utilize the luminance correction and fuse the overlap areas smoothly, an alpha mask is constructed for each projector. This mask is multiplied with the projector output to locally attenuate the signal (Figure 5). For this, we transform the displayable area and overlap polygons back to projector coordinates,
using the inverse of the geometric transformation. The displayable area polygon functions as a main mask. In order to have two adjacent tiles blend towards each other (Figure 6), the overlap polygons are used to produce gradual fall-offs towards the sides of the projection. Finally, the mask is attenuated with the luminance model.

It is important to realize that this method only allows us to model linear effects in the luminance transfer from projector to camera. This is sufficient for a convincing stereo image, despite higher order effects that are still noticeable in the output (gamma effects and a non-linear camera response).

4.3 Viewpoint Calibration

Next to left-right separation, the perspective should be calculated correctly for each eye. To obtain an optimal depth image, we look at three things (Figure 9): the 3D viewing frustum, the distance and the fusion point of the left- and right-eye projections.

The 3D viewing frustum governs how 3D world coordinates are transformed to the 2D surface of the screen. For a tiled display, we require o-axis frustum projection (as used in traditional VR systems), set up for each projector individually. Parameters of a frustum are the location and orientation of the viewer, and the location and orientation of the screen that should receive the image. As our audience is not tracked, we require a static viewer that represents an average of the audience. Figure 9 shows how a viewer is placed at 5 meters in front of a 6 meter wide screen. A complete mathematical description on how to calculate an off-axis frustum for each projector lies beyond the scope of this paper.

The inter-ocular distance is the distance between the left and right eye. For audiences to view the screen for prolonged periods of time, it is important that this distance is calibrated on several test-viewers. If the distance is chosen to be too wide or too narrow, viewers will feel uncomfortable when viewing the screen.

The fusion point is the exact location that both eyes are looking at. Choosing the fusion point too close to the viewer will give an uncomfortable cross-eye impression, whereas a far fusion point tends to leave the viewer with no reference. For any (traditional) display, viewers will naturally place the fusion point at the surface of the display. We choose the same for the stereo tiled display uncomfortable, and place the fusion point at the center of the display surface.

4.4 Implementation

Geometric, photometric and viewpoint calibration have been implemented in our basic software layer (Aura/VIRPI [5]) that drives the ICWall tiled display. With this system, we were able to produce a convincing stereo image for several test-applications, such as Crayoland (VR test application), molecule visualization and an architectural application. A small and informal survey indicated that viewing was not considered.

5 Summary and Conclusions

This paper demonstrates how a high-resolution stereo tiled display can be realized, using low-cost commodity components. Despite the differences in alignment, color and intensity of each projector, we show that it is possible to generate a convincing stereo image.
To obtain a sub-millimeter accurate alignment, we apply a method based on matched-filter convolution. Furthermore, by modeling the luminance differences across the screen surface, we are able to equalize luminance output considerably for left- and right-eye images. Finally, calculation of the correct viewing frustum by taking into account the viewer position, the inter-ocular distance and the fusion point, makes the stereo image comfortable to view.

References


