The Varrier™ Autostereoscopic Virtual Reality Display

Daniel J. Sandin, Todd Margolis, Jinghua Ge, Javier Girado, Tom Peterka, Thomas A. DeFanti

Electronic Visualization Laboratory
University of Illinois at Chicago
dan@uic.edu

Abstract

Virtual reality (VR) has long been hampered by the gear needed to make the experience possible; specifically, stereo glasses and tracking devices. Autostereoscopic display devices are gaining popularity by freeing the user from stereo glasses, however few qualify as VR displays. The Electronic Visualization Laboratory (EVL) at the University of Illinois at Chicago (UIC) has designed and produced a large scale, high resolution head-tracked barrier-strip autostereoscopic display system that produces a VR immersive experience without requiring the user to wear any encumbrances. The resulting system, called Varrier, is a passive parallax barrier 35-panel tiled display that produces a wide field of view, head-tracked VR experience. This paper presents background material related to parallax barrier autostereoscopy, provides system configuration and construction details, examines Varrier interleaving algorithms used to produce the stereo images, introduces calibration and testing, and discusses the camera-based tracking subsystem.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism --- virtual reality

Keywords: autostereoscopic display, 3D display, virtual reality, camera-based tracking, Varrier, parallax barrier

1. Introduction

The study of stereoscopy and autostereoscopy is not new; Euclid understood back in 280 B.C. that depth perception is achieved by presenting each eye with a disparate image, and in 1903, F.E. Ives patented the first parallax stereogram. [Ives 1903] Today, autostereoscopic displays are commercially available, but most do not function as VR systems because they fail to satisfy key criteria of VR. Although the definition of VR is not universal, the authors' definition requires most, if not all, of the following: head-tracked first-person perspective, large angles of view, stereoscopic display, and real-time interactivity.

EVL has designed and produced a barrier-strip autostereoscopic display system that satisfies all of these criteria. Varrier is both the name of the system as well as the computational method used to produce the autostereoscopic imagery through a combination of a physical parallax barrier and a virtual barrier. The Varrier method is unique in that it produces interleaved left and right eye perspectives in floating point (world) coordinates by simulating the action of the physical barrier screen. In most other lenticular and barrier strip implementations, stereo is achieved by sorting image slices in integer (image) coordinates. Moreover, many autostereo systems compress scene depth in the z direction to improve image quality but Varrier is orthoscopic, which means that all 3 dimensions are displayed in the same scale. Besides head-tracked perspective interaction, the user can further interact with VR applications through hand-held devices such as a 3d wand, for example to control navigation through virtual worlds.

There are four main contributions of this paper. New virtual barrier algorithms are presented that enhance image quality and lower color shifts by operating at sub-pixel resolution. Automated camera-based registration of the physical and virtual barrier strip is performed to calibrate the tiled display quickly and accurately. Camera-based head tracking with artificial neural networks (ANNs) is used to track user head position at interactive frame rates, requiring no gear to be worn for tracking. Finally, all the system components are combined with a distributed clustered architecture and tiled parallax barrier display panels in a new VR display paradigm: tiled barrier strip autostereoscopy. This paper includes the engineering details that make the system functional, cost-effective, and reproducible.

2. Background and previous work

The CAVE [Cruz-Neira et al. 1992, 1993] and other VR technologies of the past such as head mounted displays required the user to wear stereo and sometimes tracking gear. The goal of autostereoscopic VR is to provide the same high-quality visual experience, but without the constraints of user-worn hardware.

Most autostereoscopic display technologies broadly fall into one of three categories: optical, volumetric, or parallax barrier. Optical technologies rely on optical elements to steer light to desired points in space and a stereo image is produced when the viewer's eye positions coincide with the locations of the focused light beams for the left and right perspective. The Cambridge display [Dodgson et al. 2000] is one such example. Volumetric displays such as the Depth Cube by Light Space Technologies [Sullivan 2004] produce a stereo image by stacking numerous flat displays in space to produce depth in the image. The individual displays illuminate "slices" of a 3D shape separated in space, and when all the slices are viewed simultaneously, the result is a stereo image. Parallax barriers produce stereoscopic imagery by juxtaposing stripes of left and right eye perspectives across an image. A comprehensive survey of other methods can be found in Schmidt and Grasnick [2002].

In parallax barrier technology, a planar sheet of alternating transparent and opaque regions is mounted near the display surface, offset from it by a relatively small distance. The displayed image is composed of interleaved stripes of the original left and right eye perspective images such that each eye can see through the transparent regions of the parallax barrier only those image stripes that correspond to that eye's image. Parallax barrier methods are characterized by their relative simplicity and can be

further subdivided into active and passive. An active barrier is employed in Perlin et al. [2000; 2001]

Lenticular displays are closely related to parallax barriers and function equivalently. A lenticular screen is a sheet of cylindrical lenses while a parallax barrier is a flat film composed of transparent and opaque regions. The parallax barrier used in Varrier is not as bright as a lenticular screen but is constructed easily by printing a pattern of opaque and transparent regions onto photographic film. Both methods are shown in Fig. 1. The lenticular method is exemplified by the Synthagram display. [Lipton and Feldman 2002]

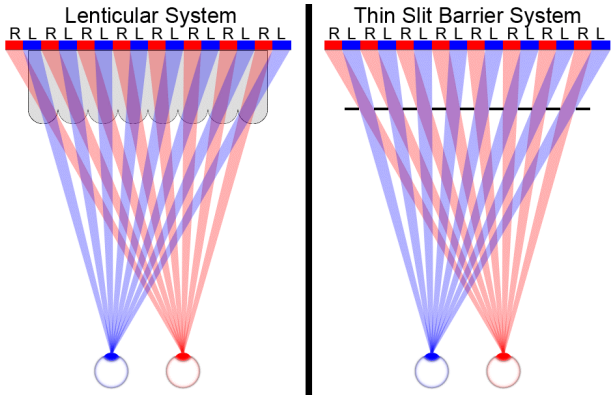

Figure 1: A lenticular screen is a sheet of cylindrical lenses while a parallax barrier is a film of thin slits. Otherwise, their functions are equivalent.

Moiré patterns may result due to the interference between the lenticular and pixel grids. Winnek [1968] patented the tilting of lenticules to minimize this effect. Later work [van Berkel 1999] confirmed this result in lenticular LCD displays and also recognized that disparity in vertical and horizontal resolution can be mitigated by the same method.

Parallax stereograms and panoramagrams both rely on parallax barriers to produce either two or multiple views, respectively, and are precursors to the Varrier method. [Sandin et al. 1989] There is a significant difference between these methods and Varrier, namely, the images in stereograms and panoramagrams are produced by a sorting process; left and right individual images are essentially cut into strips and pasted together in an interleaved order, electronically using computer graphics. This is an integer sorting operation in pixel image space, applied after the scene is rendered, as shown in Fig. 2. In the Varrier method, left and right images are interleaved through a floating point occlusion operation in world space before the scene is rendered.

Another way to categorize autostereoscopic systems is by the existence or absence of tracking technology. Examples of head-tracked autostereoscopic displays are [Son et al. 2001] and [Perlin et al. 2000, 2001]. A variety of tracking technologies exist including acoustic-inertial, infrared, and camera-based, although the ideal technology for autostereoscopy is one that does not require sensors or markers to be worn. Other panoramic displays such as the Synthagram [Lipton and Feldman 2002] are usually untracked, relying on the user to be positioned in pre-determined viewing zones. By juxtaposing many viewing zones of slightly different perspectives, the viewer may have the impression of being tracked because of limited "look-around" capability.

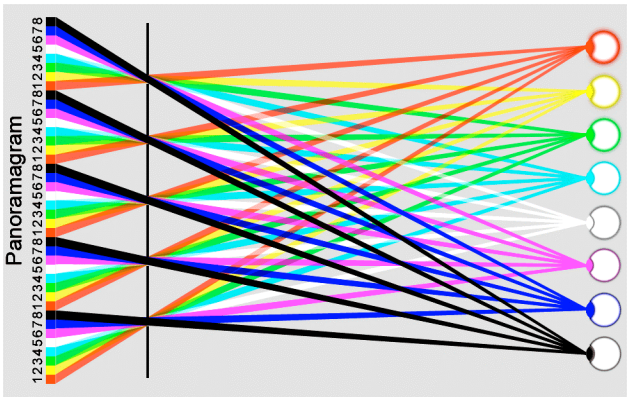

Figure 2: In a parallax panoramagram, multiple viewpoints are cut into strips and pasted together in interleaved order to produce views multiplexed in space, an operation in (integer) image space.

Panoramagrams can static or dynamic, but computational complexity often exceeds real-time interactive frame rates. Also, the discrete transitions between adjacent views, called "screen-flipping," are problematic. Larger numbers of smaller viewing zones reduce screen flipping, but exacerbate the performance issues. Screen flipping is also reduced when the views are highly correlated, so blending views, overlapping views, and compressing scene depth are techniques used to remedy the problem. However, blending or overlapping views increases cross-talk between left and right eye channels, commonly called "ghosting," while reducing scene depth diminishes the 3d effect.

3. Varrier concept

The Varrier method, first published in [Sandin et al. 2001] uses the OpenGL depth buffer to interleave left and right eye perspectives into one rendered image. A virtual parallax barrier is modeled within the scene and corresponds to the physical barrier. For the remaining discussion, the term *virtual linescreen* refers to this virtual barrier; *physical linescreen* refers to the physical barrier, and *scene* is the remainder of the virtual world excluding the virtual linescreen.

The Varrier concept is illustrated in the side-by-side top views of Fig. 3. The right side depicts real or physical space, with the actual positions of the eyes at the top of the figure. Near the bottom of the figure is the physical linescreen, and at the far bottom is the LCD panel. This situation is replicated in virtual space in the left side of Fig. 3 by locating left and right projection points coincidental with the eye positions and a virtual linescreen identical to the physical linescreen. A 3d model comprising the scene is also shown. In the virtual space, left and right projections of the virtual linescreen and scene are computed and various pixels on the LCD are illuminated. Then, in real space, those pixels distribute their light through the physical linescreen, which directs the light back to the eyes. When the virtual linescreen and virtual projection points are correctly registered with the physical linescreen and eye positions, the result is an autostereoscopic image.

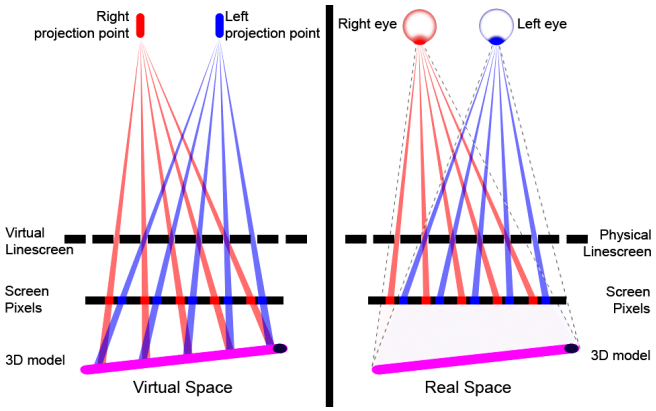

Figure 3: The Varrier concept is illustrated in two steps, computation in virtual space (left) followed by viewing in physical space (right). In virtual space, perspective sight lines are directed from the projection positions to the scene objects which are occluded by the virtual linescreen to produce a correctly interleaved image. The displayed image is then directed in real space through the physical linescreen to the eyes.

The purpose of the virtual linescreen is to occlude scene objects without being visible itself. It is drawn only into the depth buffer, followed by the scene, drawn in the usual way. This is performed for both eye viewpoints before a swap buffers command is issued, making the image visible. If the virtual linescreen were modeled in the same position as the physical linescreen, (near the LCD display panel as shown in Fig. 3) it could only occlude scene objects that appear behind it in the virtual world. Often scene objects appear in front of the display and these need to be occluded by the virtual linescreen as well. The solution is to perspectively transform the virtual linescreen to be near the eye. It is scaled and translated along perspective lines to be a minimal distance in front of the eye; otherwise the algorithm proceeds normally. This transformation is discussed in detail in [Sandin et al. 2001].

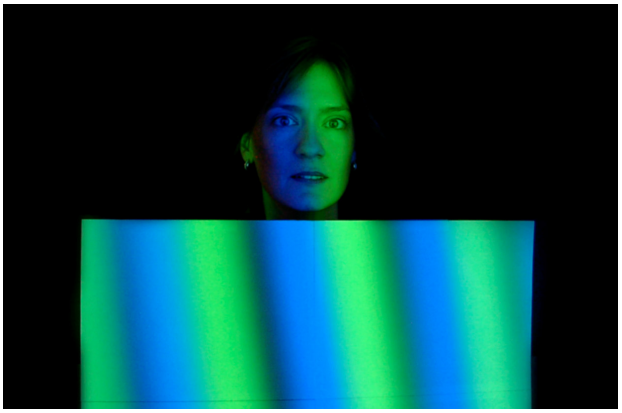

Figure 4: The stereo images are focused on the viewer's face and on a large white card that she is holding. The left eye scene is a single green polygon and the right eye scene is a single blue polygon. The images are steered correctly in space and coincide with the eye positions over a large working range, repeating laterally outward as a function of the physical parallax barrier.

The result is that beams of light are steered in space and made to intersect the eyes of the viewer, as shown in Fig. 4. The left and right perspective viewpoints correspond to the left and right eye positions correctly because the interleaving process is a floating point computation that uses at least two pixels per eye, as required by the Nyquist Sampling Theorem. Fig. 4 shows that two images correspond correctly with the eye positions by drawing each viewpoint as a separate color. Unused image pixels remain as darker *guard bands.* Fig. 4 also shows that parallax barrier displays generate secondary, tertiary, etc. views that repeat laterally outward. The edges of the viewing zones in Fig. 4 are tilted at the same angle as the physical linescreen, and this does not pose a problem as the width of the viewing zones is larger than the entrance pupil of the eye.

4. Varrier algorithms

Three different algorithms can be used to implement the Varrier concept and are identified by the number of virtual linescreen and scene passes required per eye to accomplish the interleaving process. Each algorithm affects image quality or performance differently. The first was previously published by Sandin et al. [2001]

The 1 linescreen pass / 1 scene pass algorithm is as follows:

- left eye:
  1. Clear color and depth buffer
  2. Draw linescreen from left eye perspective into depth buffer only
  3. Draw scene from left eye perspective into both color and depth buffer
- right eye:
  4. Clear depth buffer only
  5. Draw linescreen from right eye perspective into depth buffer only
  6. Draw scene from right eye perspective into color and depth buffer
- swap buffers

This algorithm is the most efficient of the three, but creates color artifacts because it assumes that pixels are homogenous, when in fact pixels are composed of RGB sub-pixels. Color banding has been studied in the context of lenticular panoramic displays. [van Berkel 1999] The physical linescreen disperses colors in space causing visible color banding, as seen in Fig. 5. One solution is to occlude per-component, repeating three times for RGB components and shifting the linescreen 1/3 of a pixel between each pass, also shown in Fig. 5.

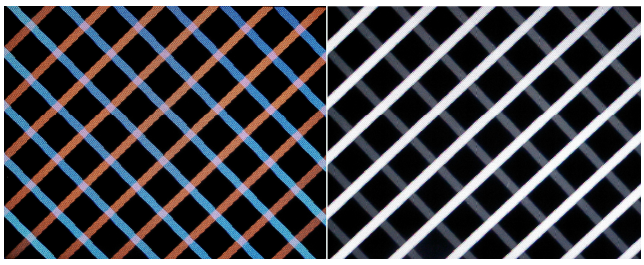

Figure 5: The 1/1 algorithm causes color shifts that are minimized by the 3/3 algorithm. The test pattern contains diagonal bars in opposite directions for each eye. The bars should be white; colors are shifted in the left image (1/1 algorithm) to red and blue. The right image shows the result of the 3/3 algorithm.

The 3 linescreen passes / 3 scene passes algorithm is as follows:

- left eye:
  1. Clear color and depth buffer
  2. Draw red component of linescreen from left eye perspective into depth buffer and red component of scene into color and depth buffer
  3. Shift linescreen 1/3 pixel and draw green component of linescreen from left eye perspective into depth buffer and green component of scene into color and depth buffer
  4. Shift linescreen 1/3 pixel and draw blue component of linescreen from left eye perspective into depth buffer and blue component of scene into color and depth buffer
- right eye:
  5. Clear depth buffer only
  6. Perform steps 2, 3, 4 from right eye perspective
- swap buffers

A comparison of 1 pass vs. 3 passes is shown in Figs. 5 and 6 for two different test patterns. In both cases, the patterns should be pure black and white, and the 3-pass algorithm produces less color shift and improves guard band effectiveness. However, the 3-pass algorithm is computationally expensive because the scene must be drawn three times for each eye, instead of once for each eye.

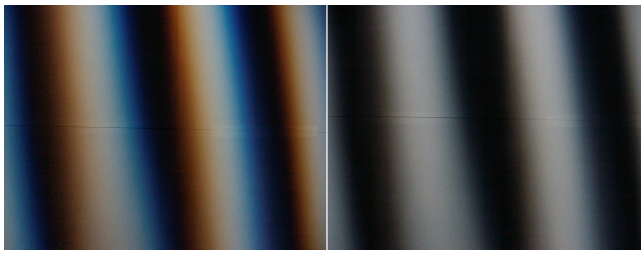

Figure 6: This is the same test pattern as in Fig. 4, but with black and white polygons used for right eye and left eye. The left image (1/1 algorithm) produces color shifts, which are improved in the right image (3/3 algorithm).

The same result is attained more efficiently by realizing that only one scene pass is required for each eye if multiple linescreen passes occur *after* the scene pass. Performance is improved because the number of linescreen polygons (approximately 700) is small compared to the average scene. The depth buffer is still used, but in an unconventional way. The 4 linescreen / 1 scene pass algorithm performs 3 main steps for each eye, in this order: draw the scene conventionally, modulate the scene with virtual linescreens, and protect the required zones with the depth buffer.

The 4 linescreen passes / 1 scene pass algorithm is as follows:

- left eye:
  1. Clear the color and depth buffers to far clipping plane
  2. Enable depth test and depth write and draw scene from left eye perspective
  3. Clear depth buffer to near clipping plane and disable depth test
  4. Draw linescreen red component from left eye perspective, shifted +1/3 pixel, into color buffer only
  5. Draw linescreen green component from left eye perspective, unshifted, into color buffer only
  6. Draw linescreen blue component from left eye perspective, shifted –1/3 pixel, into color buffer only
  7. Draw line screen from left eye perspective into depth buffer only at far clipping plane, using a slightly narrower slit than used for steps 4-6b
- right eye:
  8. Enable depth test and depth write and draw scene from right eye perspective
  9. Draw linescreen red component from right eye perspective, shifted +1/3 pixel, into color buffer at near clipping plane
  10. Draw linescreen green component from right eye perspective, unshifted, into color buffer at near clipping plane
  11. Draw linescreen blue componen from right eye perspective, shifted –1/3 pixel, into color buffer at near clipping plane
- swap buffers

Steps 1 and 2 are the conventional way to render a scene. Step 3 "protects" the entire scene from being overwritten by clearing the depth buffer to the near plane, but permits steps 4-6 to overwrite channels that will be later used for the right eye. Then steps 4-6 "black out" stripes for later use by the right eye, but the process is performed color component-wise as in the previous 3/3 algorithm. Step 7 "unprotects" those stripes by setting their depth values to the far plane. The process is then repeated for the right eye, but this time using the near plane instead of the far. Linescreens written at the near and far planes are perspectively transformed to those locations. The slit size used in the linescreen for step 7 is approximately 85% of the original size, for empirical reasons. Step 7 does not need to be repeated for the right eye because the drawing cycle is complete.

The 3/3 and 4/1 algorithms reduce color banding by functioning at sub-pixel resolution, as shown in Fig. 7. The sub-pixels of the Varrier LCDs are organized in vertical columns with R,G,B arranged from left to right. On the left side of Fig. 7, a single pixel-sized slit directs colors, with red and blue penumbral fringing. (as in Fig. 6 left) This is the case in the original 1/1 algorithm. On the right side of Fig. 7, the virtual slits are shifted while sub-pixels are illuminated component-wise, effectively reducing the overall slit size without reducing intensity of individual color components. The net result is that light is more focused as it is directed to the eyes, and the corresponding right side of Fig. 6 shows color banding is reduced. Note that the depth buffer does not limit sub-pixel precision although it contains only one sample per pixel, because the modulation of the scene is performed in the color buffer. The purpose of the depth buffer is to protect one eye's channel from being overwritten by the other eye.

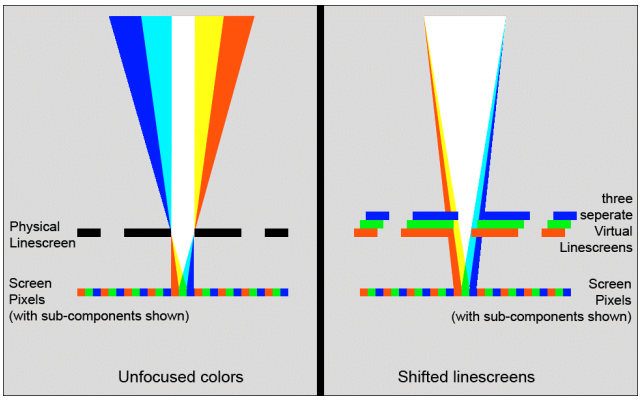

Figure 7: The causes of color banding and a solution are shown. The left side shows a pixel-size virtual and physical slit and produces color banding as outer components (red and blue) spread. The right side shows three virtual slits that are shifted as color components are drawn. Colors are focused and directed to the eyes with little color banding.

5. System configuration

The Varrier display is a 35 panel tiled system driven by a Linux cluster. Two display panels are powered by one computation node via a dual-head Nvidia Quadro FX3000 graphics card. One additional node serves as the master for the entire system. The 35 panel system is composed of 19 nodes, each containing dual Intel Xeon processors, connected by Gigabit Ethernet. Applications are built around the CAVELib™ platform. Inter-node communication is accomplished using the distributed CAVELib architecture which provides services for communicating tracking and synchronization information across a large number of screens, and TeraVision multicasting [Singh et al. 2004] is used to communicate application data.

The display panels are mounted in a semicircular arrangement to partially encompass the viewer, affording approximately 120° - 180° field of view. The number of panels is scalable so that coverage up to 360° is theoretically possible. The total pixel count of the system is 11200 x 6000, or approximately 67 Mpixels. However, the linescreen duty cycle is 77% opaque in the horizontal direction, so the net resolution is approximately 2500x6000, or 15 Mpixels. Images and specifications of the system are in Figs. 8, 9, and Table 1.

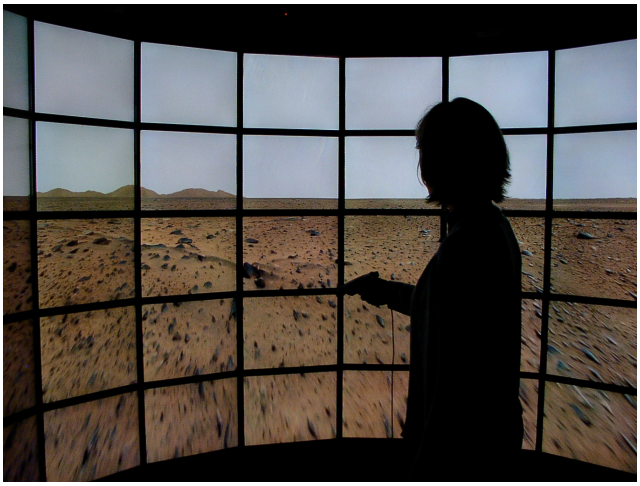

Figure 8: The Varrier display has 35 panels mounted in a semi-circular arrangement to provide wide angles of view in an immersive VR environment.

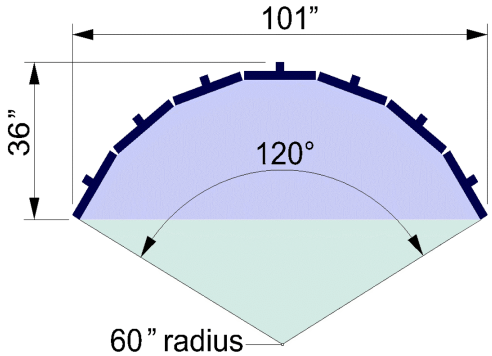

Figure 9: The system footprint is shown. The ideal user location, or sweet spot, is at the center of the 60 inch radius of curvature of the panels, although the user is free to move within an area approximately 32 inches wide by 48 inches deep.

Feature	Value
panel configuration	35 panels (5 high x 7 wide)
radius of curvature of display	60 in. (1.52 m)
angular difference between columns	20 degrees
panel size	16 in. x 12 in. (20 in. diagonal)
panel resolution	1600 x 1200 pixels
total gross resolution	11200 x 6000, 67 Mpixel
total net resolution	2500 x 6000, 15 Mpixel
overall size	101 in.(2.54m)W x 90 in.(2.29m)H
LCD pixel pitch	.010 in. (.254 mm)
linescreen pitch	.0442 in. (1.123 mm) (22.6055 lines / in.)
linescreen duty cycle	77.78% opaque, 22.22% transparent
linescreen angle	7.82 degrees from vertical
glass thickness	.126 in. (3.20 mm)
air space	.355 in. (9.02 mm)
glass refractive index	1.51
optical thickness	.438 in. (11.125 mm)
minimum view distance	40 in. (1.02 m)
maximum view distance	88 in. (2.24 m)
optimal view distance	64 in. (1.64 m)
working width	32 in. (.8m)
interocular distance	2.5 in. (6.35 cm)

Table 1: System specifications are listed.

An individual display panel is a stock NEC2080 LCD monitor removed from its plastic housing and a parallax barrier is affixed to the front. The parallax barrier is constructed by printing a pattern of black rectangular strips on a transparent film and then laminating the film to a thin piece of glass substrate to provide strength. The modifications to the LCD panel are relatively inexpensive and easy to perform, resulting in a panel assembly as shown in Fig. 10.

The physical linescreen is intentionally mounted so that the lines are not vertical and does not require pixel or sub-pixel registration with the LCD grid, as in most other lenticular and barrier strip systems. Two advantages are gained by titlting the linescreen. Moiré patterns caused by interference between the linescreen and the pixel grid are converted from highly visible vertical bars to a fine diamond pattern that is much less noticeable, as in [van Berkel and Clarke, 1997] and [van Berkel, 1999]. Color shifts are also reduced because the linescreen orientation is different from the arrangement of RGB sub-pixels. The best angle of tilt is found empirically, by rating visibility of

primary (static) and secondary (dynamic) Moiré patterns and color shift for various angles, and selecting an angle that minimizes these criteria. The optimum ranges of angles are quite narrow, approximately 1° in size; best results are achieved when linescreens are mounted to within 1/4°.

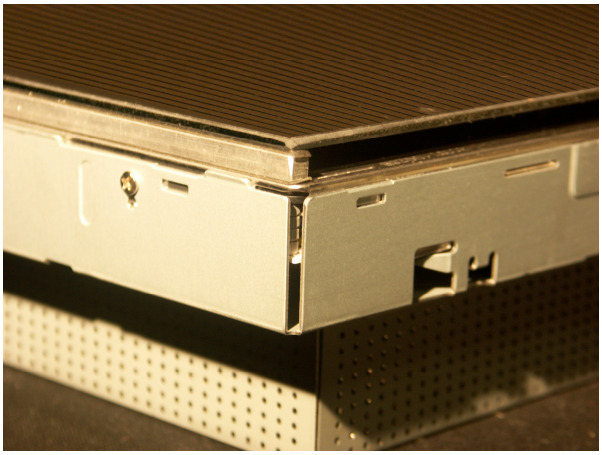

Figure 10: A completed panel assembly is ready for installation. A commodity LCD panel is removed from its housing and a linescreen assembly, consisting of aluminum spacers and a thin glass pane with a laminated film, is attached to the front.

The period or pitch of the linescreen is .0442 inches, which is just over 4 LCD pixels and near the minimum governed by the Nyquist Sampling Theorem, which implies that the linescreen pitch must cover at least 2 LCD pixels per eye. The exact linescreen pitch is slightly larger than 4 LCD pixels because it is tied to the pre-press scanner pitch used to print the film linescreen, to eliminate aliasing in the printing process. The duty cycle is computed by fixing the transparent section to the width of 1 LCD pixel, again rounded up to the nearest printer pixel to reduce printing artifacts. The actual duty cycle used is approximately 22% transparent and 78% opaque.

The theoretical limits on view distance occur when the lines of sight from left and right eyes pass through the physical linescreen and map to the same pixel on the LCD display. These conditions are shown geometrically in Fig. 11 and are solved by similar triangles to produce equations (1) and (2).

$$\text{max dist.} = t * e / s \qquad (1)$$

$$\text{min dist.} = t * (e - p) / (p - s) \qquad (2)$$

where e is the interocular distance, p is the linescreen pitch, s is the pixel pitch, and t is optical thickness = glass thickness / refractive index + air space

The resulting minimum and maximum distances are 27 in. and 100 in., respectively. The viewing range listed in Table 1 is reduced due to effective tracker coverage. (see Sect. 7) At the optical limits, interference of left and right eye channels occurs. These optical limits have been tested, but the amount of ghosting gradually increases as the limits are approached and performance there is highly subjective. The optimal view distance (sweet spot) is the average of the two limits, and can be adjusted by changing the distance between the LCD and the physical linescreen. The difference between the minimum and maximum distance can be increased by increasing the linescreen pitch, i.e., working volume and spatial resolution trade off.

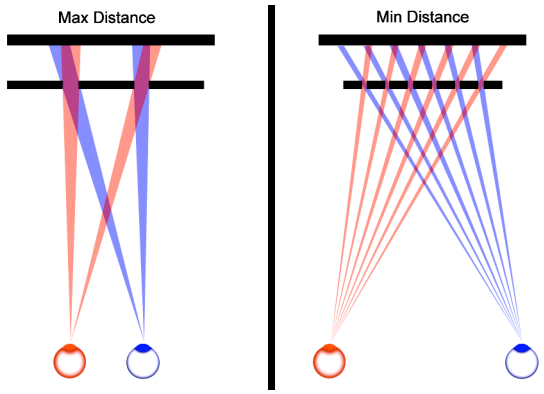

Figure 11: Limits on view distance are computed by finding distances where sight lines map to the same pixel.

Left and right, the tracking range of the system is approximately +/- 16 inches from center, while the optical performance is slightly wider, approximately +/- 24 inches from center. Outside of the optical range, ghost levels between left and right eye channels become unacceptable. Tracking, discussed in Section 7, can be expanded with more cameras but the causes of the limited optical off-axis performance have been investigated but still remain uncertain. The limitation is alleviated by the semi-circular panel arrangement because the viewer usually concentrates on-axis at a given subset of panels. However, this could be problematic in a wide, flat configuration.

The borders between individual panels of a tiled LCD display are a significant visual feature. Some years ago, EVL performed an informal simulation of tiled borders within the CAVE and determined that the effects were not detrimental to the immersive experience. Informal feedback provided by numerous users of Varrier confirms this hypothesis. Because the viewer is actively tracked, he or she is able to look around objects and borders as needed. However, to date no formal human factors studies have been published that statistically document the effects of tiled borders. One of our team members, in conjunction with members of the HCI domain, is presently performing such a study where various size physical grids are placed in front of a large stereo projection display to simulate a tiled display. Over time, manufacturers have produced LCD panels with smaller borders, and will probably continue to do so. At present however, methods for eliminating borders altogether are optical in nature, refracting light out to the edges of a panel, and these are incompatible with the optics of the Varrier system.

6. Virtual and physical linescreen registration

In parallax barrier strip autostereoscopy, registration of physical and computational parameters is critical to successful operation of the system, and can be a daunting task. Registration needs to be performed per panel, so in a tiled configuration, accurate and efficient registration procedures are mandatory. Because Varrier utilizes a virtual model of the physical linescreen, the virtual model is registered in software after the system is built to correspond with the physical barrier. This is easier than physically registering the actual linescreen with the pixel grid or sub-pixel grid during manufacturing, and is one advantage of the Varrier computational method. The process is automated using

computer vision techniques and two cameras separated by the interocular distance such that the cameras simulate a human viewer (Fig. 12). (The term *eye* and *camera* are synonymous in the following discussion) With automated registration, the entire 35 panel system is calibrated in approximately one hour.

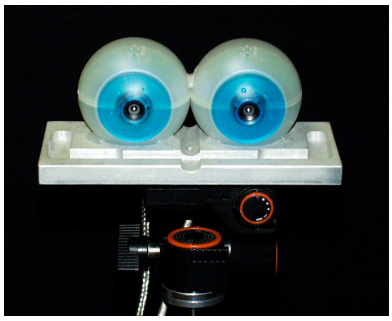

Figure 12: Two cameras mounted at the correct interocular distance simulate a human viewer.

Two calibration patterns are used to adjust the virtual linescreen parameters. The color pattern consists of a different color polygon for each eye, and the cross-bar pattern consists of orthogonal white bars at opposite angles for each eye. Image processing techniques such as edge detection and intensity thresholding are applied to the acquired images of test patterns to extract information and automatically update parameters, repeating until convergence.

The three virtual linescreen parameters are: rotation angle (corresponding to rotation of physical linescreen), position in a direction normal to the display plane (corresponding to optical thickness of physical linescreen), and position in a direction parallel to the display plane (corresponding to lateral shift of the physical linescreen).

The color pattern is used for rough calibration. Small differences of rotation and translation can cause large-scale red and blue Moiré bars, and by adjusting the parameters iteratively with the following algorithm, the Moiré gradually disappears.

This calibration method is as follows:

1. Rotate virtual linescreen until the Moiré bar angle is the same as physical linescreen angle.
2. Translate virtual linescreen in direction normal to display plane until at least one eye's image contains no Moiré bars.
3. Translate virtual linescreen in direction parallel to display plane, maximizing F = left eye's red – left eye's blue + right eye's blue – right eye's red

Usually 45-50 iterations, or less than one minute, are required to complete this phase, resulting in the patterns shown in Fig. 13 for left and right eye.

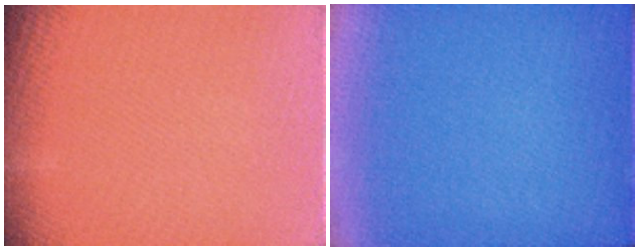

Figure 13: Left and right eye images after rough calibration; distortions still exist at the edges of the images.

Next, rotation is fixed and the best optical thickness and shift values are searched in finer steps by maximizing F = brightness – ghost. Again, computer vision methods are used to delineate bar edges in the images and determine brightness and ghost intensities.

The cross-bar pattern is used:

4. Set the step for normal and parallel translations to 1/2 the step used previously.
5. At the current normal position, search best parallel shift by maximizing F_1 = (bright angled bar intensity– ghost angled bar intensity). Maximize F_1.
6. Change normal position by one step and search parallel shift around previous best shift value. Get $\max F_2$. If $\max F_2 < \max F_1$, use current values and stop. Otherwise, let $\max F_1 = \max F_2$ and repeat from step 5.

Usually less than 60 iterations, or less than one minute, are required for steps 4-6 and the images in Fig. 14 show left and right eye results for both calibration patterns.

The registration process runs at 3 frames per second (fps). One screen is calibrated in 1-2 minutes; approximately one hour is required for the entire 35 panel system.

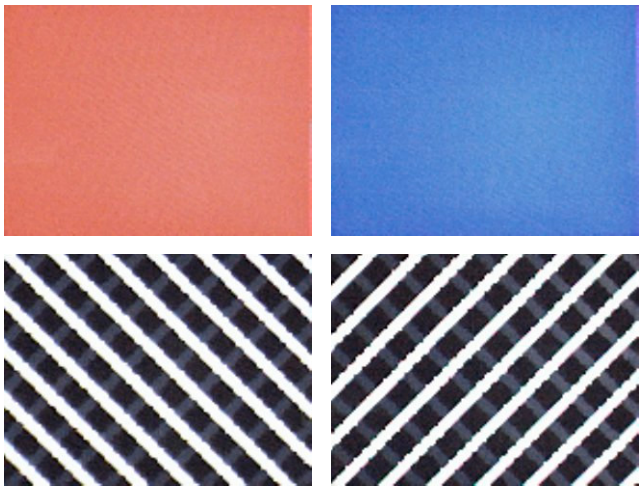

Figure 14: Color pattern and cross-bar pattern for left and right eye after registration is complete. Color pattern is uniform and cross bar pattern contains approximately 5% ghost.

The camera calibration process effectively measures the "as-built" dimensions of the system. Assuming tracking is accurate and the optical system is aberration free, this single position will optimize the system for all viewing positions. Calibrating the system from multiple viewer positions could improve performance, correcting for tracker errors and the optical aberrations that are present in the system. At various distances on center, variations of the final results were within noise levels of the system. Specifically, position normal to the display is within .5%, position parallel to display is within 2.7%; rotation angle is within 1.3%. At locations off-center, variability is higher; this is related to the limited off-center performance of the system optics, and is still being studied. Calibrating an array of positions could improve off-axis performance by finding corrective factors to apply across a wider working area, but this presently is not done.

Certain other system parameters such as screen positions and linescreen pitch are not part of the automatic registration process and are measured by other means. For example, registering virtual and physical linescreen pitch requires optical tests due to the high

degree of precision required. This parameter is critical to correct system function, and needs to be within .2 percent to reduce cross-talk levels to the 5% range. The physical linecreen pitch is determined by construction, however virtual linescreen pitch is affected by several factors such as differences in actual pixel pitch and display size vs. those reported by the LCD manufacturer. These confounding effects cause an adjustment of virtual linescreen pitch to be required. The test is performed visually: A sample linescreen film is overlaid on the display, and a visible linescreen is rendered. Virtual pitch is adjusted until the two coincide over the entire length of the display, as viewed under magnification to see the underlying sub-pixel structure. The technique is very accurate because large-scale Moiré patterns result from small errors in pitch.

The corner points of all 35 panels must also be accurately measured in 3D space; a task accomplished using a digital theodolite, a common surveying instrument, interfaced to a laptop computer. A theodolite improves efficiency and accuracy compared to other methods. Measurement using a tracker is possible and was used for earlier prototype versions. Locations in 3d space can be found within +/- 1 mm with the theodolite; in practice trackers are in the 2-3 mm range. Trackers also suffer from magnetic or acoustical interference near LCD displays. A method was developed to compute corner points in 3d space of a rectangle of known length and width from the horizontal and elevation angles provided by the theodolite, and approximately 2 hours are required to find all screen corners for the entire 35 panel system.

7. Camera-based tracking

Both VR and tracked autostereo require real-time knowledge of the exact 3d position of the eyes. In Varrier, position data is acquired using a camera-based tracking system that requires no sensors or markers to be worn, and completes the goal of freeing the user from wearing any gear to experience VR. Tracking is implemented using Artificial Neural Networks (ANNs), allowing the detection and recognition of faces in visually cluttered environments. Several ANNs per left and right camera are used for recognition, tracking, and real-time face and background training. Fast frame rates are achieved once a face is recognized, permitting real-time tracking of 120 fps at 640x480 video image resolution.

The tracking system has evolved since first published by [Girado 2004; Girado et al. 2003]. Originally, Girado proposed a supervised LAMSTAR ANN with high reliability but long training times and low frame rate performance. Varrier is very sensitive to tracker performance, and a faster method was required. Currently, two unsupervised self-organizing map (SOM) ANNs are used per camera. The recognition ANN contains 256 neurons and is used to recognize the desired face within the entire image. Once recognized, the system switches to a small, fast detection ANN with only 8 neurons. This ANN detects the face only within a small predicted region, and operates at 120 fps.

Illumination in the tracking environment is provided using six flat panel infrared (IR) illuminators with corresponding IR-pass, visible-cut filters on the cameras. This de-couples the resulting camera image intensity from room illumination and illumination from the display, both of which are variable. IR illumination variations are measured using a standard gray card and mapped accordingly.

In addition to training on a user's face, the system is trained on the background to reduce the probability of false positive results. Face recognition and tracking is based on the following steps: image acquisition, preprocessing, searching, recognition, arbitration, tracking, computation of 3d head position, and median filtering.

Two Point Grey 200 fps cameras, along with six IR illuminators, are mounted near the top of the Varrier system as shown in Fig. 15. Wide-angle lenses (f = 4 mm) provide coverage of Varrier's working area. Performance results from the current implementation demonstrate a useable camera-tracked area of 32 inches wide x 48 inches deep. The cameras' field of view extends beyond this, but detection and recognition become difficult when faces are captured from the side.

Another advantage of this system is the relatively short and easy training procedure, requiring under two minutes to train on a new face. The user slowly moves and tilts his or her head in a variety of comfortable poses while 512 frames of video are captured. Then, an ellipse is manually centered on the face and the contents of the elliptical region together with the video frames are used to automatically train the unsupervised 256-neuron recognition ANN. Once trained, the user data is stored and can be reloaded at any time.

There are still several limitations of the tracking system, and work is ongoing to solve these problems. A total of six cameras are planned to expand coverage such that tracker coverage exceeds Varrier's optical performance. Head movements faster than the predicted position can cause confusion between face and background and produce erroneous results. To improve reliability, ANNs will be distributed over a small cluster of computers dedicated to tracking. Two approaches can take advantage of the additional processors; either parallel SOMs can produce position data and a majority vote can determine the result, or more reliable supervised ANNs can be used.

Feature	Value
tracking frame rate	120 fps
recognition frame rate	6 fps
video image resolution	640x480
training time(new user)	under 2 minutes
type of prediction	always on; predicts next area of the input image to search for the face
input sensor	Kodak KAI-0340D CCD image sensor
input protocol / interface	IEEE 1394b (fast FireWire)
output protocol / interface	UDP/IP over 100Mbps Ethernet
tracking latency	81 ms end-to-end
static precision	.25 inch (6 mm) in x,y 1.0 inch (25 mm) in z
working volume	32 inches(.8m) left-right, 48 inches(1.2m) front-back, 48 inches(1.2m) top-bottom

Table 2: Tracker performance is summarized.

A final limitation is latency. Performance results indicate that current end-to-end latency is 81 ms, from head movement to visible movement on the display. This is measured using the technique from [He at al. 2000], at a rendering rate of 60 fps using the 1/1 algorithm. Further testing indicates that this latency can be further divided into: 28 ms tracking, 37 ms communication delay, and 16 ms rendering time. Since the largest component of the latency is due to communication delays required to distribute tracker data to the cluster nodes, total latency can be reduced by optimizing the communication architecture. Distribution of tracker data is currently performed by CAVELib, and performance of this and other methods needs to be carefully evaluated.

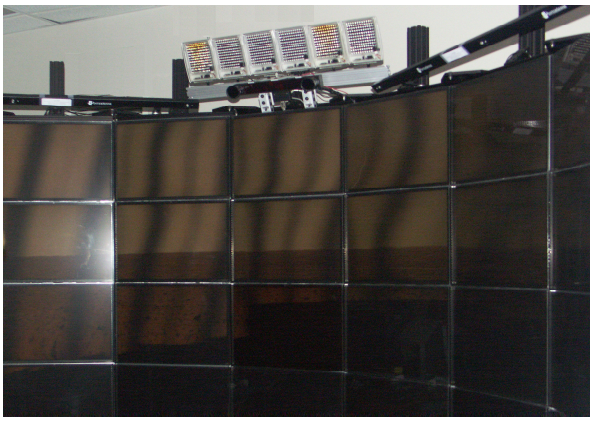

Figure 15: Camera based tracking is used on the Varrier system. Six infrared illumination panels provide controlled illumination of the tracked subjects even under varying lighting conditions. Two cameras capture the scene and artificial neural networks process the real-time images to recognize and track faces.

8. Results, conclusions and future work

Results

The Varrier system succeeds at producing 2500x6000 autostereo imagery over a 120° field of view at interactive frame rates without requiring the user to wear any stereo or tracking accessories. Images have approximately 5% ghost for ranges of scene objects from infinity to 1 ft. in front of the viewer, with the viewer free to move within a working volume approximately 32 inches wide by 48 inches deep.

Varrier is an orthostereo system, so depth dimensions are displayed in the same scale as horizontal and vertical dimensions. The contrast of the system has been measured to be over 200:1. Varrier satisfies the criteria of an autostereo VR display system, affording large angles of view, viewer-centered perspective, and autostereoscopic display, tether-less tracking, and real-time interactivity.

Image quality is further improved with enhanced interleaving algorithms, reducing color shifts and reducing cross-talk or ghosting. Because Varrier is a passive parallax barrier system built from standard display devices, the system is simple to build and cost effective. Calibration is largely automatic using computer vision methods and two cameras to simulate a human viewer. Tracking is camera-based, providing face recognition and detection using ANNs to provide the user with real-time first-person perspective without requiring the wearing of any sensors or markers.

Limitations

Off-center performance is limited, both by the tracker working range and the optical performance of the system. Tracker and system latency is noticeable, especially during moderate-speed head movements as the image darkens while the user's eyes are passing into the guard band regions before the system is able to update the images. A second viewer cannot be granted his own perspective because of cross-talk with the first viewer's images. Passive viewing by a second viewer is possible since the viewing zones are repeated, but problematic when the primary or tracked viewer moves his head, disrupting stereo for the passive viewer. Although the viewing volume can accommodate several other people as passive viewers who can see recognizable images, image quality is generally poor for non-tracked viewers.

Current and Future work

Current and future research centers around solving some of the drawbacks mentioned above. Experimentation is ongoing with complex physical linescreen patterns to direct light out in different configurations in space to permit multiple viewers to have independent perspectives and improved display clarity. Reduction of off-center viewing artifacts continues to be studied, including curving and modifying the shape of the virtual linescreen to increase off-axis performance.

Camera-based tracking is continuously being improved to include multiple cameras, improved algorithms, higher frame rates, and the extraction of head orientation information over a wider coverage area. Reduction of tracker latency is also an active topic; proposed improvements include higher speed cameras, new algorithms, distributed processing, and tighter control over distribution of positional data to cluster slave nodes.

Sub-pixel LCD organization is continued to be studied in connection with new interleaving algorithms, including the use of low-level GPU shader languages such as the CG [Fernando and Kilgard 2003] language to optimize operations on sub-pixel RGB components. Modeling the virtual linescreen as a texture is also being studied for potential use in an interleaving algorithm.

Finally, new form factors are being investigated, for example, a 30 inch single panel desktop display was recently built and is currently being tested.

9. Acknowledgement

The Electronic Visualization Laboratory (EVL) at the University of Illinois at Chicago specializes in the design and development of high-resolution visualization and virtual-reality display systems, collaboration software for use on multi-gigabit networks, and advanced networking infrastructure. These projects are made possible by major funding from the National Science Foundation (NSF), awards CNS-0115809, CNS-0224306, CNS-0420477, SCI-9980480, SCI-0229642, SCI-9730202, SCI-0123399, ANI 0129527 and EAR-0218918, as well as the NSF Information Technology Research (ITR) cooperative agreement (SCI-0225642) to the University of California San Diego (UCSD) for "The OptIPuter" and the NSF Partnerships for Advanced Computational Infrastructure (PACI) cooperative agreement (SCI 9619019) to the National Computational Science Alliance. EVL also receives funding from the State of Illinois, General Motors Research, the Office of Naval Research on behalf of the Technology Research, Education, and Commercialization Center (TRECC), and Pacific Interface Inc. on behalf of NTT Optical Network Systems Laboratory in Japan. Varrier and CAVELib are trademarks of the Board of Trustees of the University of Illinois.

10. References

CRUZ-NEIRA, C., SANDIN, D., and DEFANTI, T. 1993. Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE. In *Proceedings of ACM SIGGRAPH 1993*, ACM Press / ACM SIGGRAPH, New York. Computer Graphics Proceedings, Annual Conference Series, ACM, 135-142.

CRUZ-NEIRA, C., SANDIN, D., DEFANTI, T., KENYON, R., and HART, J. 1992. The CAVE: Audio Visual Experience Automatic Virtual Environment, *Communications of the ACM*, vol. 35, no. 6, 64-72.

DODGSON, N. A., MOORE, J. R., LANG, S. R., MARTIN, G., and CANEPA, P. 2000. A 50" Time-Multiplexed Autostereoscopic Display. In *Proceedings of SPIE Symposium on Stereoscopic Displays and Applications XI*, San Jose, California.

FERNANDO, R., and KILGARD, J. 2003. *The CG Tutorial.* Addison-Wesley.

GIRADO, J. 2004. *Real-Time 3d Head Position Tracker System With Stereo Cameras Using A Face Recognition Neural Network*. PhD thesis, University of Illinois at Chicago.

GIRADO, J., SANDIN, D., DEFANTI, T., and WOLF L. 2003. Real-time Camera-based Face Detection using a Modified LAMSTAR Neural Network System. In *Proceedings of IS&T/SPIE's 15th Annual Symposium Electronic Imaging 2003, Applications of Artificial Neural Networks in Image Processing VIII*, San Jose, California, pp. 20-24.

HE, D., LIU, F., PAPE, D., DAWE, G., and SANDIN, D. 2000. Video-Based Measurement of System Latency. *International Immersive Projection Technology Workshop 2000,* Ames, Iowa.

IVES, F.E. 1903. U.S. patent number 725,567.

LIPTON, L., and FELDMAN, M. 2002. A New Autostereoscopic Display Technology: The SynthaGram. In *Proceedings of SPIE Photonics West 2002: Electronic Imaging,* San Jose, California.

PERLIN, K., PAXIA, S., and KOLLIN, J. 2000. An Autostereoscopic Display. In *Proceedings of ACM SIGGRAPH 2000*, ACM Press / ACM SIGGRAPH, New York. Computer Graphics Proceedings, Annual Conference Series, ACM, 319-326.

PERLIN, K., POULTNEY, C., KOLLIN, J., KRISTJANSSON, D., and PAXIA, S. 2001. Recent Advances in the NYU Autostereoscopic Display. In *Proceedings of SPIE, vol. 4297*, San Jose, California.

SANDIN, D., MARGOLIS, T., DAWE, G., LEIGH, J., and DEFANTI, T. 2001. The Varrier Autostereographic Display. In *Proceedings of SPIE*, *vol. 4297*, San Jose, California.

SANDIN, D., SANDOR, E., CUNNALLY, W., RESCH, M., DEFANTI, T., and BROWN, M. 1989. Computer-Generated Barrier-Strip Autostereography. In *Proceedings of SPIE, Three-Dimensional Visualization and Display Technologies, vol. 1083*, pp. 65-75.

SCHMIDT, A. and GRASNICK, A. 2002. Multi-viewpoint Autostereoscopic Displays from 4D-Vision. In *Proceedings of SPIE Photonics West 2002: Electronic Imaging,* San Jose, California.

SINGH, R., JEONG, B., RENAMBOT, L., JOHNSON, A., and LEIGH, J. 2004. TeraVision: a Distributed, Scalable, High Resolution Graphics Streaming System, In *Proceedings of Cluster 2004*, San Diego, California.

SON, J.-Y., SHESTAK, S.A., KIM, S.-S., CHOI, Y.-J. 2001. Desktop Autostereoscopic Display with Head Tracking Capability. In *Proceedings of SPIE Vol. 4297, Stereoscopic Displays and Virtual Reality Systems VIII,* San Jose, California.

SULLIVAN, A. 2004. DepthCube Solid-State 3D Volumetric Display. In *Proceedings of SPIE Electronic Imaging 2004*, San Jose, California.

VAN BERKEL, C. Image Preparation for 3D-LCD 1999. In *Proceedings of SPIE Vol. 3639 Stereoscopic Displays and Virtual Reality Systems VI,* San Jose, California.

VAN BERKEL, C. and CLARKE, J.A. 1997. Characterization and Optimization of 3D-LCD Module Design. In *Proceedings of SPIE Vol. 3012, Stereoscopic Displays and Virtual Reality Systems IV,* San Jose, California.

WINNEK, D.F. 1968. U.S. patent number 3,409,351.