

Dynamic Mapping for Multiview Autostereoscopic Displays

Jing Liu Tom Malzbender Siyang Qin Bipeng Zhang Che-An Wu James Davis

University of California Santa Cruz

June 09, 2014

TR Number: UCSC-SOE-14-16

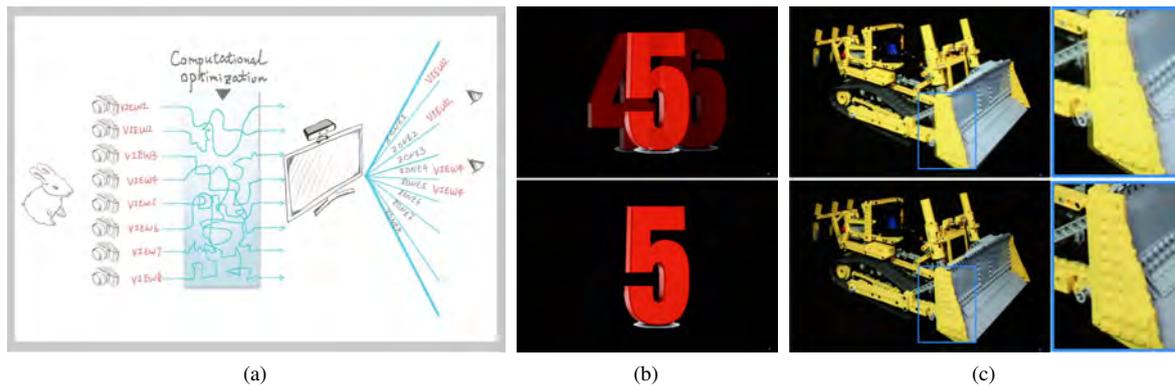


Figure 1: (a) We introduce a software computational stage for dynamic remapping of views before they are used to drive multiview autostereoscopic displays. This remapping replaces static one-to-one mapping of views to hardware display zones, and allows minimization of error due to crosstalk at multiple user eye locations considering the measured optical characteristics of the display. (b,c) Photographs comparing one-to-one mapping above to dynamic mapping below. Notice that in both the numbered test pattern, and the bulldozer scene, substantial crosstalk is visible among neighboring views in the original display and is substantially reduced using our method.

Abstract

Multiview autostereoscopic displays have several image artifacts which prevent widespread adoption. Crosstalk between adjacent views is often severe, stereo inversion occurs at some head positions, and legacy 2-view content is difficult to display correctly. We introduce a method for driving multiview displays, dynamically assigning views to hardware display zones, based on potentially multiple observer's current head positions. Rather than using a static one-to-one mapping of views to zones, the mapping is updated in real time, with some views replicated on multiple zones, and some zones left blank. Quantitative and visual evaluation demonstrates that this method substantially reduces crosstalk.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Three-Dimensional Graphics and Realism—Display Algorithms

1. Introduction

Stereoscopic display technology provides a 3D viewing experience, giving a closer reproduction of the physical world. 3D displays have had important impacts in scientific visualization, engineering, and entertainment. Glasses-free 'Autostereoscopic' 3D displays produce 3D scenes without requiring viewers to wear stereo glasses. The display presents different views to observers in a discrete set of spatial zones.

However, user experience in commercially available autostereoscopic displays suffers from three main issues. First, the display may exhibit significant crosstalk, where an observer in a given zone will also see contributions from the views intended for neighboring zones. Crosstalk is a widely recognized problem in multiview autostereoscopic research and can cause visual discomfort for viewers watching such a display [KT04]. Second, the observer experiences stere-

o inversion in positions which place the left and right eyes in neighboring repetitions of the display zones. This forces the user to adjust their head position until the proper 3D effect is apparent. Third, multiview displays are not directly compatible with the vast majority of existing stereo 3D content, which is recorded and stored in a 2-view format. The most common solution is to simply map the two available views onto alternating display zones, but this exacerbates the crosstalk and inversion issues above.

In this paper we propose software dynamic mapping of views to physical display zones. Rather than using a predetermined and static one-to-one mapping of views to zones, views are dynamically allocated to zones based on current eye positions. By replicating some views in multiple zones, and leaving some zones blank, crosstalk can be greatly reduced. In addition, dynamic allocation of views to zones allows stereo inversion artifacts to be eliminated and 2-view content to be displayed comfortably to viewers without glasses.

We have implemented a prototype system using a commercially available 8-zone lenticular display. We measure and compare crosstalk between the original display and when using dynamic mapping. We find crosstalk to be substantially reduced using our method. The cost of these improvements is a requirement for head position tracking, which we implement with a 3D camera and standard computer vision, for multiple viewers.

The primary contribution of this paper is a software-based approach to improve image quality on multiview autostereoscopic displays. We support this contribution with a prototype, and both quantitative and visual evaluation.

2. Related Work

Multiview display hardware has been well studied and designs exist using lenticular screens [JJ05], parallax barriers [TTL*09], and multiple projectors [MP04]. Implementations might have as many as 60 display zones [BKM07]. Good surveys exist [SS99,UCES11,HDFP11], and advanced hardware technologies continue to be invented [WLHR12, TLEB13,MWH*13]. Our work focuses on using software to ameliorate some of the deficiencies common to many of the most common hardware platforms.

Eye tracking has been used to improve the functioning of hardware, often mechanically moving optical elements to steer display zones to the current eye position. In ATTEST, lenticular screens with two views are mechanically adjusted according to the viewer's position [RdBF*02]. The projects MUTED and HELIUM3D replace a conventional backlight with novel steering optics and a horizontally scanned light valve so that zones can be directed to appropriate viewers' eyes [BSS*10]. Liou et al. use a synchro-signal LED scanning backlight to reduce crosstalk [LLH11]. Stolle et al. also introduce an electronically steerable backlight [SOB*08].

These systems all use head tracking to directly manipulate hardware components, dynamically redirecting zones to a moving viewer. Our use of eye tracking differs fundamentally from these techniques. Our method uses eye tracking to dynamically map views to the static zones supported by the display hardware. We avoid modifying the physical device, and are largely agnostic to the specific technology employed.

Parallax barrier technology normally uses a static printed barrier, but it is possible to replace this with a dynamic barrier updated in response to user eye position. Both single user [PPK00] and two user [PKG*07] systems have been demonstrated. Randomized barrier arrays were used in the Random Hole Display [NF09], and methods that optimize image quality for multiple viewers were developed for this display [YSF10]. These techniques are specific to this technology since they focus on updating the barrier pattern.

Most similar to our work, Boev et al. proposes a method of optimizing quality and brightness for a single user [BGGE08]. Earlier, Woodgate et al. had developed the PIXCON LCD panel, focusing primarily on novel hardware but proposing electronic tracking [WEH*98]. Both of these works suggest that content swapping based on a viewer's eye position is possible, but both focus on a single user, and neither demonstrates a functioning system. Our work introduces an algorithm suitable for multiple users, a functioning prototype, and quantitative measurements of results.

Heide et al. develop optimization methods for driving compressive displays that sample lightfields [HWRH13]. Their work focuses on a custom hardware design with considerably more flexibility than the commercially available lenticular screens that are the focus of our work. However, our work is related to theirs in that it views the output of the display as a function of a controllable set of inputs, and seeks an optimum solution for that input.

A number of authors have addressed image quality by modifying the content prior to display. Zwicker et al. provide a frequency analysis and prefilter for anti-aliasing [ZMD-P06]. Masia et al. propose a light field retargeting method using optimization [MWA*13], and Didyk et al. combine phase based video magnification and antialiasing into a single filtering process [WRDF13, DSAF*13]. These methods often produce a set of greatly improved static views which could be used in conjunction with this work.

3. Our Method

Autostereoscopic screens: Multiview autostereoscopic displays can be built based on a wide variety of optical principles, however all share a common viewing geometry. The hardware supports a finite number of display zones, each of which directs the screen image in a different angular direction. The screen is driven with a set of imagery rendered or photographed from slightly different views. Under normal operation views are mapped to zones in a static one-to-one

manner. The observer’s two eyes are in different zones, and thus receive different views, resulting in stereo perception. Figure 1a gives an abstract diagram of the screen, augmented to show dynamic mapping as presented in this work.

Unfortunately the hardware usually supports a limited number of zones, and does not have sharp transitions between them. For example, we use a display that has 8 zones, and relatively broad transitions. This results in substantial crosstalk between views, as shown in Figure 2. The average intensity of each view, as seen from a range of horizontal positions is plotted. For example, when a user’s eye is placed at position 150mm, this is intended to be zone 5. However in addition to view 5, they will see each of view 4 and view 6 at 40% intensity, resulting in visible double images.

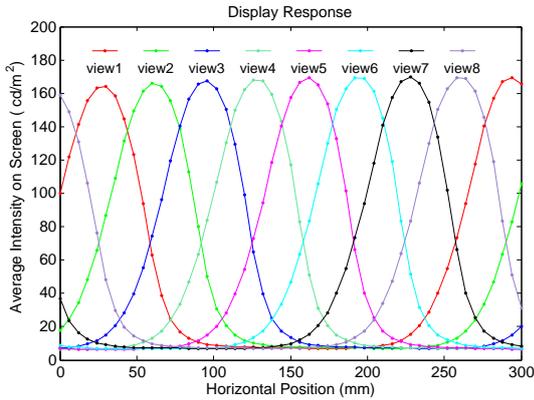


Figure 2: Multiview displays often have severe crosstalk. The average visible intensity from each zone was measured at many positions horizontally. Note that all positions have significant contribution from at least two display zones.

An additional complication is that not only do adjacent view zones of multiview displays overlap producing crosstalk, driving the zones themselves is neither independent nor linear. For example, the curves in Figure 2 are produced when driving view zones one at a time while measuring screen brightness. However, it is not true that driving Zone1 and Zone2 simultaneously would result in a final intensity which is a linear combination of the values measured for each respective zone. In general, it is necessary to know the optical transfer function of the specific display, whether through calibration or theory.

Dynamic mapping: Crosstalk can be substantially reduced by using eye position as an input and dynamically mapping views to physical display zones. Central to our approach is avoiding crosstalk by ensuring that adjacent view zones that contribute to a single user’s eye are driven by weighted copies of an identical image. This simple approach allows the bleedthrough from adjacent physical display zones to still occur, but it does not lead to image degradation. Our method is illustrated in Figure 3. In this

simple example with a single viewer, we drive both Zone2 and Zone3 with weighted copies of View3. Similarly, it is not necessary to display a view in every zone, and the zone between two eyes can be left blank in order to suppress crosstalk.

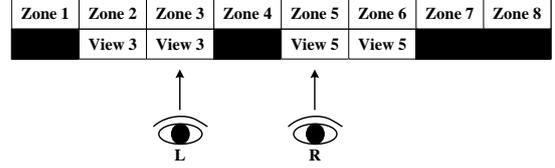


Figure 3: Our method of dynamic mapping decouples desired views from physical zones. Rendering the same view in several zones and leaving blank zones between eyes can significantly reduce crosstalk. The mapping from views to zones is updated in realtime based on eye position.

Multiple viewers may be positioned such that a more complex optimization is needed for determining what image is best displayed in each zone. However the core of our method is the simple observation that our goal is to optimize image quality at the eye positions, not to optimize quality everywhere, and driving the display with something other than a one-to-one mapping of views to zones often results in the minimum error.

Minimizing Error: The error we wish to minimize, and thus our strategy for dynamic mapping must account for two primary goals. First we should minimize crosstalk error. Second, we should account for content sources which have a different number of views than the display.

Whether the source images, $\mathbf{V}(\mathbf{x})$, are virtual or captured from cameras, a change in viewer position, \mathbf{x} , induces a parallax change in the image viewed. Since the display device has discrete zones, the *views* are often discretized and denoted as $\mathbf{V} = [V_1, V_2, V_3, \dots, V_K]$. The hardware device displays different images angularly, in each of a set of N zones. The images which are input to the device are denoted $\mathbf{Z} = [Z_1, Z_2, Z_3, \dots, Z_N]$. Current devices assume that the number of available views, K , is equal to the number of display zones, N . On existing displays views are mapped to zones in a one-to-one fashion, such that $\mathbf{Z} = \mathbf{V}$.

The display device frequently does not closely match the idealized display of discrete zones with sharp boundaries. Neighboring zones exhibit crosstalk, and each display zone influences a wide region of space with maximum intensity in the center of its intended range, and less intensity as the eye position moves into neighboring zones.

The eye image, \mathbf{E} , actually observed is a function of both the multiple zone images displayed on the device, \mathbf{Z} , and the eye position, \mathbf{x} . That is, at each position of user space, $\mathbf{E}(\mathbf{x}) = \mathcal{E}(\mathbf{Z}, \mathbf{x})$.

Traditionally there is a direct spatial relationship between

views, $\mathbf{V}(\mathbf{x})$, and the desired images at the users eye position, $\mathbf{D}(\mathbf{x})$. However a more complex *view selection policy* is possible. We might, for example, want to ignore spatial relationship to the scene and simply specify that the desired image at the left eye position, $\mathbf{D}(\mathbf{x}_L)$, should always be the left eye view, V_L , in a legacy 2-view movie. Similarly, we might want to detect possible multi-user conflicts in 8-view content and shift the desired views for individual users, even though that would change their virtual viewing direction slightly. In general the view selection policy depends on the application and user preferences for degradation when the device can not provide ideal images. We discuss one possible policy for 2-view content in Section 5.2.

Since the display device has crosstalk and other deficiencies we do not simply drive the closest hardware zones with the desired images, $[\mathbf{D}(\mathbf{x}_1), \mathbf{D}(\mathbf{x}_2), \dots]$. In general, some other set of images will produce actual observed images which are closer to those that are desired. We refer to this selection of what to drive the device with as the *display algorithm*. Perceptually based metrics for evaluating the difference between two images exist [WBSS04]. In this work, we obtain acceptable results by simply minimizing the squared image intensity error between what is desired, and what the device actually produces.

$$\operatorname{argmin}_{\mathbf{Z}} \sum_{m=1}^M [\mathcal{E}(\mathbf{Z}, \mathbf{x}_m) - \mathbf{D}(\mathbf{x}_m)]^2 \quad (1)$$

4. Implementation Details

Experimental Setup: Our experiments were conducted on a lenticular-based multiview autostereoscopic 3DTV made by Alioscopy. The display uses a slanted lens array as an optical filter, affixed to a normal LCD screen to distribute separate images in each direction. The hardware has 8 horizontal display zones defined. The manufacturer has calibrated the device for optimum performance at 140cm away from the display, and we place our viewing couch at approximately this distance. With this positioning, the display has zones spaced approximately 32mm apart. The average interocular distance in adults is 63mm, corresponding to a two zone separation between eyes on this display [Dod04]. However the statistical range of ocular separation is 50-75mm, and some viewers will thus occasionally have eyes in neighboring zones.

The setup of our prototype can be seen in Figure 4. Eye positions are tracked by either a pair of webcams or a Kinect 3D camera on top of the display and the information is used for dynamic mapping of views in realtime. Eye tracking is well studied, with surveys available [HJ10]. Our eye tracker with webcams is implemented with a decision cascade of Haar basis functions [VJ01]. We observe the Kinect based tracker to be more robust with real users but either provides acceptable performance.

In order to record experimental results, this setup is augmented with a printed image of a face and a pair of cameras



Figure 4: Our prototype consists of an 8-zone display with an eye tracker mounted below it. Multiple viewers sit on a couch and a printed face with cameras mounted behind the eyes is used to record results as they would have been seen by a viewer.

placed behind the eyes to capture the view as it would be seen by a live observer. When direct comparisons of quality are needed from repeatable positions, the printed face is mounted on a motor control rail. The known location of the rail is used to insure proper comparisons and repeatability between algorithms, but ignored by the display algorithms.

Calibration: Autostereoscopic display devices typically have optical arrangements more complicated than standard 2D displays. The crosstalk between zones, as well as any spatial or radiometric nonlinearities must be taken into account. These issues are encapsulated in the display transfer function, $\mathcal{E}(\mathbf{Z}, \mathbf{x})$, which describes what the eye actually sees at position, \mathbf{x} , when the display zones are driven with a set of images, \mathbf{Z} .

We sample the space of both \mathbf{Z} and \mathbf{x} , recording the actual observed intensities at all screen pixels using a camera. We store the samples and evaluate $\mathcal{E}(\mathbf{Z}, \mathbf{x})$ by linear interpolation. We have found this simple method empirically sufficient. We have sampled \mathbf{x} as densely as 6mm and as coarsely as 32mm, and \mathbf{Z} with as few as 9 samples and as many as 729 samples in a particular location. In all cases our method offers improvements over the static display method.

Display Algorithm: Global optimization over all possible zone images would be prohibitively slow. Our display requires eight input images each with millions of independent pixels. Fortunately we don't need to find an optimum solution, only one that produces low errors, so we use heuristics to prune the search space, and optimize in the more constrained space.

First we assume that the only reasonable value for each input zone image Z_1, \dots, Z_N is a weighted linear combination of the the desired images at all of the eye locations $\mathbf{D}(\mathbf{x}_1), \dots, \mathbf{D}(\mathbf{x}_M)$. This corresponds to assuming we should drive the display with things we want to see, not some other image entirely.

$$Z_i = w_{i1} \cdot \mathbf{D}(\mathbf{x}_1) + w_{i2} \cdot \mathbf{D}(\mathbf{x}_2) + \dots + w_{iM} \cdot \mathbf{D}(\mathbf{x}_M) \quad (2)$$

Since there are N display zones and M eye positions, this assumption defines a weight matrix, \mathbf{w} , with $N * M$ terms which encodes the possible solution space.

To further reduce the dimensionality of the search space, for each zone, Z_i , we check to see whether it contributes less than 15% of the full energy to each eye position \mathbf{x}_j , that is $\max(\mathbf{E}_i(\mathbf{x}_j)) < 15\%$. If the contribution is small then we set $w_{ij} = 0$ in \mathbf{w} .

We now substitute our restricted definition of \mathbf{Z} from Equation 2 into the general minimization defined in Equation 1, and minimize over the elements of matrix \mathbf{w} which were not set to 0.

$$\operatorname{argmin}_{\mathbf{w}} \sum_{m=1}^M [\mathcal{E}(\mathbf{Z}, \mathbf{x}_m) - \mathbf{D}(\mathbf{x}_m)]^2 \quad (3)$$

We use a very simple gradient descent solver to find the solution. To increase efficiency we subsample all images to 80x40, and in practice this minimization converges fast enough for realtime operation with two viewers. Since this is a research prototype we did not attempt to optimize the code and thus three or more users causes the display to have a slight lag when head positions change.

An important insight is that our *display algorithm* is rearranging the limited hardware resource rather than directly improving the hardware quality of the display. Any optical imperfections which lead to crosstalk remain, but they are guided by our *display algorithm* so that visual image quality at the eye positions are optimized. Figure 5 compares the error distribution in the traditional one-to-one mapping to that using our dynamic mapping display algorithm. Notice that the traditional mapping has roughly equally distributed error in all viewing locations. Our method redistributes the error so that image quality at eye positions is improved. This comes at the cost of increasing error in regions without viewers.

5. Results

5.1. Crosstalk Reduction

One of the primary issues affecting adoption of autostereoscopic displays is crosstalk. We evaluated our method both visually and quantitatively.

Visual comparisons were made using both imagery intended for multiview devices like ours as well as with test patterns which place a unique numeric image in the corresponding zone. Figure 6 uses photographs to compare the device using static one-to-one mapping with the device using dynamic mapping. Notice that one-to-one mapping produces an image with noticeable crosstalk, while dynamic mapping provides a clean image with little ghosting from other views. Figure 1(b) provides an additional test pattern, and Figure 1(c) an additional test scene.

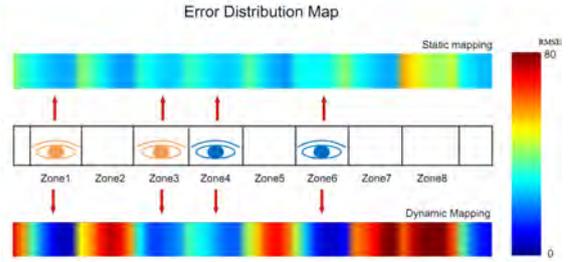


Figure 5: The error distribution of static one-to-one mapping is approximately equal in all possible viewer locations. In contrast, dynamic mapping reduces error at eye positions, redistributing it to locations without a viewer. RMSE of pixel intensities between target and observed images is used. In this visualization dark blue is low error and dark red is high error. Notice that dynamic mapping has substantially lower error at eye positions.

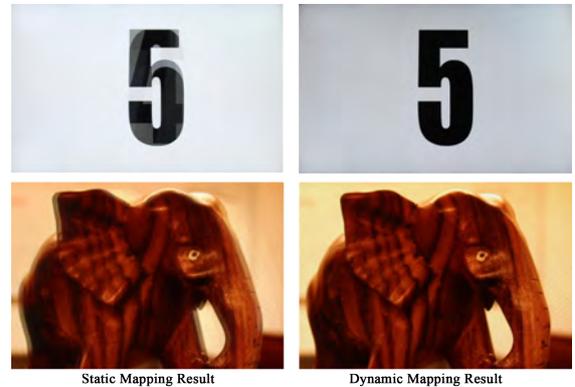


Figure 6: Visual results using static and dynamic mapping are shown. In the numeric test pattern note that static one-to-one mapping has noticeable crosstalk, with multiple numbered views visible. The overall gain of the white background also appears different because of this crosstalk. The elephant example also shows crosstalk in this close up view.

As the number of viewers increases, dynamic mapping will find it harder to place desired images into zones in a way that preserves quality everywhere. Given enough viewers the mapping will revert to one-to-one mapping since this is a good solution for optimizing many eye locations.

In order to analyze the performance of our method as the number of viewers increases, we randomly sampled possible head locations for a set of M viewers, and computed the average RMSE error across the eyes in each configuration. We repeated the experiment 100 times for each M , and plot the median error in Figure 7. The shaded bands show the range between the 25th and 75th percentile configurations. Notice that for a small number of viewers dynamic mapping clearly outperforms one-to-one mapping. As the number of

viewers grows, the performance reverts back to the quality of existing displays. However, the median performance for randomly positioned viewers stays below what is achievable with static mapping even when the number of eyes exceeds the number of display zones. This is because many possible head configurations do not interfere and dynamic mapping is able to find a good solution.

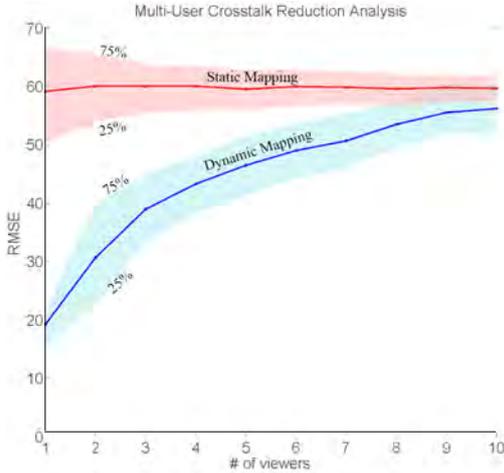


Figure 7: Comparison of median error for random head positions between static one-to-one mapping and dynamic mapping as the number of viewers increases. Dynamic mapping has much lower error for a small number of viewers and reverts back to one-to-one mapping as the number of viewers grows.

Dynamic mapping has also been evaluated against static one-to-one mapping when deviating from the ideal distance, shown in Figure 8. Crosstalk is measured as the ratio of observed intensity from unintended zones to the intensity from the intended zone. Note that the original display has a high crosstalk level of 80% even at the ideal viewing distance, where each neighboring zone is 40% as bright as the intended zone. This rises to a crosstalk of 180% as the viewer moves away from this ideal distance. In this evaluation using a single viewer dynamic mapping has a much lower crosstalk of below 10% at the ideal distance, increasing to 60% as the viewer moves. As the number of viewers increases, dynamic mapping will degrade, eventually returning to a solution equivalent to static mapping.

Since the original static display method degrades rapidly away from its ideal viewing distance, all direct comparisons in this paper were performed at the ideal distance to provide a fair comparison.

To provide more insight into how views are mapped into zones, Figure 9(a) shows an example with two viewers. In this case the viewers share an eye position so the solver is able to find a weight matrix \mathbf{w} in which some zones are left

black in order to suppress crosstalk. Close up photographs showing the result are on the right. Note that crosstalk is eliminated at all eye positions. Figure 9(b) is a more challenging case because eye position B is no longer identical with eye position C, but it is close enough to conflict. In this case no crosstalk free solution is available and the display algorithm produces a weight matrix which contains substantial mixing for the zones which impact B and C. This mixing has lower RMSE than static one-to-one mapping, but it is still far from ideal. In the photographs at right we can see that A and D, which are well separated from other eyes, are substantially improved as expected. Eye position C has somewhat less crosstalk than the static mapping, while position B has crosstalk at approximately the same level.

The example of eye positions in Figure 9(b) is a case when a more sophisticated view selection policy would help. We use a policy that sets the desired view for all eyes, including B and C, to the views corresponding to their virtual location, $\mathbf{D}(\mathbf{x}_B) = V_5$ and $\mathbf{D}(\mathbf{x}_C) = V_6$. This leads to crosstalk since Zone5 and Zone6 are not well separated and need to produce different views. It is possible to define a view selection policy which would note the conflict and shift the desired virtual view of the second viewer to the left such that $\mathbf{D}(\mathbf{x}_C) = V_5$ and $\mathbf{D}(\mathbf{x}_D) = V_7$. This would resolve the conflict and the display algorithm would be able to find a weight matrix that provides crosstalk free views to all eyes, at the cost of positional correctness for the second viewer. Since there is a tradeoff among degradations available, the correct view selection policy is application dependent.

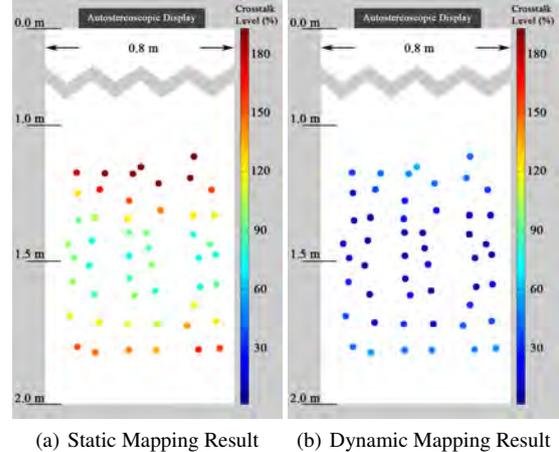


Figure 8: Perceived visual image quality as measured by the level of crosstalk at each position in 3D space in front of the autostereoscopic display. (a) Even at the ideal viewing distance, one-to-one mapping has severe crosstalk in perceived images. (b) Dynamic mapping has reduced crosstalk everywhere in the space. In this visualization dark blue is low crosstalk and red is high crosstalk.

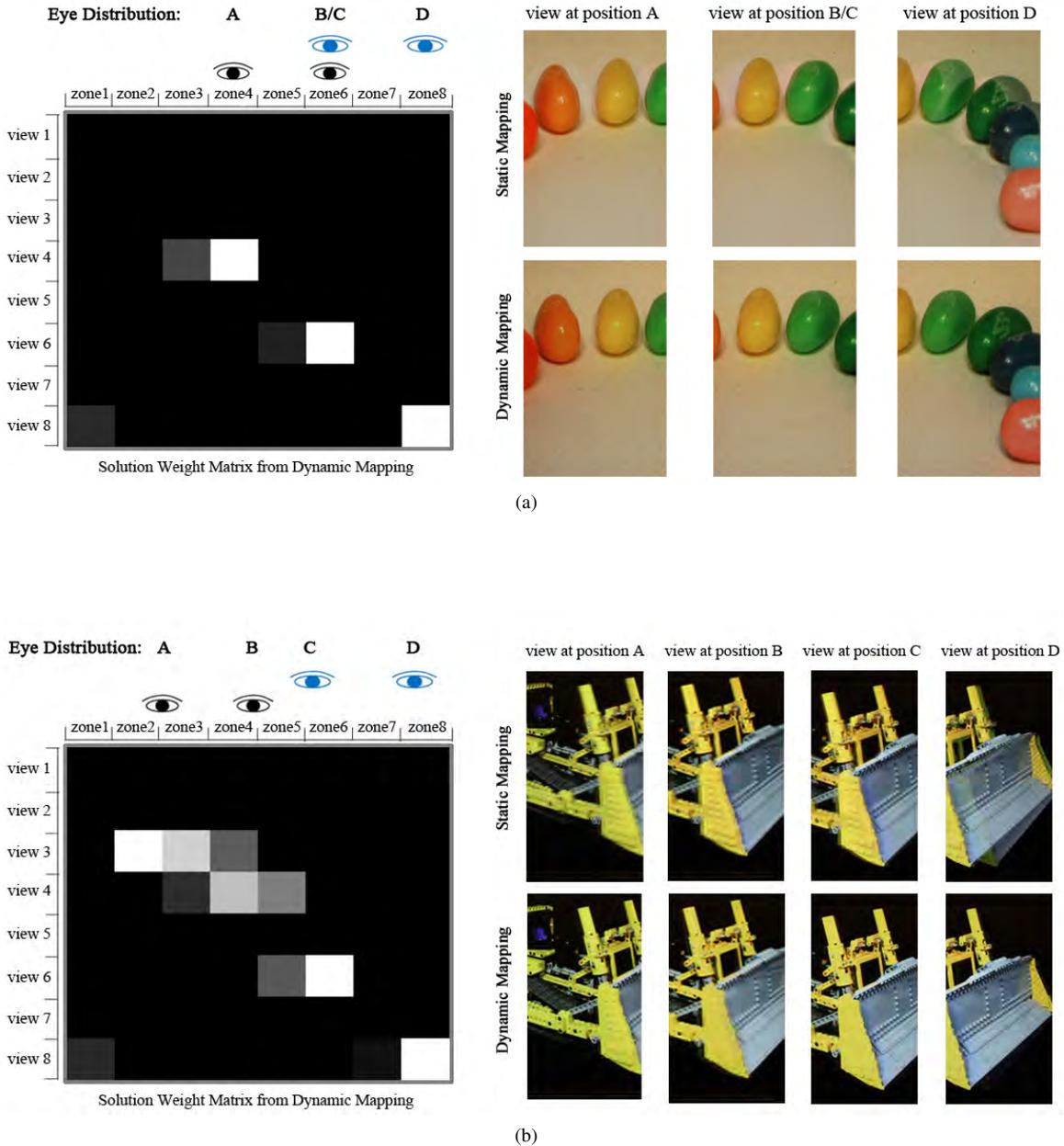


Figure 9: Examples of dynamic mapping for two users. Eye locations and the weight matrix w provided by our solver are shown on the left, with comparison photographs of the screen on the right. (a) In this example the viewers have overlapping but compatible eye locations, so the solver finds a solution with no crosstalk. In the close up cropped images at right notice that all three eye positions are improved. (b) In this example one of the viewers has moved to the left such that B and C are now in conflicting positions. The solver provides a nearly crosstalk free solution for positions A and D, but no crosstalk free solution is possible for B and C. The images at right reveal that while A, C, D are all improved, position B has crosstalk roughly comparable to static mapping.

5.2. Legacy 2-view content

Most existing content for 3D stereo viewing was captured and stored with only 2-views. Autostereoscopic displays require multiview input. Existing displays often resort to using a fixed pattern for mapping views to zones. This strategy produces some viewing locations which create a correct stereo percept and some locations with stereo inversion.

Another approach is to employ view synthesis methods to produce enough new viewpoints to drive the display [SMD*08, FWL*11, DSAF*13]. These methods are compatible with our work since they produce a set of input views. In addition, with dynamic mapping the number of synthesized views need not match the number of device zones, and if eye tracking is available, computation can be reduced by synthesizing only the required viewpoints.

Dynamic mapping also allows 2 channel input to be used directly, without synthesizing additional views, and we concentrate on this case in our discussion below. The desired view at the left eye, wherever it happens to be, is set to the left view provided by the content, $\mathbf{D}(\mathbf{x}_L) = V_L$. This is repeated for the right eye. This allows legacy stereo content to be played on autostereoscopic displays without additional processing steps, which may distort the artistic creator’s intention.

The advantage of dynamic mapping in a simple single user case is shown in Figure 10. The best static mapping for 2-view content on our particular display is $\mathbf{Z} = [V_L V_L V_R V_R V_L V_L V_R V_R]$, and thus is used in this example. The observed views for a user in different positions in front of the display are shown allowing the methods to be compared. Notice that with a static mapping the user sometimes sees a correct crosstalk free display, sometimes sees crosstalk, and sometimes sees stereo inversion with the left and right images swapped. This requires users to find the correct location for viewing and then keeping their head still to avoid ruining the effect. In contrast, dynamic mapping allows the user to position themselves in any location and move as needed, since the correct crosstalk free left and right images are always available.

As the number of viewers increases there will eventually be conflicts in which some zones are expected to display both V_L and V_R . These conflicts are unrelated to the displays transfer function, and will not be resolved adequately by the display algorithm’s minimization. Consider a perfect multiview display device with no crosstalk. When a single position \mathbf{x}_j is occupied by the left eye of one viewer and the right eye of another we have $\mathbf{D}(\mathbf{x}_j) = V_L$ and $\mathbf{D}(\mathbf{x}_j) = V_R$, thus the least RMSE is achieved by providing a combination of the two images even though this is not visually pleasing. The failure here is in the *view selection policy*, not the display algorithm. A single position should not be expected to display both left and right images.

When conflict is unavoidable the correct view selection

policy will depend on the preference of the display designer, and multiple policy options are available. We prefer a policy that favors 2D viewing to stereo inversion, since we feel that 2D is less objectionable. Figure 11 shows an example with two viewers. There is no mapping of views to zones which will provide 3D to both viewers. However Zone3 and Zone5 can be set to allow viewer1 to see 3D, while Zone7 is set to provide a 2D display to viewer2, rather than stereo inversion.

One possible view selection policy for 2-view content that supports our preference is provided here. Given an arbitrary configuration of multiple viewers we enumerate all possible assignments of V_L and V_R to $Z_1 \cdots Z_N$. We disallow configurations which result in stereo inversion to any viewer. Among the remaining possible assignments, we select the one that results in the maximum number of viewers obtaining a 3D view. The remaining viewers will see 2D. Note that it is always possible to completely avoid stereo inversion by setting all zones to V_L , at the cost of lost 3D for all viewers. However in most cases a better option is available.

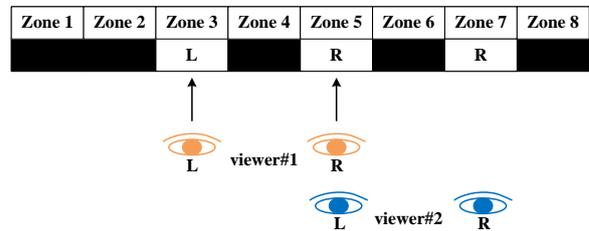


Figure 11: When dynamically mapping legacy 2-view content, the view selection policy must determine how to handle conflicts. In this case no solution exists to provide 3D stereo views to both viewers. We implement a policy that provides viewer#2 with a 2D view, instead of the more objectionable alternative, stereo inversion.

We analyze this policy in Figure 12. To make clear that the issues involved are unrelated to crosstalk, this analysis is provided for an ideal multiview display. We generate 100 random viewer configurations for different numbers of viewers and select the best zone assignment in each case. The plot reports the percentage of total viewers who see 3D. A static mapping policy results in 50% of viewers seeing 3D and 50% seeing stereo inversion. In comparison, our policy results in a larger percentage of viewers seeing 3D. In addition, the degraded viewers see 2D instead of uncomfortable stereo inversion.

6. Limitations

The dynamic mapping method depends on the robustness of the eye tracker. Fortunately, eye tracking for multiple viewers has matured to be fairly robust, but we still see failures occasionally. We could add detection of such failures and apply conservative mappings such as the conventional static

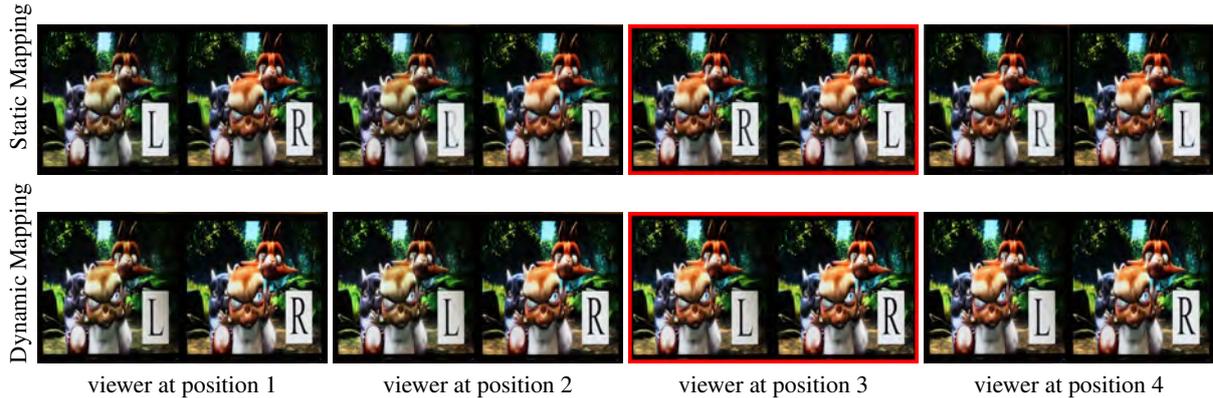


Figure 10: A static LLRLLRR mapping of legacy 2-view content onto our 8-view display results in stereo inversion for some eye positions, with left and right images swapped. Dynamic mapping provides the correct view to each eye regardless of viewer position.

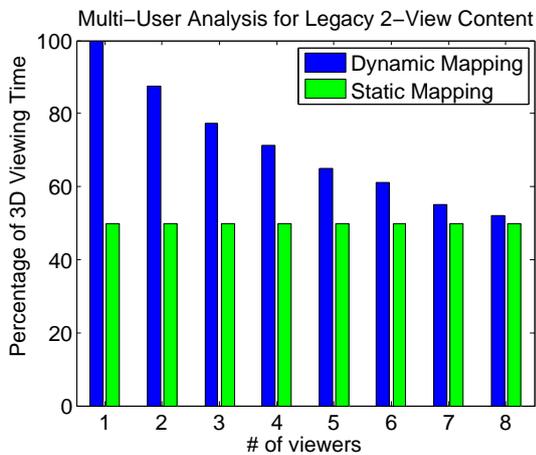


Figure 12: Analysis of percentage of 3D viewing with legacy 2-view content, an ideal crosstalk free display, and multiple viewers. Static mapping results in 50% of positions which provide 3D viewing, and 50% which provide stereo inversion. For up to eight viewers, dynamic mapping results in a greater percentage of viewers in random configurations seeing 3D. In addition dynamic mapping avoids stereo inversion for those with degraded views, providing 2D viewing instead.

assignment. Importantly, our method is robust to minor jitter in tracked position since the primary aim is to select zones at a relatively coarse scale.

We have implemented our method in real time for only two viewers, and provided offline analysis for more viewers. Since our code is not fully optimized, but performance is adequate, we do not believe that computation is a fundamental limitation.

Although we believe our method is suitable for a range of hardware devices, the display must have sufficient zones to allow meaningful dynamic mapping. A 2-view autostereoscopic display would certainly be insufficient. Fortunately the trend among designers of devices appears to be towards greater numbers of viewing zones.

Our implementation uses a simple solver for the display algorithm and investigates view selection policies only for legacy 2-view content. Future work could include more sophisticated algorithms for both. For example, the display algorithm could use a spatially varying minimization, rather than the uniform weights now applied to entire zones. In addition, when multiview content is available, we currently set the desired view for each eye to the one which corresponds to its virtual location. However this insures crosstalk when eyes from two viewers are in immediately adjacent zones. A view selection policy could adjust the *desired* view for eyes in neighboring zones to avoid this crosstalk in many cases.

7. Conclusion

In this paper, we propose a simple software-based method to improve the imaging quality of multiview autostereoscopic displays. We have evaluated the method both visually and quantitatively, and found it effective at reducing crosstalk. Since the method dynamically maps views with respect to device specific hardware zones, it can also be used to address other common difficulties such as displaying legacy 2-view content on multiview displays and eliminating stereo inversion.

References

- [BGGE08] BOEV A., GEORGIEV M., GOTCHEV A., EGIAZARIAN K.: Optimized single-viewer mode of multiview autostereoscopic display. In *Proc. of EUSIPCO* (2008). 2

- [BKM07] BALOGH T., KOVÁCS P., MEGYESI Z.: Holovizio 3D display system. In *Proceedings of the First International Conference on Immersive Telecommunications* (2007), ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), p. 19. 2
- [BSS*10] BRAR R., SURMAN P., SEXTON I., BATES R., LEE W., HOPF K., NEUMANN F., DAY S., WILLMAN E.: Laser-based head-tracked 3D display research. *Display Technology, Journal of* 6, 10 (2010), 531–543. 2
- [Dod04] DODGSON N. A.: Variation and extrema of human interpupillary distance. In *Proceedings of SPIE* (2004), vol. 5291, pp. 36–46. 4
- [DSAF*13] DIDYK P., SITTHI-AMORN P., FREEMAN W., DURAND F., MATUSIK W.: Joint view expansion and filtering for automultiscopic 3D displays. *ACM Transactions on Graphics (TOG)* 32, 6 (2013), 221. 2, 8
- [FWL*11] FARRE M., WANG O., LANG M., STEFANOSKI N., HORNING A., SMOLIC A.: Automatic content creation for multiview autostereoscopic displays using image domain warping. In *Multimedia and Expo (ICME), 2011 IEEE International Conference on* (2011), IEEE, pp. 1–6. 8
- [HDFP11] HOLLIMAN N. S., DODGSON N. A., FAVALORA G. E., POCKETT L.: Three-dimensional displays: a review and applications analysis. *Broadcasting, IEEE Transactions on* 57, 2 (2011), 362–371. 2
- [HJ10] HANSEN D. W., JI Q.: In the eye of the beholder: A survey of models for eyes and gaze. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 32, 3 (2010), 478–500. 4
- [HWRH13] HEIDE F., WETZSTEIN G., RASKAR R., HEIDRICH W.: Adaptive image synthesis for compressive displays. *ACM Transactions on Graphics (TOG)* 32, 4 (2013), 132. 2
- [JJ05] JOHNSON R., JACOBSEN G.: Advances in lenticular lens arrays for visual display. In *Proceedings of SPIE* (2005), vol. 5874, p. 587406. 2
- [KT04] KOOI F., TOET A.: Visual comfort of binocular and 3D displays. *Displays* 25, 2-3 (2004), 99–108. URL: <http://linkinghub.elsevier.com/retrieve/pii/S0141938204000502>. 1
- [LLH11] LIU J., LEE K., HUANG J.: Low crosstalk multi-view tracking 3-D display of synchro-signal LED scanning backlight system. *Display Technology, Journal of* 7, 8 (2011), 411–419. 2
- [MP04] MATUSIK W., PFISTER H.: 3D TV: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. In *ACM Transactions on Graphics (TOG)* (2004), vol. 23, ACM, pp. 814–824. 2
- [MWA*13] MASIA B., WETZSTEIN G., ALIAGA C., RASKAR R., GUTIERREZ D.: Display adaptive 3D content remapping. *Computers & Graphics* 37, 8 (2013), 983–996. 2
- [MWH*13] MAIMONE A., WETZSTEIN G., HIRSCH M., LANMAN D., RASKAR R., FUCHS H.: Focus 3D: Compressive accommodation display. *ACM Transactions on Graphics (TOG)* 32, 5 (2013), 153. 2
- [NF09] NASHEL A., FUCHS H.: Random hole display: A non-uniform barrier autostereoscopic display. In *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2009* (2009), IEEE, pp. 1–4. 2
- [PKG*07] PETERKA T., KOOIMA R. L., GIRADO J. I., GE J., S D. J., JOHNSON A., LEIGH J., SCHULZE J., DEFANTI T. A.: Dynallax: Solid state dynamic parallax barrier autostereoscopic VR display. In *In Proceedings of IEEE Virtual Reality 2007* (2007), IEEE, pp. 155–162. 2
- [PPK00] PERLIN K., PAXIA S., KOLLIN J. S.: An autostereoscopic display. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (2000), ACM Press/Addison-Wesley Publishing Co., pp. 319–326. 2
- [RdBF*02] REDERT A., DE BEECK M., FEHN C., IJSSELSTEIJN W., POLLEFEYS M., VAN GOOL L., OFEK E., SEXTON I., SURMAN P.: Advanced three-dimensional television system technologies. In *3D Data Processing Visualization and Transmission, 2002. Proceedings. First International Symposium on* (2002), IEEE, pp. 313–319. 2
- [SMD*08] SMOLIC A., MULLER K., DIX K., MERKLE P., KAUFF P., WIEGAND T.: Intermediate view interpolation based on multiview video plus depth for advanced 3D video systems. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on* (2008), IEEE, pp. 2448–2451. 8
- [SOB*08] STOLLE H., OLAYA J.-C., BUSCHBECK S., SAHM H., SCHWERDTNER A.: Technical solutions for a full-resolution autostereoscopic 2D/3D display technology. In *Proceedings of SPIE* (2008), International Society for Optics and Photonics, pp. 1–12. 2
- [SS99] SEXTON I., SURMAN P.: Stereoscopic and autostereoscopic display systems. *Signal Processing Magazine, IEEE* 16, 3 (1999), 85–99. 2
- [TLEB13] TRAVIS A., LARGE T., EMERTON N., BATHICHE S.: Wedge optics in flat panel displays. *Proceedings of the IEEE* 101, 1 (2013), 45–60. doi:10.1109/JPROC.2011.2176689. 2
- [TTL*09] TSAI R., TSAI C., LEE K., WU C., LIN L., HUANG K., HSU W., WU C., LU C., YANG J., ET AL.: Challenge of 3D LCD displays. In *Proc. SPIE* (2009), vol. 7329, pp. 732903–1. 2
- [UCES11] UREY H., CHELLAPPAN K., ERDEN E., SURMAN P.: State of the art in stereoscopic and autostereoscopic displays. *Proceedings of the IEEE* 99, 4 (2011), 540–555. 2
- [VJ01] VIOLA P., JONES M.: Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (2001), vol. 1, IEEE, pp. I-511. 4
- [WBSS04] WANG Z., BOVIK A. C., SHEIKH H. R., SIMONCELLI E. P.: Image quality assessment: From error visibility to structural similarity. *Image Processing, IEEE Transactions on* 13, 4 (2004), 600–612. 4
- [WEH*98] WOODGATE G., EZRA D., HARROLD J., HOLLIMAN N., JONES G., MOSELEY R.: Autostereoscopic 3D display systems with observer tracking. *Signal Processing: Image Communication* 14, 1 (1998), 131–145. 2
- [WLHR12] WETZSTEIN G., LANMAN D., HIRSCH M., RASKAR R.: Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting. *ACM Trans. Graph. (Proc. SIGGRAPH)* 31, 4 (2012), 1–11. 2
- [WRDF13] WADHWA N., RUBINSTEIN M., DURAND F., FREEMAN W. T.: Phase-based video motion processing. *ACM Transactions on Graphics (TOG)* 32, 4 (2013), 80. 2
- [YSF10] YE G., STATE A., FUCHS H.: A practical multi-viewer tabletop autostereoscopic display. In *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on* (2010), IEEE, pp. 147–156. 2
- [ZMDP06] ZWICKER M., MATUSIK W., DURAND F., PFISTER H.: Antialiasing for automultiscopic 3D displays. In *Proceedings of the 17th Eurographics conference on Rendering Techniques* (2006), Eurographics Association, pp. 73–82. 2

Correction Page

An error was pointed out in our description of [NF09] and [YSF10]. The paragraph implies that The Random Hole Display used a dynamic barrier, but it in fact used a static barrier and then optimized the display for tracked viewer position. That work is similar to this TR in spirit, since it also optimizes an autostereoscopic display's contents based on the active viewpoints. The reader of this work is encouraged to read that work as well. The paragraph from the related work section is included below:

" Parallax barrier technology normally uses a static printed barrier, but it is possible to replace this with a dynamic barrier updated in response to user eye position. Both single user [PPK00] and two user [PKG*07] systems have been demonstrated. Randomized barrier arrays were used in the Random Hole Display [NF09], and methods that optimizes image quality for multiple viewers were developed for this display [YSF10]. These techniques are specific to this technology since they focus on updating the barrier pattern. "