Improving Visual Quality of View Transitions in Automultiscopic Displays

Song-Pei Du[1,2] Piotr Didyk[1] Frédo Durand[1] Shi-Min Hu[2] Wojciech Matusik[1]

[1]MIT CSAIL [2]TNList, Tsinghua University, Beijing

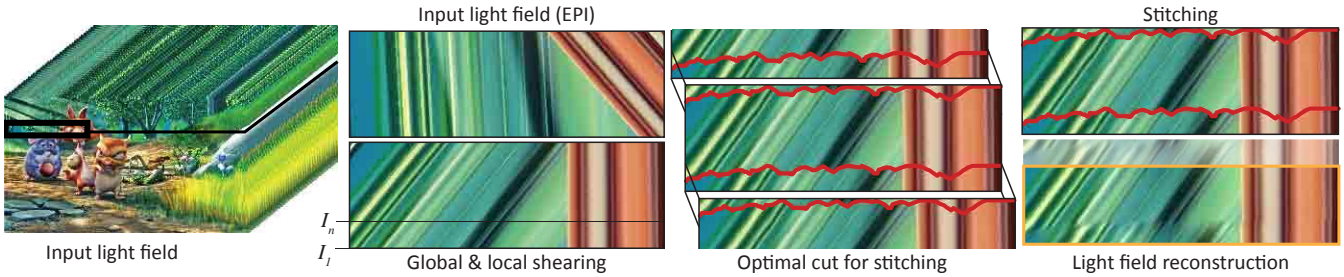

Figure 1: *To reduce artifacts caused by the limited angular view coverage of an automultiscopic display, our technique performs light field manipulation to improve its continuity. It takes as an input the original light field and performs a global shearing followed by a local shearing to align the scene around the screen plane and provide a good structure alignment between the first and the last views shown on the display, i. e., I_1 and I_n. Next, replicas of the light field are overlapped, and an optimal stitching cut is found. As the last step, the method performs the stitching along the cut in the gradient domain and reconstructs the image. The part of the light field marked in orange can be then shown on the screen.*

Abstract

Automultiscopic screens present different images depending on the viewing direction. This enables glasses-free 3D and provides motion parallax effect. However, due to the limited angular resolution of such displays, they suffer from hot-spotting, i. e., image quality is highly affected by the viewing position. In this paper, we analyze light fields produced by lenticular and parallax-barrier displays, and show that, unlike in real world, the light fields produced by such screens have a repetitive structure. This induces visual artifacts in the form of view discontinuities, depth reversals, and excessive disparities when viewing position is not optimal. Although the problem has been always considered as inherent to the technology, we demonstrate that light fields reproduced on automultiscopic displays have enough degrees of freedom to improve the visual quality. We propose a new technique that modifies light fields using global and local shears followed by stitching to improve their continuity when displayed on a screen. We show that this enhances visual quality significantly, which is demonstrated in a series of user experiments with an automultiscopic display as well as lenticular prints.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image generation—display algorithms,viewing algorithms;

Keywords: multi-view display, lighfield, view transitions

Links: DL PDF WEB

1 Introduction

Multi-view autostereoscopic (or automultiscopic) displays provide an immersive, glasses-free 3D experience, which gives them the potential to become the future of television and cinema. By showing a different image depending on the viewer's position, they reproduce both binocular and motion parallax depth cues. This is typically achieved by adding a parallax barrier [Ives 1903] or a lenticular sheet [Lippmann 1908] atop a high-resolution display, which trades some of the spatial display resolution for angular resolution. In order to provide an adequate binocular parallax, a sequence of images is shown within the primary field of view (Figure 2). Beyond it, the same sequence repeats, forming additional viewing zones and extending the effective field of view. The main advantage of such solutions is that they have the potential for providing immersive glasses-free 3D for multiple users everywhere in front of the screen. We believe that such "stereo free-viewing" is a requirement for 3D displays to succeed – think of a family watching a 3DTV at home. However, a problem arises when the left and right eyes fall into different zones (Figure 2, viewpoint B). In this situation, a depth reversal, discontinuities, and excessive disparities occur. Moreover, the reversed depth usually creates a conflict between occlusion depth cue and binocular disparity. All these effects lead to significant quality reduction for non-optimal viewing positions. We refer to this phenomenon as *transitions*.

The artifacts due to the limited extent of viewing zones in the current displays are widely recognized as a significant shortcoming. This restricts the usage of such screens in home applications and large scale visualizations. Existing solutions [Peterka et al. 2008; Yi et al. 2008; Ye et al. 2010] are based on hardware extensions, including head-tracking and dynamic parallax barriers. Although they can reduce the problem, they are suitable only for a small number of viewers (1 – 3). Furthermore, the additional hardware and the need for real-time processing, which depends on the current viewers' position, make these approaches hard to implement in commercial devices such as 3DTVs.

In this paper, we demonstrate that multi-view content offers enough degrees of freedom to improve its quality by only modifying the displayed views. To this end, we first analyze light fields produced

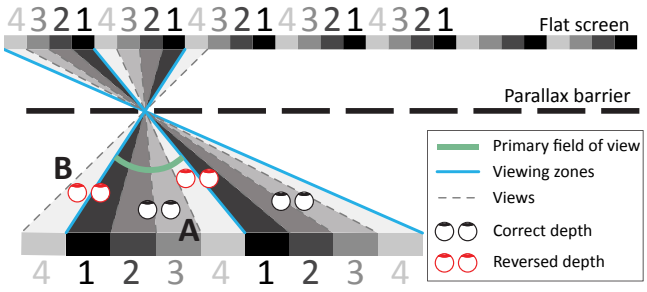

Figure 2: *Illustration of a 4-view automultiscopic display with parallax barrier. 3D viewing is correct within the marked viewing angle. The discontinuities between viewing zones (views 1 and 4) lead to visual artifacts.*

by traditional lenticular as well as parallax barrier automultiscopic displays, and give a unified and comprehensive explanation of transitions in the light field space. Based on this analysis, we propose a new method which optimizes input images to improve the perceived quality at the transitions. In contrast to previous hardware solutions, our optimization does not require knowledge about viewer's position, which makes the technique suitable for an arbitrary number of observers. It also does not require any hardware modifications and can be used as a pre-processing step. To the best of our knowledge, this is the first technique of this kind. We demonstrate the results for static images and video sequences using both parallax barriers and lenticular sheets. To further validate the quality improvement, we present a series of user experiments that analyze the advantages of the optimized content.

2 Related Work

Besides the hardware solutions mentioned in the previous section, the problem of transitions has not been addressed before. Our work is mostly related to light field processing and manipulation techniques. It also takes advantage of a wealth of techniques for seamless image and video compositing.

Light Field Processing and Manipulation A light field is a continuous function that represents radiance emitted from the scene [Levoy and Hanrahan 1996]. Due to the discrete nature of acquisition and display stages, light fields are usually aliased. Several techniques have been developed to correctly reconstruct light fields from recorded data [Isaksen et al. 2000; Stewart et al. 2003] and avoid both spatial and inter-view aliasing on automultiscopic displays [Zwicker et al. 2006; Konrad and Agniel 2006; Didyk et al. 2013]. To further adjust content to a particular device, a few techniques for multi-view content depth manipulation have been proposed [Zwicker et al. 2006; Didyk et al. 2012; Masia et al. 2013]. They focus on depth manipulations to achieve the best trade-off between the blur introduced by inter-view antialiasing and presented depth. To better adjust light fields to different screens, Birklbauer et al. [2012] proposed a retargeting technique for changing the size of a displayed light field. Also resolution limitation in light field reproduction has been recently addressed. Tompkin et al. [2013] introduced a novel technique to increase the resolution of lenticular prints by optimizing lenslet arrays based on the input content. With an increasing interest in light field capture and display, several techniques for manipulating and editing such content have been also introduced. These include methods for light field morphing [Zhang et al. 2002], deformation [Chen et al. 2005], and compositing [Horn and Chen 2007]. Lightfields provide also a great flexibility in the context of stereoscopic content production. Kim et al. [2011] demonstrated a technique for generating stereo image pairs with a per-pixel disparity control where each view is defined as a 2D cut through the 3D lightfield volume.

Seamless Image/Video Compositing To avoid transitions, we want to assure that the light field produced by an automultiscopic display is continuous. To achieve this goal, our work is inspired by the recent advances in seamless image and video editing. In particular, we rely on image stitching techniques [Levin et al. 2004; Jia et al. 2006; Jia and Tang 2008; Eisemann et al. 2011], which can combine different images into one natural-looking composition. Creating continuous light fields is also related to work on video textures [Schödl et al. 2000; Agarwala et al. 2005], where the goal is to create sequences that can be played continuously and indefinitely. We also adapt work on video retargeting by Rubinstein et al. [2008]. All of these techniques heavily rely on gradient-based compositing [Pérez et al. 2003; Agarwala 2007] and graph cut methods [Kwatra et al. 2003], which are also used in our technique.

3 Autostereoscopic Transitions

A standard autostereoscopic screen allows for displaying different views which are visible only from the corresponding locations. If the views are displayed densely enough (i. e., with a high angular resolution), each eye can receive a different view, which leads to stereoscopic viewing. Due to the limited resolution of display panels, such screens can display only a limited number of views, for instance, the high-end automultiscopic Philips BDL5571VS/00 has 28 views. This allows for reproducing only a small part of the light field observed in the real world. In this section, we analyze the light field produced by such displays and show how this limitation impacts the perceived quality. We restrict our discussion to a parallax barrier displays; however, the same analysis holds for lenticular based systems.

3.1 Scene vs. Display Light Field

A light field is a function that describes the light traversing a scene. In this work, our goal is to analyze the light field produced by automultiscopic displays. For this purpose, it is enough to consider a light field as a four-dimensional function $\mathscr{L}$ parametrized using two parallel planes (s,t) and (u,v). In such parametrization, $\mathscr{L}(s,t,u,v)$ corresponds to the image value obtained by intersecting a scene with a ray originating from the first plane at the location (s,t) and passing through the second plane at the location (u,v). For visualization purposes, we limit our discussion to epipolar-plane images (EPI) which are 2D slices through the 4D light field (i. e., parameters t and v are fixed). Such images correspond to a stack of 1D skylines captured from different viewing locations along the horizontal direction. In particular, every point in the scene corresponds to a line in the image, and the slant of this line encodes the depth of the point. Figure 3 shows a simple scene with a corresponding light field. Although the light field is limited in the figure, it extends further along the s and u axes.

To visualize the light field produced by an automultiscopic display, we embedded the screen in the scene (Figure 3, top). For every slit of the automultiscopic display, only a small range of directions can be shown (green cones). The signal shown in these cones is also repeated at other locations along the s axis (red cones), although it does not corresponds to these locations. This creates repetitions in the light field created by the screen. The colored boxes in Figure 3 demonstrate how a fragment of the original light field (middle, dashed line) is encoded in the panel (top), and how this fragment forms replicas in the screen light field (bottom). The repetitive structure of this light field creates discontinuities (bottom, dashed line),

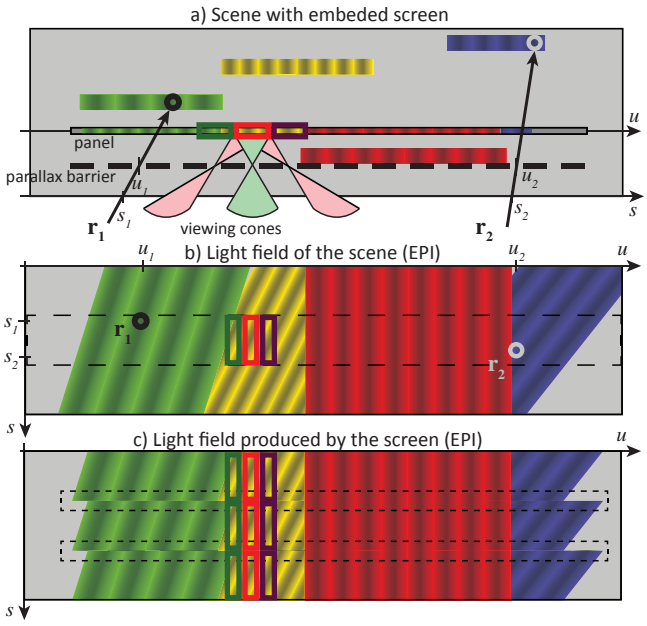

Figure 3: *A simple scene (top) and the corresponding light field (middle). Rays* $\mathbf{r}_1$ *and* $\mathbf{r}_2$ *show the relationship between the scene and the light field representation. To visualize the relation between the scene light field and the light field produced by the display, we embedded the screen in the scene (top). For simplicity, we assumed its alignment with the u axis. The bottom image shows the light field that is produced by the display. Due to the limited angular coverage of the display with views (green cone), the screen is able to reproduce only a part of the original light field (marked with the dashed line in (b)). Beyond this range, the screen creates replicas of the light field, which results in discontinuities (dashed line in (c)).*

which can significantly affect the quality of perceived images.

3.2 Repetitive Light Field and Quality

The images observed on an automultiscopic screen correspond to a cut through the light field created by the screen. For example, for a viewing location located on the *s* axis, the view is a horizontal cut ($\mathbf{x}_1$ and $\mathbf{x}_2$ in Figure 4). As the observer moves along the *s* axis, different skylines of the EPI are observed. Whenever the viewer moves away from the *s* axis, the observed image no longer corresponds to a skyline, but to a slanted line ($\mathbf{x}_3$ and $\mathbf{x}_4$ in Figure 4).

The repetitive structure of the light field produced by an automultiscopic display may lead to visual artifacts. For example, when a view corresponds to a slanted line in the EPI, it may cross several replicas of the original light field. This creates a discontinuity in the perceived image at locations that correspond to the boundaries of the replicas. Furthermore, when the observer moves, these artifacts become more apparent as they change their location. Please see our interactive demo and supplementary video for a visualization.

The discontinuities have also a significant influence on depth perception. In EPIs, the depth of the scene is encoded by the slops of lines that correspond to the same points in the scene. In contrast, the perceived depth is related to the slope of the line that passes through the intersections of the line corresponding to a given point in the scene and the lines corresponding to the left- and right-eye view (Figure 5). In the case when both eyes see the same replica

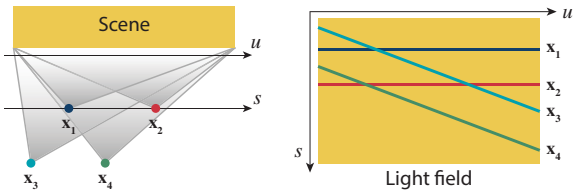

Figure 4: *The figure visualizes a scene together with its light field in the two-plane parametrization. Images observed on the screen correspond to linear cuts through the epipolar-plane images. While views located on the s axis correspond to horizontal lines, viewing positions, which are away from the axis, result in slanted lines. As slants encode the distance to the scene, two viewing positions located at the same distance result in parallel lines.*

of the light field (Figure 5a), the perceived depth is correct. However, when the views correspond to different replicas, the estimation of the depth may be wrong (Figure 5b). In particular, the sign of the slope changes, creating a depth reversal or excessive disparities, which may lead to viewing discomfort [Shibata et al. 2011]. Depending on the viewing position, the depth reversal can be observed in the entire image or only in some parts (Figure 5c). In Figure 6, we provide an example of a stereoscopic image with and without a depth reversal. For an interactive demonstration please see our demo.

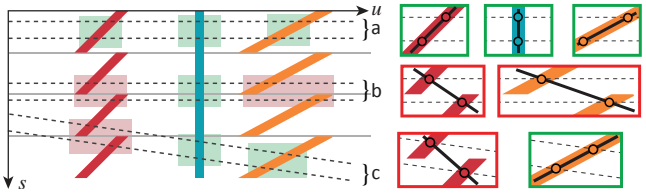

Figure 5: *A simple light field produced by an automultiscopic display showing three objects at different depths with three different stereoscopic viewing locations indicated by pairs of dashed lines (each line corresponds to a view for one eye). Insets on the right present close-ups at the light field. The relative slopes of the black solid lines with respect to the dashed lines correspond to the perceived depth. Green and red colors indicate correct (green) and wrong (red) depth reproduction. When both eyes look at the same replica (a), the slopes (depth) estimated by the observer are correct. When both eyes see different replicas (b), the estimated slopes change the sign, and reversed depth is produced for the whole image. When the observer moves away from the plane s, the lines corresponding to the views are slanted and depth may be estimated incorrectly for some parts of the image (c).*

4 Our Method

We introduce the idea of modifying multi-view content to reduce the artifacts caused by the discontinuities between viewing zones of an automultiscopic display. Our key idea is that by applying subtle modifications to the input content, we can improve the continuity of the light field at those locations, thereby hiding the display imperfections.

4.1 Light Field Shearing

Our first observation is that in specific cases, discontinuities in a light field can be completely removed if the multi-view content is carefully designed. For example, we can utilize a repetitive struc-

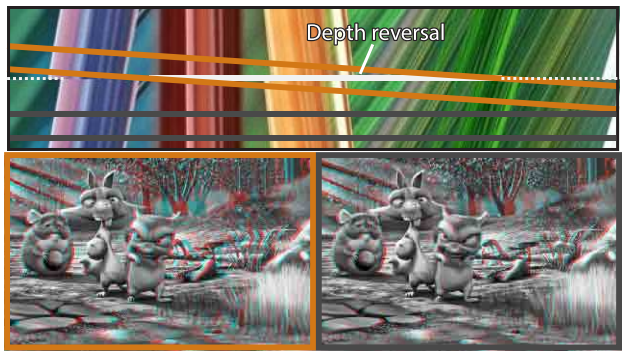

Figure 6: *The figure presents stereoscopic images (in anaglyph colors) with (left) and without (right) depth reversals. The image on the left has a depth reversal as the views are composed of two different replicas of the original light field (EPI on top).*

ture in a scene. Figure 7 shows a light field produced by an automultiscopic display for a scene with a periodic pattern located at a certain depth (left). By applying a small horizontal shear to the original light field, a new light field without any transitions can be obtained (right). As the slope of each line corresponds to scene depth, such

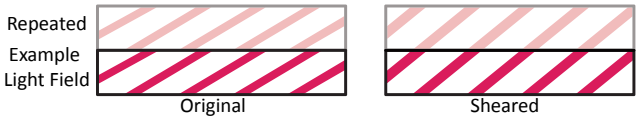

Figure 7: *For a simple planar pattern, the discontinuities in the original light field can be removed by applying a horizontal shear.*

an operation corresponds to re-positioning the entire scene along the depth direction. Although this modifies the absolute depth, it does not significantly affect local depth changes, which dominate depth perception [Brookes and Stevens 1989]. Inspired by this observation, our first step towards reducing discontinuities in a light field is a global horizontal shear followed by small local shears that further improve results. The difference between local and global shears is visualized in Figure 8.

Figure 8: *Visualization of global and local shear: input light field on the left, output on the right. The arrows indicate the magnitude of the shear applied at each location.*

Global shear is defined by one value s, which encodes the amount of shear that needs to be applied to the last view of the light field shown on a screen to best match the first one. Instead of modifying individual EPIs separately, we apply the shear to the entire 3D light field, and compute the optimal shear on full 2D views using the following formula:

$$\arg\min_{s} \frac{1}{N_p} \sum_{(x,y)} Q(I_1, I_n, x, y, s) \qquad (1)$$

where I_1 and I_n are the first and last views presented on the screen, N_p is the total number of pixels, and Q is a matching error between the local neighborhood of a pixel (x,y) in I_1 and the neighborhood of $(x+s,y)$ in I_n. Here, we adopt a matching function introduced by Mahajan et al. [2009], which was originally proposed for optical flow correspondence:

$$Q(I_1, I_n, x, y, s) = \sqrt{\frac{\nabla I_1(x,y) - \nabla I_n(x+s,y)^{\,2} + 0.5\, I_1(x,y) - I_n(x+s,y)^{\,2}}{\sigma(I_1,x,y)\;\;\sigma(I_n,x+s,y)}} \qquad (2)$$

where ∇I is the gradient of image I and $\sigma(I,x,y)$ represents the standard deviation in the 9 9 neighborhood of pixel (x,y) in view I. To find the best s we iterate over integer values in the range between s_{min} and s_{max} and choose a value that results in the smallest value of the matching function. We found that values $s_{min} = -200$ and $s_{max} = 200$ were sufficient for all results presented in this paper.

Local Shears The optimization defined by Equation 1 finds a large global shear that minimizes the matching error between the first and the last view. To further improve the continuity of the light field, we can refine it using small local shears. The idea is similar to finding an optical flow between the two views and minimizing the difference between them using a warp guided by the flow field. Such a solution, however, would lead to flattening the entire scene. To prevent this, we restrict the local warps to be small, which results in matching similar structures instead of the same objects. Finding a dense correspondence between views may also introduce an additional problem of disocclusions. This may lead to significant compression and stretching artifacts during the warping. To avoid these problems we define a regular grid (20 20), and find the optimal shears only for the grid points. This allows us to find good shear magnitudes that vary smoothly across different locations. Later, the coarse grid is warped to improve the continuity of the light field. During this process, the problematic regions are filled-in using the neighboring signal.

The problem of finding the optimal local shear can be formulated as a minimum graph cut. To this end, for each grid point (i,j) we create multiple nodes (i,j,s), where s spans the whole range of integer values from $[s_{\min}, s_{\max}]$, and corresponds to different magnitudes of shear considered at each location. In our implementation we use $s_{\min} = -10$ and $s_{\max} = 10$. The edges in the graph are between (i,j,s) and $(i,j,s+1)$, and encode the cost of the shear s at the position (i,j). The cost is defined as $E(i,j,s) = Q(I_1, I_n, i, j, s)$. In order to find a cut that defines the optimal shears, we add a source and a target node (S,T) to the graph, which are connected to all $(i,j,s_{\min})$ and $(i,j,s_{\max})$ respectively. Additionally, to guarantee that the cut is continuous and passes through every position (i,j) only once, we adapt the idea of forward edges introduced by Rubinstein et al. [2008], and add additional edges with an infinite cost (Figure 9, right). For more details please refer to the original method. After finding the optimal cut of the graph, the amount of shear at the position (i,j) is defined by the location of the cut, i. e., if the cut goes through edge (i,j,s) $(i,j,s+1)$, the optimal shear for position (i,j) is s. In order to apply optimal shears to the light field, we first propagate them from I_n to all views using linear interpolation, assuming that view I_1 receives zero shear. Then, every view is separately sheared by warping the grid together with the underlying view.

4.2 Light Field Stitching

Shearing techniques that are discussed in the previous section can already align the structure of the repeating light field fragments. However, sharp color differences can still remain visible. In order to reduce them, we apply an additional compositing of repeating light field structures in a gradient domain. Inspired by image/video stitching and retargeting techniques [Jia et al. 2006; Jia and Tang 2008; Rubinstein et al. 2008; Eisemann et al. 2011], we use a similar technique to further hide the transitions. To this end, we first create two copies of the original light field and overlap them by m

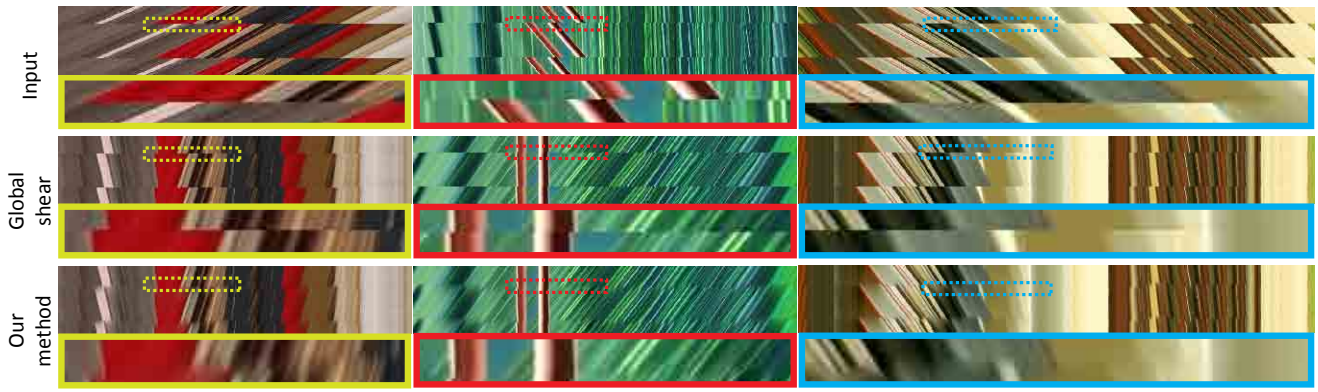

Figure 10: *Examples of light fields produced by our technique. From the top: input light fields, only global shearing applied, our full method, insets with close-ups at the discontinuities.*

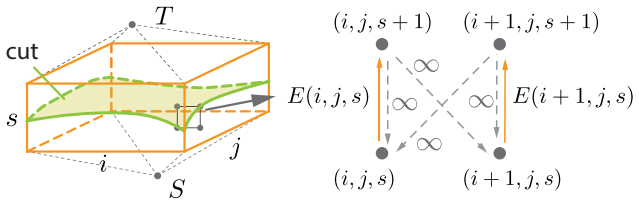

Figure 9: *The local warping could be regarded as a continuous cut of the function E (left). The weights on edges are shown on the right. For simplicity, we show the forward edges (dashed lines) only in 2D. In general, forward edges are added in the $s-x$ and $s-y$ planes.*

views along the s direction. Then, we find a cut through the overlapping part, which provides us with a surface where both replicas fit best. This cut, similarly to finding good shears, can be found using the graph cut technique. We first transform the overlapping light field volume into a graph, where each voxel (s,u,v) corresponds to a node. The edges between (s,u,v) and $(s+1,u,v)$ encode the cost of the cut between these two voxels. The goal of this cost is to penalize large differences in gradients between the overlapping replicas, and we express it as:

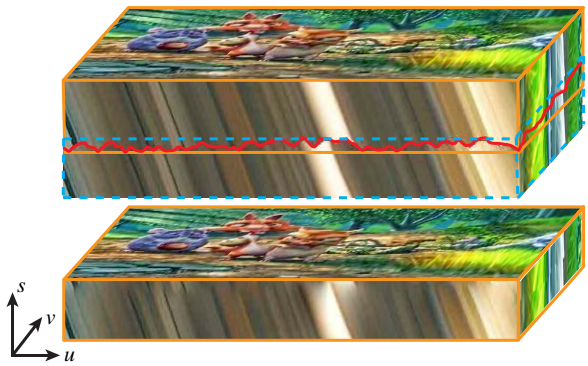

Figure 11: *Light field stitching. We overlap light field copies and find the optimal cut. The reconstructed light field is shown at the bottom.*

$$C(u,v,s) = \nabla_{su}\mathscr{L}(s,u,v) - \nabla_{su}\mathscr{L}(n-m+1+s,u,v) + \nabla_{su}\mathscr{L}(s+1,u,v) - \nabla_{su}\mathscr{L}(n-m+2+s,u,v) \quad , \tag{3}$$

where $\nabla_{su}\mathscr{L}$ is the (s,u) component of the light field gradient, n is the total number of views, and m is the number of views that are overlapped. (s,u,v) and $(n-m+1+s,u,v)$ as well as $(s+1,u,v)$ and $(n-m+2+s,u,v)$ are positions that are directly overlapping. Similarly to the construction of the graph for the local shearing, we add forward edges with an infinite cost as well as a source and a target node to perform minimal graph cut. After finding the optimal cut in the graph, we stitch gradients of the overlapping light field replicas along the cut, and compute the full light field by reconstructing each EPI separately using Poisson reconstruction [Pérez et al. 2003]. The whole process is visualized in Figure 11. The width of the overlap m controls the number of views that are affected by the method. From our experience, around 4 views are necessary to create a good transition between different viewing zones. Therefore, for a display with 8 views we use $m = n/2$. For displays that offer higher numbers of views, smaller m can be used. In our results, we use $n/2$ and $n/4$.

4.3 Extension to Videos

So far we have described our method only for static light fields. The direct extension of the shearing and stitching to videos includes a computation of a minimal graph cut for a 4D volume and Poisson reconstruction in 3D. To avoid high computation costs, a full computation can be performed for every k-th frame. In our experiments we use $k = 50$. Later, the shears as well as the cuts can be linearly interpolated for the remaining frames. Besides better performance, such a solution provides temporally coherent results (see supplementary materials).

5 Results

We have tested our technique on a variety of images and videos. Figure 10 shows comparison of the individual EPIs for several scenes. Compared to the original light field and to one where only global shear was applied, our full technique provides much smoother results. In many high frequency regions, our method successfully finds the local repetitive structures and eliminates transitions. The stitching propagates transitions optimally into different views in different regions, making them less pronounced. Additionally, in Figure 12, we provide a comparison of four views that are generated using our method to a version where only global shear is applied to align both light fields in the same way around the screen. Our results present smoother transitions with less pronounced discontinuities and depth reversals as well as fewer diagonal strips.

Figure 12: *The comparison of the results obtained using full technique and only global shearing on an automultiscopic screen (simulation). For better perception of depth, we remove the colors and show a monochrome anaglyph; please refer to the supplementary materials for full results.*

Performance Processing one multi-view image composed of 100 views (1200 800 pixels) using our Matlab implementation takes 1 minute. This includes 5 seconds for shearing and stitching, and 55 seconds for Poisson reconstruction which is the main bottleneck as it needs to be performed for each epipolar plane image separately. Processing 80 frames of a multi-view video in resolution 800 540 takes almost 1 hour, which is also limited by the Poisson reconstruction. We believe that the performance of our technique can be greatly improved. For example, the method is highly parallelizable, i. e., every shot can be processed separately. Also, for slowly changing scenes, the full computation can be performed for fewer frames. These improvements together with a GPU implementation, can reduce the computation time significantly.

6 Evaluation

In order to evaluate the quality improvement provided by our technique, we conducted user experiments in which we first compared performance of our automatic global shear with manual adjustment done by users and then evaluated the full technique. In all experiments 16 participants took part. They had normal or corrected-to-normal vision and were tested for stereoblindness.

6.1 Manual Adjustment vs. Global Shear

The global shear adjusts the position of the entire scene with respect to the screen plane. A similar, manual correction is a common practice to reduce the need of inter-view antialiasing [Zwicker et al. 2006; Didyk et al. 2013] and visual discomfort [Shibata et al. 2011]. To validate our approach, we compared it to the same operation performed manually. To acquire the optimal correction, three video sequences (Figure 13) were presented to each participant. The sequences were displayed using a 42-inch 8-view automultiscopic Newsight Japan display. The participants were asked to sit 1.5 m away from the screen, which is the optimal viewing distance for this display, and adjust the global depth of the scene until the best viewing quality is achieved. From collected data, the average adjustment for each scene was computed. The same content was processed using our global shear. The resulting adjustments were compared with the results from the user study. The small differences and relatively high inter-subject variability (Table 1) suggest that the difference between the global shear and the manual adjustments is small. Although it is unclear whether this observation holds in general, in the further evaluation we decided to compare our full technique only to the global shear as both of them provide an automatic solution.

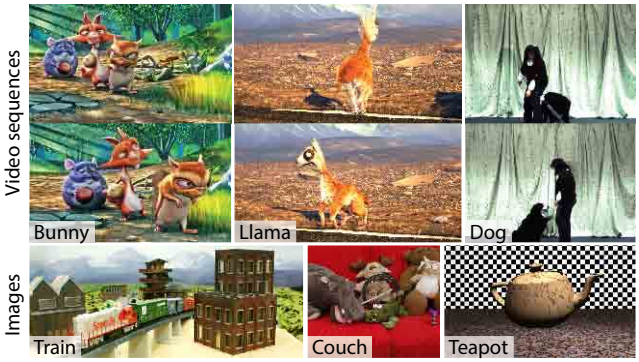

Figure 13: *Examples used in our experiments.* COUCH *[Kim et al. 2013],* TRAIN *and* DOG *were captured using a camera array.* BUNNY, TEAPOT *and* LLAMA *were rendered.*

Scene	Δ adjustment	σ	*SEM*
BUNNY	0.5 px	1.3	0.3
LLAMA	-0.2 px	0.7	0.17
DOG	-0.32 px	0.6	0.15

Table 1: *Statistics for the manual adjustment of the content.* Δ *adjustment represents the difference between correction provided by our global shear and manual adjustment provided by users. The difference is expressed as a change of the disparities between neighboring views and measured for Full HD resolution. Additionally, standard deviation* (σ) *and standard error of the mean (SEM) are provided.*

6.2 Global Adjustment vs. Full Technique

In the second experiment, we compared the results obtained using our automatic global shear with sequences produced using the full technique. In order to evaluate how our method performs for different viewing locations, 8 viewing positions were tested. In each trial, participants were asked to stay in one location which was marked on the floor. They were presented with both versions of the content, and could switch between them using a keyboard. The task was to judge which one provided a better viewing experience. There were no other specific instructions provided to the observers. Each of them performed the task for every location and for all videos. The arrangement of the experiment and the results for all the viewpoints are shown in Figure 14. Although the participants were instructed

to remain in the same locations, some of them were adjusting their head position to find a better viewing spot. We consider this to be a natural scenario for watching such content. In 68% of cases, partic-

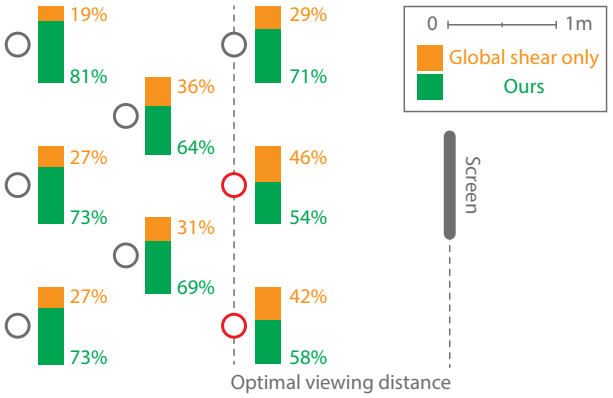

Figure 14: *Results of the experiment for automultiscopic displays. The numbers correspond to percentage of people who chose our full technique (green) and the content processed using only global shear (yellow). The places for which the results were not statistically significant are marked in red.*

ipants preferred sequences processed using the full technique. This demonstrates that besides improving view transitions our method does not introduce temporal artifacts which would reduce the quality of the content. To test significance of the results we performed a series of binomial tests. All results were significant ($p < 0.05$) except for two viewing positions located in the optimal distance from the screen. In this experiment, all views were affected by our technique. We believe that for displays with larger numbers of views, this will not be required, and better quality can be achieved.

We also conducted a similar experiment with 4 different static light fields: TRAIN, COUCH, TEAPOT and BUNNY (Figure 13). As previously, each was prepared in two versions: one processed with the full technique, and the other with the global shear only. Both versions were downsampled to 18 views, printed at 720 DPI and glued to lenticular sheets with 40 lenses per inch with a viewing angle of 25 . For better comparison, both versions were printed on a single sheet. We also produced additional sheets where the versions were swapped for randomization. In every trial, one stimulus was shown to a participant who could look at each lenticular sheet from different angles and distances. They were allowed to take as much time as they wanted to investigate the print. Afterwards, they were asked to decide which version offered a better viewing experience. Similarly to previous experiments, we did not provide any further guidelines, so every participant was free to use any criteria to judge the quality. The results of the experiment are shown in Figure 15. In more than 80% of trials, participants preferred results produced using our technique. To verify significance of of our results, we computed a binomial test for each scene separately. For all examples the obtained p-value was below 0.05.

7 Discussions

In our approach, we take two main steps to process the light field: shearing (global and local) and stitching. To better justify the role of each operation, we provide examples of our technique with certain steps omitted in Figure 16.

Our technique cannot remove transitions and depth reversals completely as this would require complete flattening of the scene, i. e., falling back to 2D viewing. However, due to our stitching across

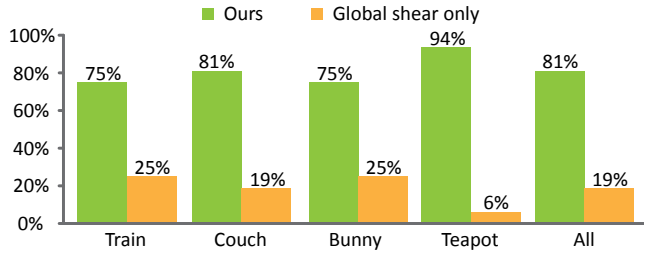

Figure 15: *Results of our experiment with lenticular prints. For each scene, the bars represent a percentage of subjects who judged the quality of our (green) and global shear (yellow) as higher.*

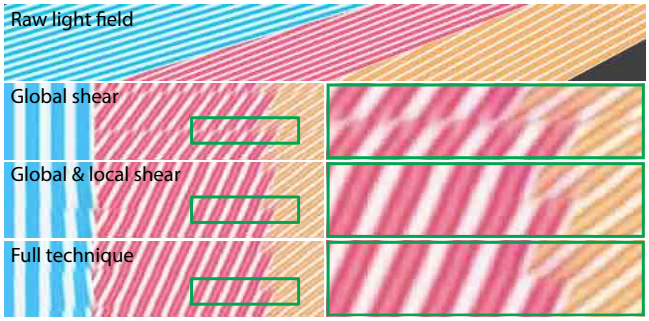

Figure 16: *Results of our technique with certain steps omitted. The global shear of the original light field reduced excessive disparities and provided a good alignment of the blue lines. Local shearing significantly improved the discontinuities between the red lines. In this case, this is achieved by locally sacrificing the continuity of the blue lines, which does not in uence significantly the overall quality across all views. The complete technique was able to match the red and yellow lines to further improve the continuity between the replicas. The Poisson reconstruction was applied in all examples.*

many views (Section 4.2), the remaining artifacts are distributed across different views. In contrast to the original content, where transition areas are large and have obvious structure, the artifacts that remain after our technique was applied are small and local. To demonstrate this, we followed suggestions by Filippini et al. [2009] that some stereoscopic effects can be explained by local cross-correlation. Consequently, we used this approach to compute depth for sequences before and after our techniques was applied (Figure 17). Although it is an objective comparison, it cannot be treated as an accurate prediction of what the observer perceives. From our experience, small reversals are usually less objectionable. This can be explained by the fact that people are less sensitive to high frequency depth variations.

Although distributing transitions across different views may affect the sweet-spot viewing, it was not manifested in our experiments. We believe that for displays with more views (e. g., Philips BDL2331VS/00 has 28 views), the stitching can be performed only on a small part of the light field near viewing zone boundaries. In our examples for $m = n/4$, the resulting light field contains $3/4$ of all views, and our stitching step affects $1/4$ of them. As a result, the content of $2/3$ of all views shown on the screen remains unchanged. To avoid limiting the number of input views that are shown on the screen, view synthesis techniques can be used to create additional views for the purpose of stitching (Figure 18).

Depending on the display (e.g., number of views) and the viewing conditions (e.g., range of viewing distances) different sets of views might be needed. To avoid performing expensive computation on

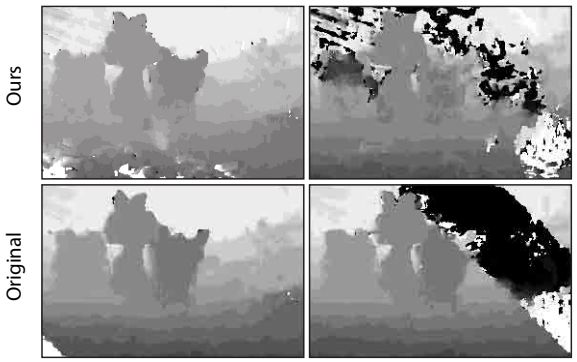

Figure 17: *Perceived depth computed using cross correlation for two original views, and the same views processed using our technique. While the original content produces pronounced depth reversals, our technique provides views for which resulting depth contains less structured depth errors.*

Figure 18: *Here, the original light field (*100 *views) is expanded to* 200 *views using image-based warping. After stitching, the resulting light field has the same number of views as the input content. Please refer to the supplementary materials for the complete light field.*

the display, several versions of the content can be prepared and delivered to the client. In the case of streaming systems, the correct version can be sent alone.

Our technique benefits from repetitive patterns, which are common in many natural scenes (e. g., trees, grass, ground, clouds). As can be seen in Figure 10, our local shearing finds good matches and improves the quality of the light field. For some cases our method has limited options, but it can reduce the transitions problem by placing these objects close to the screen plane, which is also desirable for eye strain reduction.

8 Conclusions

In this paper, we analyzed the light field created by standard automultiscopic screens. First, we showed that the limited angular coverage results in a periodic structure of the light field with discontinuities between repeating fragments. Then, we explained how this leads to a number of artifacts such as view discontinuities, depth reversals, and excessive disparities. To the best of our knowledge, we are the first to address these issues together with their analysis and explanation. To overcome these limitations, we presented the novel idea of modifying the 3D content shown on automultiscopic screens. Our method improves the continuity of the light field, which leads to significant visual quality improvements. The performance of the method was demonstrated on static images as well as videos, and validated in user experiments. Additional advantage of our technique is device- and view-independence, i. e., it does not utilize any information about display type or viewers' positions. These together with the fact that it is a purely software solution make it attractive as a pre-processing step for a wide range of applications.

Extending our method to a full-parallax display is an exciting avenue for future work and we consider this a non-trivial problem. First, the analysis of the problem needs to be extended from 2D EPI images to 3D. Then, the technique would need to enforce the repetitive structure in both horizontal and vertical directions. Our technique does not apply directly to multi-layer displays; however, we believe that some ideas from our work could be used to expand their small field of view. Another interesting possibility would be to combine such manipulations with depth remapping methods and inter-view antialiasing as well as to improve performance for real-time applications. We believe that our technique will be beneficial not only for 3DTV applications and 3D visualizations, but also for large scale projector-based cinema systems.

Acknowledgments

We would like to thank Katarina Struckmann and Krzysztof Templin for proofreading, and the Anonymous Subjects who took part in our experiments. This work was partially supported by NSF IIS-1111415, NSF IIS-1116296, Quanta Computer, the National Basic Research Project of China (Project Number 2011CB302205), the Natural Science Foundation of China (Project Number 61272226/61120106007), the National High Technology Research and Development Program of China (Project Number 2013AA013903), and Research grant of Beijing Higher Institution Engineering Research Center.

References

AGARWALA, A., ZHENG, K. C., PAL, C., AGRAWALA, M., COHEN, M., CURLESS, B., SALESIN, D., AND SZELISKI, R. 2005. Panoramic video textures. *ACM Trans. Graph. 24*, 3, 821–827.

AGARWALA, A. 2007. Efficient gradient-domain compositing using quadtrees. *ACM Trans. Graph. 26*, 3, 94.

BIRKLBAUER, C., AND BIMBER, O. 2012. Light-field retargeting. *Computer Graphics Forum 31*, 2pt1, 295–303.

BROOKES, A., AND STEVENS, K. A. 1989. The analogy between stereo depth and brightness. *Perception 18*, 5, 601–614.

CHEN, B., OFEK, E., SHUM, H.-Y., AND LEVOY, M. 2005. Interactive deformation of light fields. In *Proceedings of the 2005 symposium on Interactive 3D graphics and games*, ACM, 139–146.

DIDYK, P., RITSCHEL, T., EISEMANN, E., MYSZKOWSKI, K., SEIDEL, H.-P., AND MATUSIK, W. 2012. A luminance-contrast-aware disparity model and applications. *ACM Trans. Graph. 31*, 6, 184.

DIDYK, P., SITTHI-AMORN, P., FREEMAN, W., DURAND, F., AND MATUSIK, W. 2013. Joint view expansion and filtering for automultiscopic 3D displays. *ACM Trans. Graph. 32*, 6, 221:1–221:8.

EISEMANN, M., GOHLKE, D., AND MAGNOR, M. 2011. Edge-constrained image compositing. In *Proceedings of Graphics Interface 2011*, Canadian Human-Computer Communications Society, 191–198.

FILIPPINI, H. R., AND BANKS, M. S. 2009. Limits of stereopsis explained by local cross-correlation. *Journal of Vision 9*, 8–8.

HIRSCH, M., WETZSTEIN, G., AND RASKAR, R. 2014. A compressive light field projection system. *ACM Trans. Graph. 33*, 4, 1–12. to appear.

HORN, D. R., AND CHEN, B. 2007. Lightshop: interactive light field manipulation and rendering. In *Proceedings of the 2007 symposium on Interactive 3D graphics and games*, ACM, 121–128.

ISAKSEN, A., MCMILLAN, L., AND GORTLER, S. J. 2000. Dynamically reparameterized light fields. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, ACM Press/Addison-Wesley Publishing Co., 297–306.

IVES, F. E., 1903. Parallax stereogram and process of making same. U.S. Patent 725,567.

JIA, J., AND TANG, C.-K. 2008. Image stitching using structure deformation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 30*, 4, 617–631.

JIA, J., SUN, J., TANG, C.-K., AND SHUM, H.-Y. 2006. Drag-and-drop pasting. *ACM Trans. Graph. 25*, 3, 631–637.

KIM, C., HORNUNG, A., HEINZLE, S., MATUSIK, W., AND GROSS, M. 2011. Multi-perspective stereoscopy from light fields. *ACM Trans. Graph. 30*, 6, 190.

KIM, C., ZIMMER, H., PRITCH, Y., SORKINE-HORNUNG, A., AND GROSS, M. 2013. Scene reconstruction from high spatio-angular resolution light fields. *ACM Trans. Graph.*.

KONRAD, J., AND AGNIEL, P. 2006. Subsampling models and anti-alias filters for 3-D automultiscopic displays. *Image Processing, IEEE Transactions on 15*, 1, 128–140.

KWATRA, V., SCHÖDL, A., ESSA, I., TURK, G., AND BOBICK, A. 2003. Graphcut textures: image and video synthesis using graph cuts. *ACM Trans. Graph. 22*, 3, 277–286.

LEVIN, A., ZOMET, A., PELEG, S., AND WEISS, Y. 2004. Seamless image stitching in the gradient domain. In *Computer Vision-ECCV 2004*. Springer, 377–389.

LEVOY, M., AND HANRAHAN, P. 1996. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, ACM, 31–42.

LIPPMANN, G. 1908. Épreuves réversibles donnant la sensation du relief. *Journal of Physics 7*, 4, 821–825.

MAHAJAN, D., HUANG, F.-C., MATUSIK, W., RAMAMOORTHI, R., AND BELHUMEUR, P. 2009. Moving gradients: a path-based method for plausible image interpolation. *ACM Trans. Graph. 28*, 3, 42.

MASIA, B., WETZSTEIN, G., ALIAGA, C., RASKAR, R., AND GUTIERREZ, D. 2013. Display adaptive 3D content remapping. *Computers & Graphics 37*, 8, 983–996.

PÉREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. *ACM Trans. Graph. 22*, 3, 313–318.

PETERKA, T., KOOIMA, R. L., SANDIN, D. J., JOHNSON, A. E., LEIGH, J., AND DEFANTI, T. A. 2008. Advances in the Dynallax Solid-State Dynamic Parallax Barrier Autostereoscopic Visualization Display System. *IEEE Transactions on Visualization and Computer Graphics 14*, 487–499.

RUBINSTEIN, M., SHAMIR, A., AND AVIDAN, S. 2008. Improved seam carving for video retargeting. *ACM Trans. Graph. 27*, 3, 16:1–16:9.

SCHÖDL, A., SZELISKI, R., SALESIN, D. H., AND ESSA, I. 2000. Video textures. In *Annual Conference on Computer Graphics*, SIGGRAPH '00, 489–498.

SHIBATA, T., KIM, J., HOFFMAN, D., AND BANKS, M. 2011. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision 11*, 8, 11:1–11:29.

STEWART, J., YU, J., GORTLER, S. J., AND MCMILLAN, L. 2003. A new reconstruction filter for undersampled light fields. In *Proceedings of the 14th Eurographics workshop on Rendering*, Eurographics Association, 150–156.

TOMPKIN, J., HEINZLE, S., KAUTZ, J., AND MATUSIK, W. 2013. Content-adaptive lenticular prints. *ACM Trans. Graph. 32*, 4, 133:1–133:10.

YE, G., STATE, A., AND FUCHS, H. 2010. A practical multi-viewer tabletop autostereoscopic display. In *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*, IEEE, 147–156.

YI, S.-Y., CHAEAND, H.-B., AND LEE, S.-H. 2008. Moving parallax barrier design for eye-tracking autostereoscopic displays. In *3DTV Conference*.

ZHANG, Z., WANG, L., GUO, B., AND SHUM, H.-Y. 2002. Feature-based light field morphing. *ACM Trans. Graph. 21*, 3, 457–464.

ZWICKER, M., MATUSIK, W., DURAND, F., AND PFISTER, H. 2006. Antialiasing for automultiscopic 3D displays. In *Proceedings of the 17th Eurographics conference on Rendering Techniques*, Eurographics Association, 73–82.