

A hybrid multi-view and eye-tracked transparent auto-stereoscopic display for augmented reality

Charlie Schlick^{1,2}, Thomas Crespel^{1,3}, Xavier Granier^{3,4} and Patrick Reuter^{1,2};

¹Univ. Bordeaux, Inria, and ²Bordeaux INP, CNRS (LaBRI UMR 5800), and ³Institut d'Optique Graduate School - CNRS (LP2N UMR 5298), and ⁴CNRS (Archéosciences Bordeaux UMR 6034); Talence, France

Abstract

We put forward the use of transparent 3D displays for augmented reality. For a glasses-free experience with autostereoscopy and a large viewing area, we study the use of a recent transparent display with multiple discrete and horizontally adjacent viewing zones. Although promising, this display cannot directly be used for augmented reality due to inconsistencies within and between the discrete viewing zones. In this work, we propose to overcome this limitation by tracking the user's eyes for ensuring continuous transitions, thus making the display feasible for augmented reality. In particular, we compensate the intensity variations, we ensure a consistent horizontal parallax within and between the adjacent viewing zones, and we add vertical parallax. In this way, the display becomes a transparent augmented window that can be used for various augmented reality applications. We present results on a display with 5 viewing zones for three different use cases, evaluate the appropriateness, discuss the limitations, and show future directions.

Introduction

Augmented reality merges virtual worlds into our real-world environment. Among the different approaches, the most widely known ones are the augmentation of live-acquired video streams on hand-held displays such as tablets, or the use of head-attached displays based on transparent optical combiners to merge the real and virtual worlds. Both approaches require the user to wear or hold special devices, and in the case of hand-held displays, the user does not really see the real world, but its video stream on a 2D opaque screen. Another approach is spatial augmented reality [4], also known as projection mapping, where a projector directly projects images on the surface to augment. However, no augmentation can be done "in the air", since a receiving surface has to be present. Transparent 3D displays [6], as part of spatial optical see-through displays [4], are a path to explore in order to provide an augmentation of real objects beyond the projection on their surface, and without requiring that users carry dedicated equipment.

The goal is to reproduce as many depth cues as possible, and especially binocular vision and full parallax, while still being able to produce the virtual, environment-dependent content in real-time. There are two major interrelated challenges for this augmented reality approach [5]. First, it requires a display that enables 3D automultiscopy and transparency at the same time, and second, producing multi-view images has to be feasible in real-time. Indeed, multi-view displays, such as for example tensor displays [24], microlens arrays [1], a high number of projectors [16], or a rotating screen [7], require the rendering of a full light-field, and thus image generation for a massive number of different viewpoints. This involves a high computational cost and can thus hardly be done in real-time with satisfying spatial and angular resolutions on commodity hardware.

In this work, we assume that high framerates and high resolution images for each viewpoint still need to reduce the number

of viewpoints. This is the case of a recent transparent display that is based on multiple projectors for creating discrete and horizontally adjacent viewing zones [6]. Although promising for augmented reality, as is, such a display lacks consistency in intensity, presents no vertical parallax, and horizontally, there are visible transitions between the viewing zones. With wearable displays, these issues can be solved using tracking devices. In our case, we want natural augmentation without wearables. Starting from a transparent display with multiple views, such as [6], we add less efficient but remote eye tracking systems to perform the required adjustments. In particular, we compensate the intensity variations, we ensure a consistent horizontal parallax within and between the adjacent viewing zones, and we add vertical parallax. This is possible by adjusting the viewing zones to be very slightly overlapping, and by adjusting the images for ensuring continuous transitions according to the user's eye positions. Altogether, this leads to a transparent, glasses-free display providing a large viewing area (eyebow) with continuous and consistent parallax, while limiting the number of projectors. This approach requires only a few images to be rendered, and it is thus suitable for interactive applications. Consequently, the contributions of this work are as follows:

- we propose hardware and software solutions for real-time adjustments of image intensity,
- we demonstrate that simple eye-tracking can be used to ensure continuous transitions for intensity and views,
- we thus show that the combination of eye-tracking and a recent 3D transparent display provides the user with an improved experience of glasses-free augmented reality, and
- we illustrate these features on some new potential applications for such a display.

With the characteristics of transparency, 3D multiscopy, interactivity, and thanks to the eye tracking, the display becomes a 3D augmented window on the real world.

Related work

3D displays try to reproduce as many depth cues as possible, and in particular binocular disparity and the resulting convergence, motion parallax, and accommodation. Among the different types of 3D displays, the most available ones today are autostereoscopic displays based on parallax barriers [23]. On the other hand, holographic displays aim at reproducing the full depth of a virtual scene with an interference-based approach, but they involve cost-intensive computations [13]. Another promising 3D display approach are light-field displays, either based on dense projector arrangements [3], multiple layers [24], or a 2D array of lenses [1]. In the latter integral imaging approach, the number of views that are displayed simultaneously can be massive, and if multiple rays from the same point enter the eye, then the eye is forced to accommodate to see it sharp, solving the vergence-accommodation conflict (VAC) [10, 21]. However, the involved computation is very high, and these displays are based

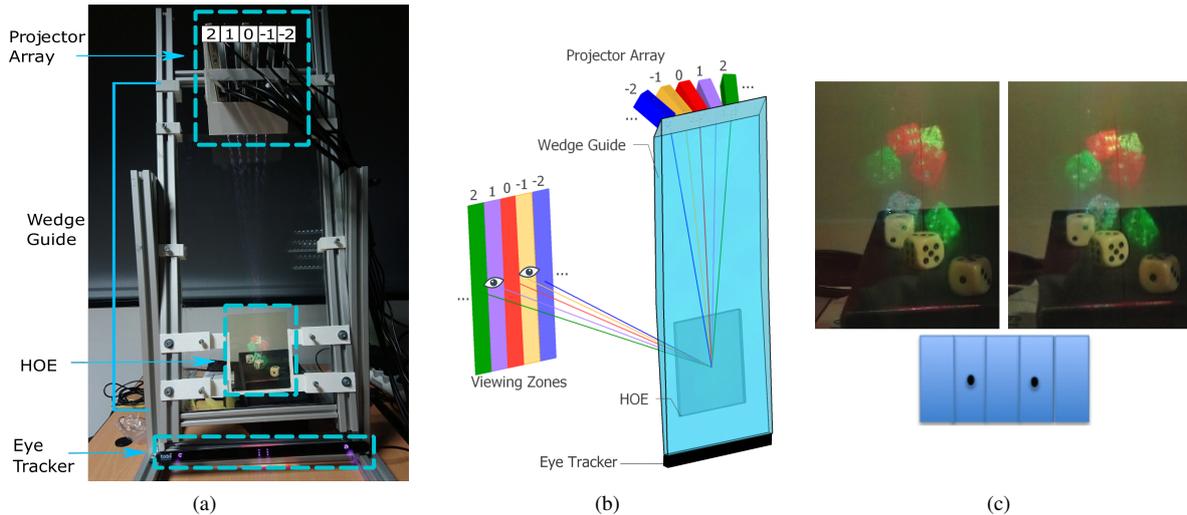


Figure 1. (a-b) Overview of our setup: the transparent autostereoscopic display, where 5 projectors labelled $-2..2$ are coupled into a wedge-shaped light guide and a HOE scatters the light towards independent respective viewing zones $-2..2$, and the eye tracker. (c) Top: the two images with a slightly different perspective for the left and right eye, creating a stereoscopic effect. Bottom: the eyes are positioned in viewing zones -1 and 1 , among the 5 different viewing zones that compose the entire designated viewing area.

on conventional optics that make them difficult to adapt for a transparent autostereoscopic display.

Nevertheless, adaptations for transparency have been proposed for lenticular-based [22] and parallax barrier autostereoscopic displays [11]. However, diffraction effects and low transmittance can be observed. Concerning projection-based approaches, the ASTOR display [19] uses two projectors and a reflective holographic combiner to create two viewing zones. Hong et al. [12] proposed a multi-projection system with a reflective transparent anisotropic diffuser. The autostereoscopic transparent display [6] is also based on multi-projection and avoids free-space projection by guiding light from the projectors in a wedge guide. A transmissive holographic optical element (HOE) redirects light from the projectors into the viewing directions, and lets ambient light pass through, thus making the display transparent. While offering a larger viewing zone over previous systems for a 3D experience, all these displays suffer from limitations in terms of the vergence-accommodation conflict, and in terms of remaining optical aberrations.

Eye tracking has been used to improve displays for 3D perception. In [20], a regular 2D display is used in combination with head tracking for rendering the correct images according to the measured viewpoint, thus fulfilling the motion parallax depth cue. However, it does not provide any binocular cues. In [9], a similar idea is used for introducing parallax into video see-through augmented reality. For near-eye displays, eye tracking can be used for foveated rendering [8]. It has also been used to improve parallax-barrier displays: updating the barrier according to the user's position, and updating the images, results in a higher freedom of movement [25]. For an autostereoscopic multi-view display, eye tracking has been used to fuse overlapping viewing zones and to minimize crosstalk [15]. For a compressive light field display, eye tracking has been used to extend the field-of-view [17], and for autostereoscopic projector array setups with dense horizontal parallax, it can add vertical parallax [14].

Setup proposal

As a setup for our augmented reality approach, we propose to modify an autostereoscopic multi-view transparent display (e.g., [6]), combine it with a simple eye tracking system,

and adjust the rendered images accordingly (Figure 1).

In [6], the user perceives autostereoscopy when both eyes are in a well-defined area. This designated area in front of the display has multiple, horizontally adjacent viewing zones that are designed so that both eyes of the user are always in different viewing zones. In each viewing zone, an image can be seen, together with the real world behind the display. The viewing zones are addressed by computer generated images of the same 3D scene, rendered from different viewpoints to create the stereoscopic effect. This is illustrated in Figure 1(c), where the images for both the left and right eyes are shown, together with their positions in the designated viewing area. Since the real world can be perceived through the transparent display, interactive 3D augmented reality applications may become possible. Despite the fact that such a display is already an improved experience for natural augmented reality, there are remaining drawbacks. First, as for most multi-view displays, the optical paths for the different views may vary, resulting in different more or less important intensity differences. This is especially the case when an HOE is used. Second, like in many autostereoscopic displays, it does not solve the vergence-accommodation conflict. Finally, there are some inconsistencies when the user's eyes are not in the exact centres of the viewing zones for which the images are calculated. Consequently, when the user moves his/her head, the virtual information may still be misaligned with the real world to augment, and transitions are perceived when the user's eyes move to different viewing zones.

In this work, we reduce some limitations by adjusting both the display design and the rendered images. For the display, we introduce slightly overlapping viewing zones, especially for the intensity compensation. We also use simple remote eye tracking to adapt the projected images accordingly in order to explore ways of extending the display's capabilities. As detailed in [6], the initial display has an effective size of 13cm height and 10cm width. The optimal viewing distance is 50cm, and it creates 5 viewing zones that are parallel for a horizontal-only parallax. Each viewing zone is 10cm high and 3cm wide, which corresponds to about half the interpupillary distance. In contrast to [6] where the viewing zones are perfectly adjacent, we adjust the viewing zones to be very slightly overlapping in order to produce

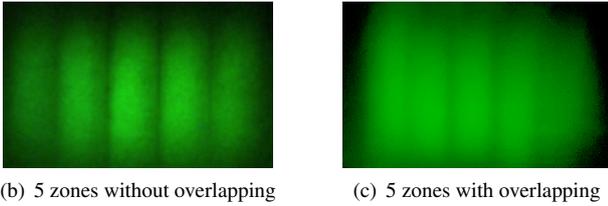
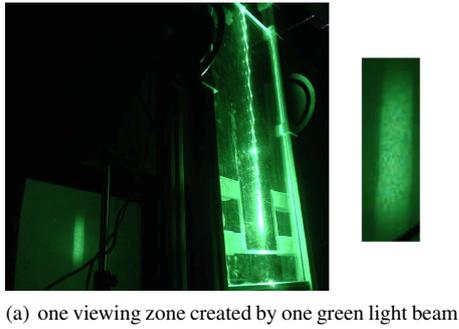


Figure 2. Observed intensity variation on a diffuse screen placed at the viewing zones. (a) The entire setting for one green laser source, where the HOE reconstructs one viewing zone. (b) The reconstruction of the 5 viewing zones at the screen for the 5 projectors, with no overlapping as in [6]. (c) The reconstruction of the 5 viewing zones with 6mm overlapping.

a higher resulting intensity at the edges of the viewing zones. This increase is very important for relative adjustments of intensity in software, since the intensity of the images can only be decreased. Our experiments have shown that an overlapping of 6mm at the viewing distance provides the best trade-off between the intensity and the slight reduction of the width of the viewing zones, as shown in Figure 2.

As an eye tracker, we use the Tobii 4C working at a 90 Hz data frequency. It is located under the display (see Figure 1) and adjusted so that it can precisely track both user’s eyes everywhere in the viewing zones. We use a prior calibration procedure where the user has to position one eye at different key positions of the viewing zones, and in particular, the centre of the overlapping zones, and the exterior limits. For generating the images from the 3D virtual scene, we position the virtual cameras according to the tracked eye positions and the calibration. Our implementation is based on OpenGL, and the image adjustments for the light guide are done directly in a rendering pass on the GPU.

Consistent intensity

We adjust the intensity of the display to make it consistent in the designated viewing area. Indeed, as pointed out in [6], the intensity of the display is neither uniform over all viewing zones, nor uniform within each viewing zone. This is related to the Bragg efficiency of the involved HOE in the display. The intensity in the viewing zones can be measured by putting a diffuse screen at the location of the viewing zones, hence 50cm in front of the display, and by observing the reconstructed viewing zones. In Figure 2(a), we show the path of a green light beam from the entrance of the wedge-shaped light guide to the one resulting viewing zone created by the HOE. This illustrates the total internal reflection in the light guide, and it shows that a point on the HOE spreads its energy toward the full viewing zone. Especially in the close-up, we can observe the non-uniformity within this one viewing zone. In a multi-projector setup, the efficiency is best for the central projector position, and then it decreases for the outer projectors, as illustrated in Figure 2(b) for rendered green projector images.

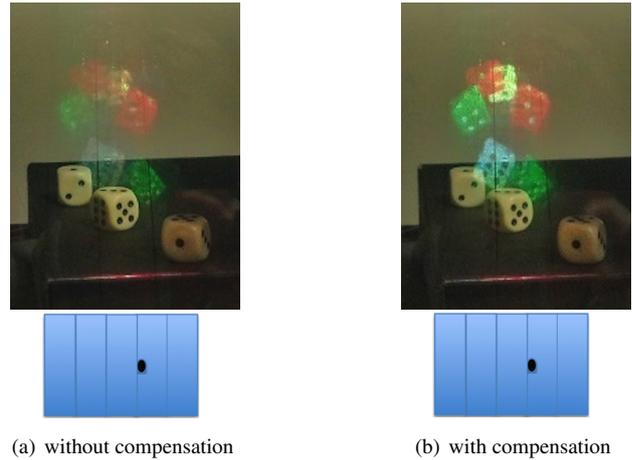


Figure 3. Intensity compensation. At the edge of a viewing zone, the intensity compensation results in a brighter image.

When using the display, we observed that most of the time, the intensity variation becomes only noticeable at the very edge of each viewing zone. In order to compensate this intensity variation, we use eye tracking and adjust the luminosity of all pixels in the rendered images according to the position of both user’s eyes in the viewing zones. As we defined the viewing zones to slightly overlap, we can produce a higher intensity at the edges of the viewing zones as the two images are accumulated, at the cost of a smaller horizontal range, as illustrated in Figure 3.

Note that the HOE in the display is colour-multiplexed, and as the HOE efficiency varies depending on the colour channels, the intensity can be adjusted colour-channel wise.

Consistency for multiple views

In this section, we present how we adjust the display to different interpupillary distances; and how we add full parallax.

True interpupillary distance

In a transparent autostereoscopic display such as [6], for a user to perceive the stereo effect, the width of each viewing zone is related to the user’s interpupillary distance (IPD), and the head cannot be tilted as this decreases the horizontal component of the distance. Otherwise, when a user moves, both eyes may not enter in adjacent viewing zones simultaneously (see Figure 4), and hence the received images are inconsistent. This is quite constraining, especially because the size of the viewing zones has to be determined in advance and baked in the hardware. As the IPD slightly varies depending on the user, as is, the display is user dependent. In our new solution, by tracking both eyes of the user, we know their locations and thus the concerned viewing zones of the user. Consequently, we can generate the images in the viewing zones according to the tracked eyes and thus the user’s true IPD. This also makes it possible that the user can tilt her head, as long as both eyes remain in different viewing zones.

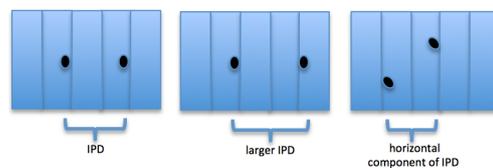


Figure 4. (left) IPD matches for the display, (middle) a user with a larger IPD, (right) tilting the head results in a smaller horizontal component of IPD.

Full parallax

The images for each viewing zone are generated as renderings of a 3D scene from different virtual camera locations, and then adjusted to compensate the distortions introduced by the optical system. In [6], each virtual camera is at the centre of each viewing zone. In this way, it is capable of displaying a discrete parallax over the viewing zones. However, in each viewing zone, the same image is displayed regardless of the exact positions of the user's eyes within the viewing zone, resulting in a noticeable transition from one viewing zone into another. Moreover, in [6], there is no vertical parallax.

Eye tracking allows us to modify the images displayed by changing the position of the virtual cameras according to the exact position of the user's eyes within the viewing zones. Consequently, we obtain a smooth, continuous horizontal parallax without the noticeable transitions between the viewing zones. Moreover, we take into account vertical parallax, as has been done before for opaque displays [14]. The full parallax is particularly important for augmented reality, where the received rendered images and the real world have to be precisely aligned for all viewpoints in the well-defined designated viewing area.

Obviously, the images of the scene have to be rendered in real-time according to the positions of the user's eyes, and as always in such eye tracking approaches, the lower the latency, the better the experience. This latency depends on the frequency of the eye tracking system, and on the time to render the images.

Use cases for augmented reality

In this section, we present some use cases of the display for augmented reality, each one putting forward a different characteristic of our approach. The first, classical use case consists in a 3D augmentation of a real scene where real objects are augmented by their virtual counterparts (Figure 1(c)). As the images are calculated according to the exact positions of the user's eyes, the real world and the virtual objects are correctly aligned for all viewpoints in the viewing area, thus providing consistency and autostereoscopy. Note that, like in any augmented reality application, occlusions have to be taken into account explicitly. If the geometry of the real scene is known, the occlusion of virtual objects by real objects is obtained by simply not rendering the occluded parts. On the other hand, managing the occlusion of real objects by virtual objects strongly depends on the relative brightness of the real scene with respect to the display. For example, in controlled lighting conditions, a projector can be used to illuminate only the parts of the scene that are not occluded [18][2].

We devised two other use cases that put forward our approach, and they concern the virtual pointing to different, real-world locations behind the transparent display. Indeed, it is very difficult for one person standing in front of a window to make another person understand a farther location that he/she is pointing at on the window, due to the different perspectives. At the moment, our employed display is too small for such a multi-user experience, but on a smaller scale, similar difficulties appear due to the different perspectives of each user's eye, and, of course, when the user moves within the designated viewing area. When the geometry of the real scene is known, the transparent multi-view display makes it possible to point at the same location visible from different perspectives at the same time.

In the first pointing use case, we use 3D arrows to point on real objects (Figure 5(a)). In the example, we point at three dice located at different distances from the user. With the consistent intensity and especially the full parallax, the autostereoscopy of our approach is highlighted. On the downside, the vergence-accommodation conflict of the employed transparent display be-

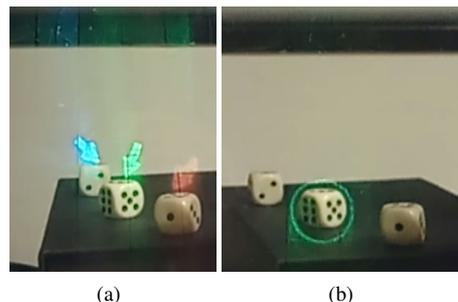


Figure 5. (a) Pointing at real objects with virtual arrows at the correct distance. (b) Encircling a real object in the plane of the display.

comes apparent as the virtual positions of the 3D arrows are behind the display, but they are shown on the display.

In the second pointing use case, we virtually encircle the real objects in the plane of the transparent display and thus at a single depth (Figure 5(b)). In this way, the distance to the virtual object (the circle), and the distance to the display, correspond. The user focuses mostly on the real object, and the circle guides the viewing direction, just like when using your finger for pointing at a distant location.

Conclusions

In this paper, we have shown how a transparent multi-view display, in combination with eye tracking, can be used to derive augmented reality applications. We presented a concrete implementation, by slightly modifying an existing display [6] and by adding a commercial eye tracker. By adjusting the rendered images, we obtain a glasses-free autostereoscopic display with a large field of view while limiting the number of projectors, yet still providing a continuous and consistent parallax.

Only a few larger viewing zones are required, and they are still sufficiently dense so that a user receives a different image in each eye. Hence, only few images have to be rendered, thus making our approach feasible for interactive applications. We have shown concrete use cases, one for a 3D augmentation of a real scene, and two for using the transparent display for virtually pointing to different, real-world locations.

We would like to point out that neither the employed multi-view display alone, without modifications and eye-tracking, nor the eye-tracking alone with a classical transparent display, could provide such an experience. The novelty is that we only require a smaller number of larger viewing zones for a larger eyebox, while still being autostereoscopic and transparent. Another advantage of our approach is that it scales well for more projectors for obtaining an even larger eyebox. Note also that the compensation of the intensity provides a solution for other applications and displays, since this is a problem that is inherent to HOEs.

Our approach inherits many characteristics of the employed transparent autostereoscopic display (e.g., the number of projectors and their resolutions, the effective display size corresponding to the HOE size, the size of the viewing zones, and the vergence-accommodation conflict) and of the eye tracking systems (e.g., frequency and accuracy). In the future, we want to run some experiments on a higher-end eye tracker. Indeed, in fast head movements, the latency of the tracker becomes apparent. Adjusting the tracking accuracy and frequency together with the number of viewpoints would also help in studying the best tradeoff between these criteria. On a longer term, we would like to test our approach on a larger scale display, especially for deriving scenarios for multiple users.

References

- [1] Hidenobu Arimoto and Bahram Javidi. Integral three-dimensional imaging with digital reconstruction. *Optics letters*, 26(3):157–159, 2001.
- [2] Giovanni Avveduto, Franco Tecchia, and Henry Fuchs. Real-world occlusion in optical see-through ar displays. In *23rd ACM Symposium on Virtual Reality Software and Technology*, pages 1–10, 2017.
- [3] Tibor Balogh, Tamás Forgács, Tibor Agács, Olivier Balet, Eric Bouvier, Fabio Bettio, Enrico Gobbetti, and Gianluigi Zanetti. A scalable hardware and software system for the holographic display of interactive graphics applications. In *Eurographics (Short Presentations)*, pages 109–112, 2005.
- [4] Oliver Bimber and Ramesh Raskar. *Spatial Augmented Reality: Merging Real and Virtual Worlds*. A. K. Peters, Ltd., 2005.
- [5] Géry Casiez, Xavier Granier, Martin Hachet, Vincent Lepetit, Guillaume Moreau, and Olivier Nannipieri. *Towards VE that are More Closely Related to the Real World*, chapter 4, pages 217–245. John Wiley & Sons, Ltd, 2018.
- [6] Thomas Crespel, Patrick Reuter, Adrian Travis, Yves Gentet, and Xavier Granier. Autostereoscopic transparent display using a wedge light guide and a holographic optical element: implementation and results. *Appl. Opt.*, 58(34):G293–G299, Dec 2019.
- [7] Osman Eldes, Kaan Akşit, and Hakan Urey. Multi-view autostereoscopic projection display using rotating screen. *Opt. Express*, 21(23):29043–29054, Nov 2013.
- [8] Chen Gao, Yifan Peng, Haifeng Li, and Xu Liu. Toward low-computation light field displays by foveated rendering. In Bernard C. Kress and Christophe Peroz, editors, *Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR) II*, volume 11765, pages 254 – 259. International Society for Optics and Photonics, SPIE, 2021.
- [9] Alex Hill, Jacob Schiefer, Jeff Wilson, Brian Davidson, Maribeth Gandy, and Blair MacIntyre. Virtual transparency: Introducing parallax view into video see-through AR. In *10th IEEE International Symposium on Mixed and Augmented Reality*, pages 239–240. IEEE, 2011.
- [10] David M Hoffman, Ahna R Girshick, Kurt Akeley, and Martin S Banks. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision*, 8(3):33–33, 2008.
- [11] Jong-Young Hong, Chang-Kun Lee, Soon gi Park, Jonghyun Kim, Kyung-Hoon Cha, Ki Hyung Kang, and ByoungHo Lee. See-through multi-view 3D display with parallax barrier. In Liang-Chy Chien, Sin-Doo Lee, and Ming Hsien Wu, editors, *Advances in Display Technologies VI*, volume 9770, pages 11 – 18. International Society for Optics and Photonics, SPIE, 2016.
- [12] Jong-Young Hong, Soon-Gi Park, Chang-Kun Lee, Seokil Moon, Sun-Je Kim, Jisoo Hong, Youngmin Kim, and ByoungHo Lee. See-through multi-projection three-dimensional display using transparent anisotropic diffuser. *Opt. Express*, 24(13):14138–14151, Jun 2016.
- [13] Tsubasa Ichikawa, Takuo Yoneyama, and Yuji Sakamoto. CGH calculation with the ray tracing method for the fourier transform optical system. *Opt. Express*, 21(26):32019–32031, Dec 2013.
- [14] Andrew V Jones, Koki Nagano, Jing Liu, Jay Busch, Xueming Yu, Mark T Bolas, and Paul Debevec. Interpolating vertical parallax for an autostereoscopic three-dimensional projector array. *Journal of Electronic Imaging*, 23(1):011005, 2014.
- [15] Sung-Kyu Kim, Seon-Kyu Yoon, and Ki-Hyuk Yoon. Crosstalk minimization in autostereoscopic multiview 3D display by eye tracking and fusion (overlapping) of viewing zones. In *Three-Dimensional Imaging, Visualization, and Display 2012*, volume 8384, page 838410. International Society for Optics and Photonics, 2012.
- [16] Jin-Ho Lee, Juyong Park, Dongkyung Nam, Seo Young Choi, Du-Sik Park, and Chang Yeong Kim. Optimal projector configuration design for 300-mpixel multi-projection 3D display. *Optics express*, 21(22):26820–26835, 2013.
- [17] A. Maimone, R. Chen, H. Fuchs, R. Raskar, and G. Wetzstein. Wide Field of View Compressive Light Field Display using a Multilayer Architecture and Tracked Viewers. *Journal of the Society for Information Display*, 2015.
- [18] Kaori Murase, Tetsuro Ogi, Kota Saito, Takahide Koyama, and Kudan Biru. Correct occlusion effect in the optical see-through immersive augmented reality display system. In *18th International Conference on Artificial Reality and Telexistence*, pages 12–19, December 2008.
- [19] Alex Olwal, Christoffer Lindfors, Jonny Gustafsson, Torsten Kjellberg, and Lars Mattsson. ASTOR: An autostereoscopic optical see-through augmented reality system. In *4th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 24–27, 2005.
- [20] Tsuyoshi Suenaga, Yoshio Matsumoto, and Tsukasa Ogasawara. 3D display based on motion parallax using non-contact 3D measurement of head position. In *17th Australia Conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future*, OZCHI '05, pages 1–4. Computer-Human Interaction Special Interest Group (CHISIG) of Australia, 2005.
- [21] Yasuhiro Takaki, Kosuke Tanaka, and Junya Nakamura. Super multi-view display with a lower resolution flat-panel display. *Optics express*, 19(5):4129–4139, 2011.
- [22] Yasuhiro Takaki and Yuta Yamaguchi. Flat-panel see-through three-dimensional display based on integral imaging. *Opt. Lett.*, 40(8):1873–1876, Apr 2015.
- [23] Cees Van Berkel. Image preparation for 3D LCD. In *Stereoscopic Displays and Virtual Reality Systems VI*, volume 3639, pages 84–91. International Society for Optics and Photonics, 1999.
- [24] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar. Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 31(4):1–11, 2012.
- [25] S. Yi, H. Chae, and S. Lee. Moving parallax barrier design for eye-tracking autostereoscopic displays. In *2008 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, pages 165–168, May 2008.