Impact of correct and simulated focus cues on perceived realism

Joseph March
Dept. of Computer Science
and Technology
University of Cambridge
United Kingdom
jgm45@cam.ac.uk

Anantha Krishnan
School of Human and
Behavioural Sciences
Bangor University
United Kingdom
a.krishnan@bangor.ac.uk

Simon J. Watt
School of Human and
Behavioural Sciences
Bangor University
United Kingdom
s.watt@bangor.ac.uk

Marek Wernikowski
Dept. of Computer Science
and Technology
University of Cambridge
United Kingdom
mw861@cam.ac.uk

Hongyun Gao
Dept. of Computer Science
and Technology
University of Cambridge
United Kingdom
hg470@cam.ac.uk

Ali Özgür Yöntem
Dept. of Computer Science
and Technology
University of Cambridge
United Kingdom
aoy20@cam.ac.uk

Rafał K. Mantiuk
Dept. of Computer Science
and Technology
University of Cambridge
United Kingdom
rafal.mantiuk@cl.cam.ac.uk

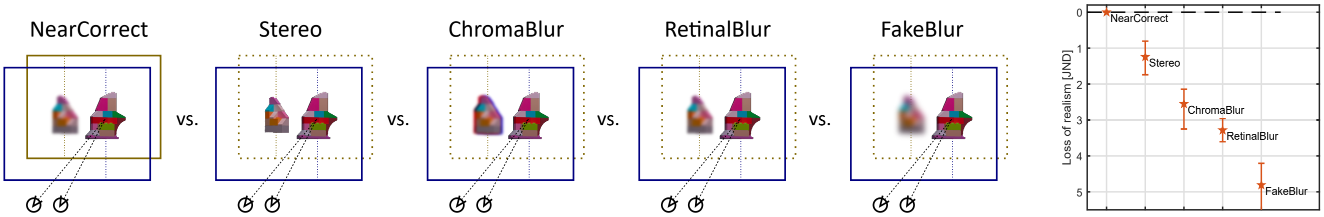

Figure 1: We compared the realism of a 3-dimensional scene reproduced with near-correct focus cues (NearCorrect)**, as a stereo image with incorrect focus cues** (Stereo)**, and as a stereo image with three types of defocus blur simulation** (ChromaBlur, RetinalBlur, FakeBlur)**. We found that focus cues have a significant effect on realism and that replacing natural defocus with depth-of-field effect simulation degrades realism instead of improving it.**

ABSTRACT

The natural accommodation of the human eye to different distances results in focus cues, which contribute to depth perception and appearance. Since focus cues are very difficult to reproduce in an electronic display, it is desirable to know how much they contribute to realistic image appearance. In this work we quantify the potential benefit of focus cues in terms of increased realism compared to regular stereo image presentation. As a secondary goal, we evaluate whether three depth-of-field rendering techniques, which reproduce defocus blur at three different degrees of accuracy, can reintroduce the benefits of focus cues. Our findings confirm the importance of focus cues for realistic image appearance, and also show that they cannot easily be substituted by depth-of-field rendering.

CCS CONCEPTS

• **Computing methodologies** → **Perception**.

KEYWORDS

computational displays, perception, visual experiments, perceptual realism, quality assessment

ACM Reference Format:
Joseph March, Anantha Krishnan, Simon J. Watt, Marek Wernikowski, Hongyun Gao, Ali Özgür Yöntem, and Rafał K. Mantiuk. 2022. Impact of correct and simulated focus cues on perceived realism. In *Proceedings of SIGGRAPH Asia.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/8888888.7777777

1 INTRODUCTION

The problem of delivering correct focus cues — blur and accommodation — has attracted much attention and motivated work on 3D display technologies, including holographic [Chakravarthula et al. 2022; Javidi et al. 2021], light-field [Huang et al. 2015; Lanman and Luebke 2013], varifocal [Akşit et al. 2017a; Konrad et al. 2016; Laffont et al. 2018] and multi-focal [Akeley et al. 2004; Chang et al. 2018; Rathinavel et al. 2018; Rolland et al. 2000] displays. It is well recognized that the lack of correct focus cues results in vergence-accommodation conflicts [Lambooij et al. 2009], which cause discomfort and fatigue [Hoffman et al. 2008; Shibata et al. 2011], reduce image quality (via diplopia, defocus blur) and affect depth perception [Watt et al. 2005]. The question we address is whether correct focus cues are also required to produce 3D imagery that appears highly realistic (i.e. close to real scenes). This

SIGGRAPH Asia, December 06–09, 2022, Daegu, South Korea

ACM ISBN 978-1-4503-1234-5/22/07…$15.00
https://doi.org/10.1145/8888888.7777777

might be expected because displaying correct focus cues will reproduce variations in retinal blur (the appearance of the scene), and induce the pattern of accommodation responses (motor output) that occur naturally with variations in depicted depth — both of which are incorrect in conventional 3D stereo. It could be, however, that those factors are barely detectable, especially in otherwise hi-fidelity images, and so have little practical importance for realism. As building a display that can deliver correct focus cues is technically challenging, and increases complexity and cost, it is important to understand the potential benefit for display quality that this additional complexity can bring.

Our secondary objective is to test whether the lack of correct focus cues can be substituted by adaptive depth-of-field (DoF) rendering. Such rendering can use eye tracking to determine the gaze point and use it to simulate realistic defocus blur that would naturally occur due to accommodation at different distances in a real-world scene. If such defocus blur simulation accounts for chromatic aberrations and compensates for natural blur from viewing the display, it can drive the accommodation mechanism, as shown by the ChromaBlur method [Cholewiak et al. 2017]. Although different aspects of DoF have been evaluated in several studies [Brooker and Sharkey 2001; Duchowski et al. 2014; Maiello et al. 2014; Mauderer et al. 2014; Zhang et al. 2015], the gain in realism has been demonstrated only for non-stereo scenes [Cholewiak et al. 2017; Hillaire et al. 2008; Mantiuk et al. 2011]. Since non-stereo images do not reproduce realistic depth, and since the effect of DoF on depth perception was found to be different for stereo and non-stereo images [Zhang et al. 2015], it is important to measure how DoF affects realism in hi-fidelity, stereoscopic scenes.

We assessed the gain in realism that results from presenting correct focus cues, while holding all other image properties constant. We used high-fidelity, high dynamic range (HDR) and steroscopic rendering of complex, real-world objects (as opposed to reduced-cue stimuli typical in vision science) so that we could determine the impact of correct focus cues while reproducing highly realistic (but isolated) objects. Together, this allowed us to determine the improvement that can be attributed to correct focus cues alone, in a practical context.

In the experiment, the observers compared a simple scene with two objects at two different depths (see Figure 1), which were rendered stereoscopically: (a) on two focal planes, delivering near-correct focus cues (NearCorrect), (b) on a single focal plane, delivering incorrect focus cues as per typical stereoscopic presentation (Stereo), (c) using the ChromaBlur method (ChromaBlur) (c) with simulated achromatic defocus blur (RetinalBlur), and (e) with an excessive (cinematic) DoF effect (FakeBlur). The realism of each rendering method was assessed in a pairwise comparison experiment. To eliminate the risk that inaccurate eye tracking could affect the results, the responses were collected both with eye tracking (free-viewing) and with controlled fixation (fixation on the near object). Our main observations and contributions are:

- Evidence that correct focus cues *do* improve the realism of stereoscopic imagery.
- The observation that focus cues cannot easily be substituted by adaptive DoF rendering, which is perceived as less realistic than even conventional stereo rendering.
- Results showing that the observers are sensitive to the type of DoF rendering and that physically accurate DoF simulation looks more realistic.

We hope that the results will further the understanding of what trade-offs are acceptable when considering the display properties required to deliver realistic content.

2 RELATED WORK

The importance of focus cues. The lack of correct focus cues introduces a vergence-accommodation conflict, which has been shown to cause discomfort and fatigue [Hoffman et al. 2008; Shibata et al. 2011], reduced binocular image quality ([Hoffman et al. 2008], and distortions in perceived depth [Watt et al. 2005]. The experiments used to show all those effects isolated focus cues by showing stereoscopic images of mostly planar stimuli on either a single or multiple focal planes. This allowed comparison of correct vs. incorrect focus cues, which we also rely on in our study. None of those works, however, used highly realistic objects or attempted to measure the influence of focus cues on realism.

Display technologies that deliver focus cues. The reproduction of focus cues on a display has attracted much research in recent years, and remains a very challenging problem. Some researchers have explored how continuous real-world variation in focus cues might be approximated in various display technologies, in conjunction with different rendering techniques. For example, work using multi-focal-planes displays has explored how to drive accommodation to intermediate distances between focal planes [MacKenzie et al. 2010], and how best to assign image intensity to focal planes to optimise image quality [Mercier et al. 2017; Narain et al. 2015]. The accuracy of focal cues can be also improved by increasing the number of focal planes, typically by temporal multiplexing [Chang et al. 2018]. This, however, increases the rendering cost, decreases the display brightness and may result in visible flicker. Correct focus cues can be produced by light-field [Huang et al. 2015; Lanman and Luebke 2013] or holographic displays [Chakravarthula et al. 2022; Javidi et al. 2021], but they require an excessively large number of addressable pixels, suffer from low spatial resolution (or field-of-view), sampling or color artifacts. There have been attempts to deliver focus cues using the configuration known as *monovision*, in which each eye is shown an image presented at different focal depth [Johnson et al. 2016; Konrad et al. 2016]. Those, however, did not show reduced fatigue, time to fuse [Johnson et al. 2016], or preference and showed only small gains in accuracy and reaction times [Konrad et al. 2016].

Vergence-accommodation conflict can be potentially eliminated in varifocal displays, in which the focal distance of a single plane is dynamically controlled, typically using adaptive optics [Akşit et al. 2017a; Konrad et al. 2016; Laffont et al. 2018]. Such a display requires low-latency, and highly accurate, eye-tracking, which is used to determine the depth of the gaze point so that the display focal plane can be dynamically adjusted to that depth. Varifocal displays do not produce correct focus cues — neither defocus blur, nor the focal depth of a 3D scene is reproduced (except for the currently displayed plane). The lack of defocus blur is typically substituted by adaptive DoF rendering, which we will discuss next.

In this work we use a multi-focal display as a way to reproduce near-correct focus cues. However, our experiment is not meant to evaluate a particular display technology but instead to assess the benefit of focus cues regardless of technology.

Defocus blur. It has been shown that defocus blur contributes to various aspects of scene perception, including perception of distance, overall scene scaling [Held et al. 2010; Vishwanath and Blaser 2010], as well being sufficient to create a sense of solid 3D space (stereopsis; [Vishwanath and Hibbard 2013]). While reproducing accommodation requires a display technology that can generate an appropriate light field, defocus blur can be potentially simulated by gaze-contigent DoF rendering. Gaze-contingent DoF in stereo presentation has been found to improve binocular fusion [Maiello et al. 2014], reduce visual discomfort [Duchowski et al. 2014] and improve performance in a task that requires depth perception [Brooker and Sharkey 2001]. Zhang et al. [2015] observed that DoF affects depth perception differently in stereo and non-stereo images. DoF has been shown to improve realism in non-stereo images [Hillaire et al. 2008; Mantiuk et al. 2011; Mauderer et al. 2014]. When DoF correctly reproduces chromatic aberrations, it can also drive accommodation and improve realism [Cholewiak et al. 2017]. Interestingly, in all those studies realism was assessed in non-stereo images, which could not reproduce realistic depth. In this work we also measure the effect of DoF rendering on realism, but use stereoscopic presentation, high-fidelity stimuli, and compare to an anchor with near-correct focus cues.

3 EXPERIMENT

The main goal of the experiment was to determine whether correct focus cues improve the realism of 3D scenes regardless of a display technology. The secondary goal was to check whether the lack of correct focus cues can be substituted by simulated DoF rendering, either approximate or physically accurate.

There is no display technology that can correctly reproduce continuous range of focus cues and, at the same time, deliver high-fidelity images (as explained Section 2). For the best approximation of both focal cues and high color fidelity, we used a multi-focal-plane display and constrained the scene so that objects were centered on the focal planes. The objects are rendered solely on the nearest focal plane, which results in focus cues that are very close to correct.

Multi-focal plane display. The experiment was conducted on a multi-focal high-dynamic-range stereo (HDRMFS) display, similar to the one used in [Zhong et al. 2021]. A simplified diagram of the display, including the locations of the focal planes, is shown in Figure 2. The display consists of two focal planes per eye, each produced by a 9.7" HDR display. The HDR displays use a combination of a 2048×1536 px LCD (LP097QX1) and a 1024×768 px DLP projector (Acer P1276). The main advantage of using HDR displays is that they introduce almost negligible black level, which otherwise can reduce image quality of a multi-focal plane display. Beam-splitters and first-surface-reflection mirrors are used to combine images from HDR displays, as shown in Figure 2. The display has the peak luminance of 4,000 cd/m^2, the black level of less than 0.01 cd/m^2. The display was calibrated to reproduce linear RGB values in the BT.709 color space and all rendering, including DoF simulation, was performed in that space.

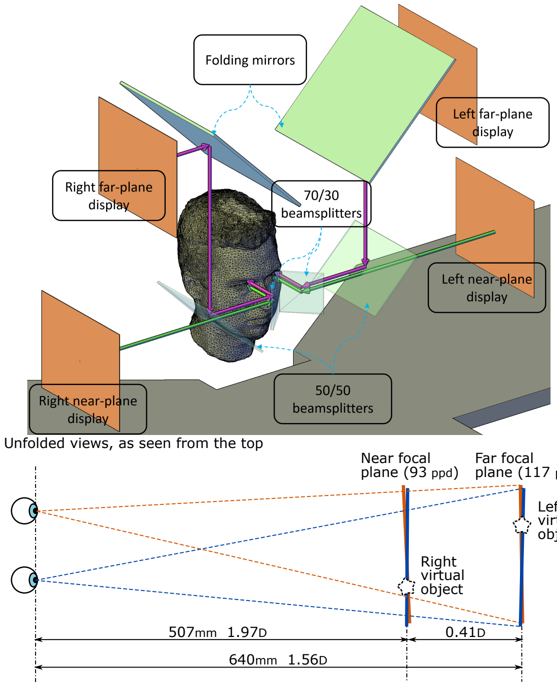

Figure 2: The schematic diagram of the multi-focal display used in our experiment. Top: the physical configuration of mirrors, beam splitters and displays. Bottom: unfolded views and viewing distances used in our experiment.

3.1 Stimuli

We wanted to use plausibly realistic objects for our study, therefore, we opted for an image-based rendering technique (lumigraphs [Gortler et al. 1996]) rather than computer graphics rendering. We also wanted to ensure accurate disparities. Therefore, the lumigraphs were rendered from the point of view of each eye, determined in a dedicated calibration procedure (explained below). The calibration also accounted for the differences in the inter-ocular distances between the observers.

In each trial, we showed a pair of objects, each rendered at a different depth, using one of the following conditions:

- NearCorrect — using both the near and the far focal planes and therefore reproducing both binocular and natural focus cues. This condition should result in (near) correct accommodation responses and defocus blur as the virtual dimensions of our presented objects lie within 25 mm of their respective focal planes.
- Stereo — using only the near plane, reproducing correct binocular cues, but incorrect focus cues for the far object.
- RetinalBlur — using only the near plane and simulating defocus blur for either near or far object, depending on the gaze direction.

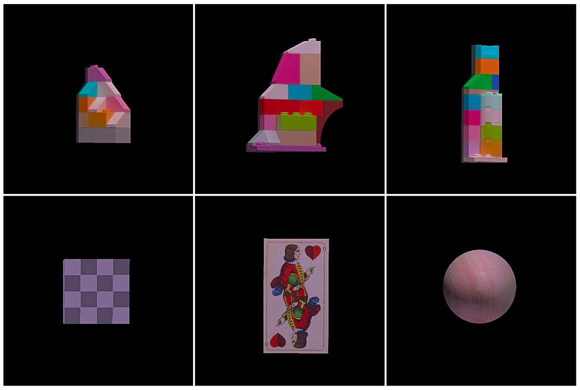

Figure 3: Renderings of the light fields used in the experiment.

- ChromaBlur — as RetinalBlur above, but the defocus blur accounts for chromatic aberrations of the eye using the method from [Cholewiak et al. 2017] and is compensated for the aberrations introduced by displayed images.
- FakeBlur — as RetinalBlur above but using an excessive amount of blur. This condition used a Gaussian blur with the standard deviation that was approximately twice the radius of the circle of confusion. The condition was meant to simulate a "cinematographic" depth-of-field effect, sometimes used in real-time rendering.

We will refer to each condition using the labels listed above. The two objects were positioned so that the center of near object lie on the near, and the center of the far object on the far display plane, as as shown in the top row of Figure 4.

Because each object was rendered on a single display plane, we were able to reproduce focus cues between the two objects but not within each object. For that reason we call our first condition NearCorrect. To minimize the focus cue inaccuracies within an object, we selected small objects with a shallow depth of up to 25 mm (0.1 D at the near and 0.06 D at far display plane). The images of the objects can be found in Figure 3. The objects were mostly built from Lego bricks, as we could easily obtain 3D models from those (using LeoCAD software[1]), which were required for lumigraph rendering. We also included a cube with a white checkerboard pattern and a playing card, both resulting in a strong chromatic aberrations in the ChromaBlur condition. The details of our light field capture rig and the photographs of the stimuli on the display can be found in the supplementary. The median luminance of each object (excluding background pixels) was set to 100 cd/m^2.

By presenting two objects in an otherwise empty environment, our experiment reduced scene complexity compared to most real-world situations, but allowed us to present focus very accurately. This trade-off is necessary to test the in-principle question of whether correct focus cues improve perceptual realism, independent of implementation-specific limitations of a given display.

DoF rendering. In all *Blur conditions, we simulate the defocus blur by preforming depth dependant filtering of the image generated by our lumigraph renderer. We filter each pixel by a blur kernel which depends on the condition type.

In both the RetinalBlur and ChromaBlur conditions, we use a cylinder function with diameter, K_d, in angles, calculated as:

$$K_\mathrm{d} = \frac{180}{\pi}\, 10^{-3}\, P\, |D_\mathrm{f} - D_\mathrm{p}|\,, \qquad (1)$$

where P is the diameter of the viewer's pupil in millimetres, D_f is the depth of the focal point in diopters and D_p is the depth of the pixel in diopters [Strasburger et al. 2018]. The pupil diameter was determined individually for each participant using the eye tracker. Note that in Equation 1, we convert from radians to degrees, and from millimetres to metres using the constants $\frac{180}{\pi}$ and 10^{-3} respectively. When filtering background (black) pixels, we set D_p to be either the depth of the near or the far plane, determined by which object the pixel is closer to. This allows us to approximate the blurred fringes of objects.

For the ChromaBlur condition, we simulate chromatic aberration using the technique outlined in [Cholewiak et al. 2017]. In our implementation, we perform defocus filtering on each color channel individually and shift the depth of the filtered pixel according to the difference between displayed and in-focus wavelengths of light. As in the original method, we approximated chromatic aberrations by superimposing the simulation for three color channels and using wavelength corresponding to the peaks of the corresponding spectral emissions (measured with spectroradiometer).

Another important distinction of ChromaBlur condition is that it compensates for the natural aberrations which the viewer's eye will introduce when viewing the image on our display. The goal is not to display retinal images (as other *Blur conditions do), but to display images that would result in a correct image on the retina. As in [Cholewiak et al. 2017], this is achieved by a deconvolution pass on the target retinal image. As such de-convolution is computationally expensive, we generate the required images offline after calibration and prior to the experiment, based on the measured pupil diameter. Our implementation used Wiener deconvolution, calculating the per channel blur kernel as in [Cholewiak et al. 2017] with the focal point is set to the near focal plane of our display. We assume additive noise with a mean of 0 and a variance of $12 \cdot 10^{-8}$.

Gaze-tracking. In order to render gaze-contingent DoF, we need to track the observer's gaze direction and estimate pupil size. We used a commercially available eye tracker (Pupil Core, PupilLabs) and its API to obtain the diameters of both pupils and their centers in camera image coordinates. However, we implemented our own calibration and mapping to the gaze direction, based on fitting multivariate polynomial functions [Duchowski 2007, ch. 14, pp. 163–166]. To estimate the pupil diameter, we used the *pye3d* model included with the PupilLabs software. The model uses the detected ellipsoid of the pupil in the image space and an estimated eye position to calculate the pupil position and diameter in the 3D space [Dierkes et al. 2018, 2019; Świrski and Dodgson 2013].

Because the arrangement of the display did not allow for tracking the pose of the head, we relied on chin- and forehead-rests to stabilize head position. To compensate for potential drift of the head orientation, the experiment allowed for a quick recalibration procedure. We artificially introduce a latency to this transition [Heron et al. 2001] in order to simulate the accommodation time

[1] LeoCAD - https://www.leocad.org/

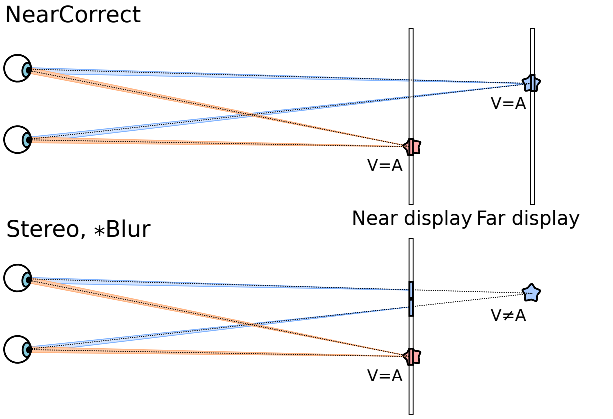

Figure 4: NearCorrect **condition was displayed on two focal planes so that vergence was consistent with the accommodation (V=A, up to the depth of our objects). All other conditions used only near display plane, introducing conflict between vergence and accommodation for the far object (V≠A).**

typical in human eyes. We do this by interpolating between two focal depths using a sigmoid function

$$\alpha = \frac{1}{e^{-30\cdot(0.3+t)}}\,, \qquad (2)$$

where α is the interpolation coefficient and t is the time from the transition start, measured in seconds. We determined the coefficients empirically in order to give a believable transition between focal depths under the viewing conditions we used in the experiment.

3.2 Experimental procedure

The experiment was split into three sessions. In the first *free-viewing* session the participants were asked to look freely at either of the two objects. Attempts to accommodate to the far object should reveal that its absolute focal distance is incorrect in all but NearCorrect conditions. The relative blur, however, should remain realistic in these conditions, to the degree that the various DoF rendering techniques are effective.

In the second *fixed-on-near* session participants were asked to look only at the near object on the right, which was reproduced at the correct focal depth. Here, the influence of the incorrect focal distance to the far object is largely removed (because no attempt is made to accommodate to it), and so any effects of focus cues should be driven almost entirely by defocus blur. We might therefore expect DoF rendering to perform closer to NearCorrect with *fixed-on-near* viewing. The second session provided a similar stimulus as a varifocal display (for all but NearCorrect condition), and eliminated the need for gaze-contingent rendering.

Because we were concerned that the observers may have moved their gaze towards the far object in the *fixed-on-near* session, we also ran a third control session for three observers. In that session we blanked the screen and displayed a fixation point at the near object's position when the eye tracker detected that the gaze was moved from that object.

For the NearCorrect condition focus cues were naturally generated by the focal planes used to display the stimuli, and eye optics, as illustrated in the top part of Figure 4. When participants looked at the far object in the *free-viewing* session, accommodation naturally caused the near object to appear blurry, and vice versa. For the Stereo and all Blur conditions, focus cues specified a single (near) plane, and so were incorrect for the far object, as shown in the bottom part of Figure 4. For all simulated blur conditions gaze tracking was used in the *free-viewing* session to determine which object should be rendered in-focus and which should be out-of-focus (or blurred). In the other two sessions (*fixed-on-near* and *fixed-on-near-and-blank*) the near object was always rendered in-focus.

We used a pairwise comparison procedure to determine the realism of each rendering condition. In each trial, the participants could switch with a key-press between the two compared conditions, shown one at a time. Both conditions contained the same pair of objects. A short 500 ms blank was introduced when switching between the conditions. The written instruction given to the participants was: *"You will be shown two scenes with two objects separated in depth and asked which of two scenes appears more realistic — in terms of the how tangible and realistic the depth between the two object looks. Note that a more realistic scene is not always a better-looking scene."*.

All possible pairs of conditions were compared (full pairwise design) and each comparison was repeated 6 times for each participant. The same pair of objects was shown for both conditions, but the pair of objects was randomized for each trial.

Calibration. Before starting the main experiment, the participants were asked to complete a quick calibration procedure, intended to determine their eye positions. They were shown two grids of vertical and horizontal lines, one on each display plane. The grids were shown to only one eye at a time. Then, participants were asked to drag with the mouse the corners of the grid on the near plane to align it with the grid on the far plane. The position of each eye was determined by finding the point closest to a bundle of lines passing through the corresponding corners of both grids in the 3D space. The grid calibration was followed by 36-gaze-point eye tracker calibration [Duchowski 2007, ch. 14, pp. 163–166].

To measure pupil diameter (needed for accurate DoF rendering), the participant was shown a sequence of different objects used in the main experiment for three minutes. The pupil diameter was measured by taking 600 samples from the eye tracker in the last minute and averaging them. The mean measured pupil diameter was 2.6±0.68 mm. Refer to the supplementary video for the recorded example session of the experiment.

Participants. 12 volunteers participated in the first session (two female), from those, 7 (all male) completed also the second session and three (all male) completed the third session. To ensure that none of the participants was presbyopic, we recruited from graduate students, aged between 23 and 28 years (median age 25 years). The participants were rewarded for their participation. All had normal stereo perception, as indicated by the Titmus stereo test. The color vision was tested Ishihara Test. One participant was a known Deuteranope (participant 10 in the supplementary). We investigated their data and decided to include it in the results as

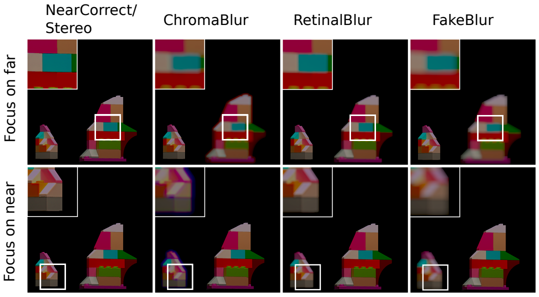

Figure 5: Cropped rendering output for regular (1st column) and simulated DoF rendering (other columns). Top: fixation at the far object. Bottom: fixation at the near object.

it did not deviate from the rest of the observers. All participants had either normal or corrected to normal vision. While our pool of participants was not demographically representative of the general population, we are not aware of any factors related to gender and age that could impact our results (except the age-related decrease in ability to accommodate). The experiment was approved by the departmental ethics board.

3.3 Results

The results of pairwise comparisons were scaled into just noticeable difference (JND) units using Bayesian maximum likelihood estimation under Thurstone's case V conditions [Perez-Ortiz and Mantiuk 2017]. The difference of 1 JND unit means that one condition is selected as more realistic than another 75% of the times. As the JND scale is relative, we set the NearCorrect condition to 0 units so that the reported scores denote the difference from that condition. The confidence intervals were computed by bootstrapping (500 samples).

The results across all participants and for both sessions are shown in Figure 6. The plots illustrate a very consistent trend of observers perceiving the Multi-focal condition as the most realistic, followed by Stereo then ChromaBlur, RetinalBlur and finally FakeBlur. All differences are statistically significant except for the differtence between RetinalBlur and FakeBlur in the *fixed-on-near-and-black* session (two-tailed Z-test on neighboring conditions). The error bars are small despite the small number of participants. This is because of the large number of repetitions (each pair compared 6 times. The results for individual participants can be found in the supplementary.

The trend of the results is similar for all three sessions. The consistency between sessions 2 and 3 (without/with blanking) suggests that the participants followed the instruction and fixated on the near object. However, fixing the gaze on the near object (*fixed-on-near*), and then adding blanking when the gaze is moved away from that object (*fixed-on-near-and-blank*) made it more difficult to discriminate between the conditions. It could be that the differences in the focus cue reproduction are the most noticeable when we allow free viewing.

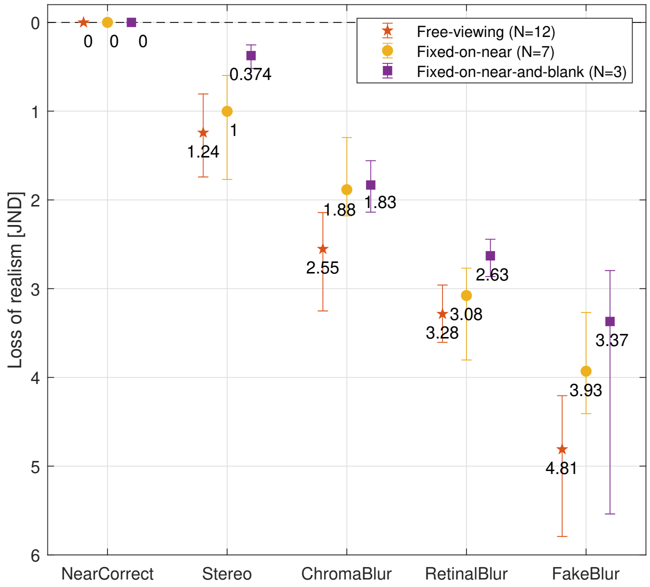

Figure 6: The loss of realism with respect to NearCorrect condition, across all three sessions of our experiment (different markers). Error bars represent 95% confidence intervals.

4 DISCUSSION

The design of a display or rendering method is often a trade-off between the cost (complexity, computation) and image quality. Realism is an important aspect of image quality, which we expect to vary less across the population than subjective preference. As the inclusion of focal depth cues adds much cost and complexity to both a display and rendering, it is important to measure how much gain in realism focus cues can bring. The three main findings from our results are as follows.

Correct focus cues improve realism. The results in Figure 6 demonstrate that NearCorrect focus cues improve perceived realism compared to the regular binocular Stereo condition. Our experiment cannot disambiguate the degree to which the gain in realism is due to correct accommodation response and/or correct defocus blur. The difference in perceived realism between the NearCorrect focus cues in comparison to the other conditions was greater in the free-viewing condition, suggesting that the accommodation response contributed to improved realism. However, this benefit could derive from sensing the accommodation response itself, or from appropriate (predictable) changes in retinal blur that result. Note that microfluctuations in accommodation during static fixation in the *fixed-on-near* session could also contribute to increased realism of the NearCorrect condition by generating correct changes in the retinal image. Whatever the underlying mechanism, over 1 JND difference shows that presentation of near-correct focus cues is judged more realistic >75% of the time. This demonstrates that the contribution of focal depth cues to realism is significant, and a highly realistic 3D display must be able to reproduce them correctly.

Accurate simulated blur is perceived as more realistic than inaccurate blur. The observers are sensitive to the type of blur used in the

DoF simulation. The exaggerated "cinematographic" blur used in the FakeBlur condition looked the least realistic. Realism improved in the RetinalBlur condition, in which the kernel simulated correct retinal blur for the measured pupil diameter. A further small gain in realism was obtained for the ChromaBlur method, which adds chromatic aberrations and compensates for natural aberrations in the eye. Our results are consistent with those of Cholewiak et al. [2017], who found that ChromaBlur rendering is judged as more realistic than achromatic blur (RetinalBlur).

Simulated blur does not improve realism over conventional stereo presentation. The previous works found that gaze-contingent DoF can improve realism over regular pinhole rendering in non-stereo presentation [Hillaire et al. 2008; Mantiuk et al. 2011]. We found the opposite effect in stereo presentation — realism was assessed to be worse instead of better than conventional stereo presentation when we introduced simulated DoF rendering. The poor performance of DoF rendering cannot be attributed to problems with eye tracking (latency, accuracy) because this result was confirmed in the second session with a fixed gaze point, which did not require eye tracking.

There are several major differences between our study and other studies that may explain the different findings. Other studies [Cholewiak et al. 2017; Hillaire et al. 2008; Mantiuk et al. 2011] used non-stereo images. The absence of binocular depth cues would be expected to substantially reduce the realism of the imagery. It is possible that effects of adding DoF blur are more pronounced in that situation (i.e. making images noticeably more 'photorealistic'). Cholewiak et al. used grayscale images, which may have enhanced the visibility of chromatic aberrations and so increased their effect. Also, the resolution and color fidelity in previous studies was lower than in ours. It is also possible that DoF rendering in those studies helped masking deficiencies of reproduced images (by blurring), causing them to be judged more realistic.

Another major difference is that other studies did not compare the presented scene to one with correct focus cues. Without a near-perfect image serving as a reference for realism, the observers could misinterpret the "realism" task and choose the condition that looked better or more interesting rather than more realistic per se. Even if the task was correctly interpreted, and the participants selected the condition they believed was more realistic, their answers may have been different if a near-perfect anchor condition was presented to them.

One limitation of both our and other DoF studies is that the defocus blur is rendered correctly only once the gaze is fixated at an object, but not when the gaze (and vergence) moves between two objects. The latency and inaccuracy of eye tracking does not allow to accurately reproduce defocus blur at every time instance. This reflects a general limitation of gaze-contingent DoF rendering. The second session with a fixed gaze eliminated this limitation, but the improvement in realism (compared to *free fixation*) was only marginal (refer to Figure 6). This suggests that even if DoF rendering simulated these aspects, it still may not match the realism of NearCorrect focus cues. Note that using rendering to achieve truly correct dynamic retinal blur (including effects of microfluctuations in accommodation) may require measuring accommodation in real time, and taking account of individual eye optics, presenting considerable technical and practical challenges. Another challenge of the methods that require deconvolution (i.e. ChromaBlur) is that they may result in negative color values, which cannot be reproduced. Because of that, a perfect simulation of retinal blur may not be physically possible.

It is important to note that simulated blur may bring other benefits, such as driving accommodation [Cholewiak et al. 2017], improving preference [Konrad et al. 2016], performance [Brooker and Sharkey 2001], the aesthetics or immersion [Hillaire et al. 2008]. Our study only informs about the decreased realism of simulated blur.

4.1 Limitations

Our current study measures the realism at a single depth separation of 0.41 D. It would be desirable to know whether the same effects can be observed at other depth separations. Although our display can accurately reproduce disparity for the continuous range of depths, it cannot reproduce focus cues for depth gradients It could be that focus cues become even more important when continuous gradients in depth, such as a ground plane, are present in the scene. Our findings could also be different if the presented objects overlap — one object was presented on the background of another. Multi-focal displays, however, cannot reproduce depth occlusions correctly [Narain et al. 2015]. This limits our ability to test whether correct focus cues improve perceived realism in more complex 3D scenes, though we are not aware of any reason why our results would not generalise to more naturalistic scenes.

5 CONCLUSIONS

Our work adds to the existing body of literature confirming the importance of focus depth cues by demonstrating that they are important for reproducing the sense of realism. The gain in realism coming from correct focus cues was statistically and practically significant (1 JND). This finding, however, should be taken with the recognition that reproducing focal cues accurately on a display is very challenging and the benefits may not compensate for the trade-off required to achieve them. For example, holographic displays are able to reproduce perfect focus cues, but their problems with color reproduction may easily degrade realism of presented scenes by more than 1 JND. We suggest that the addition of correct focus cues should not come at the expense of other important display attributes.

We also found that naturally occurring defocus blur cannot be easily substituted by synthetic DoF rendering, even when using accurate measurement of pupil size, accounting for chromatic aberrations, and for the "forward-pass" through the eye's optics. We found that synthetic blur degrades realism as compared to the stereo presentation. This could be an inherent limitation of gaze-contingent DoF methods, which cannot perfectly track accommodative state of the eye (e.g. microfluctuations), nor produce equivalent retinal images (e.g. negative values due to the deconvolution). DoF simulation is considered an important component of varifocal displays, which do not reproduce natural defocus blur, but do move the focal distance of the whole scene to match the fixation point. Our finding puts doubt on whether such DoF simulation will the bring desired benefits to those displays.

ACKNOWLEDGMENTS

We would like to thank anonymous reviewers for their valuable comments. This project has received funding from the European Union's Horizon 2020 research and innovation programme under European Research Council (ERC) grant agreement N° 725253 (EyeCode) and Marie Skłodowska-Curie grant agreement N° 76591 (RealVision). J.M. acknowledges funding from the Huawei Studentship.

REFERENCES

Kaan Akşit, Ward Lopes, Jonghyun Kim, Peter Shirley, and David Luebke. 2017a. Near-Eye Varifocal Augmented Reality Display Using See-through Screens. *ACM Trans. Graph.* 36, 6, Article 189 (Nov. 2017), 13 pages. https://doi.org/10.1145/3130800.3130892

Kaan Akşit, Ward Lopes, Jonghyun Kim, Josef Spjut, Anjul Patney, Peter Shirley, David Luebke, Steven A. Cholewiak, Pratul Srinivasan, Ren Ng, Martin S. Banks, and Gordon D. Love. 2017b. Varifocal Virtuality: A Novel Optical Layout for near-Eye Display. In *ACM SIGGRAPH 2017 Emerging Technologies* (Los Angeles, California) *(SIGGRAPH '17)*. Association for Computing Machinery, New York, NY, USA, Article 25, 2 pages. https://doi.org/10.1145/3084822.3084829

Kurt Akeley, Simon J. Watt, Ahna Reza Girshick, and Martin S. Banks. 2004. A Stereo Display Prototype with Multiple Focal Distances. *ACM Trans. Graph.* 23, 3 (aug 2004), 804–813. https://doi.org/10.1145/1015706.1015804

Julian P. Brooker and Paul M. Sharkey. 2001. Operator performance evaluation of controlled depth of field in a stereographically displayed virtual environment. In *Stereoscopic Displays and Virtual Reality Systems VIII*, Andrew J. Woods, Mark T. Bolas, John O. Merritt, and Stephen A. Benton (Eds.), Vol. 4297. 408–417. https://doi.org/10.1117/12.430841

Praneeth Chakravarthula, Ethan Tseng, Henry Fuchs, and Felix Heide. 2022. Hogel-free Holography. *ACM Transactions on Graphics* (2022). https://doi.org/10.1145/3516428

Jen-Hao Rick Chang, B. V. K. Vijaya Kumar, and Aswin C. Sankaranarayanan. 2018. Towards Multifocal Displays with Dense Focal Stacks. 37, 6, Article 198 (dec 2018), 13 pages. https://doi.org/10.1145/3272127.3275015

Steven A. Cholewiak, Gordon S. Love, Pratul P. Srinivasan, Ren Ng, and Martin S. Banks. 2017. ChromaBlur: Rendering chromatic eye aberration improves accommodation and realism. *ACM Transactions on Graphics (TOG)* 36 (2017). Issue 6. https://doi.org/10.1145/3130800.3130815

Kai Dierkes, Moritz Kassner, and Andreas Bulling. 2018. A novel approach to single camera, glint-free 3D eye model fitting including corneal refraction. In *Proc. ACM International Symposium on Eye Tracking Research and Applications (ETRA)*. 1–9. https://doi.org/10.1145/3204493.3204525

Kai Dierkes, Moritz Kassner, and Andreas Bulling. 2019. A fast approach to refraction-aware 3D eye-model fitting and gaze prediction. In *Proc. ACM International Symposium on Eye Tracking Research and Applications (ETRA)*. 1–9. https://doi.org/10.1145/3314111.3319819

Andrew Duchowski. 2007. Eye tracking methodology: Theory and practice. Springer London, Chapter 14, 163–166.

Andrew T. Duchowski, Donald H. House, Jordan Gestring, Rui I. Wang, Krzysztof Krejtz, Izabela Krejtz, Radosław Mantiuk, and Bartosz Bazyluk. 2014. Reducing visual discomfort of 3D stereoscopic displays with gaze-contingent depth-of-field. In *Proceedings of the ACM Symposium on Applied Perception*. ACM, New York, NY, USA, 39–46. https://doi.org/10.1145/2628257.2628259

Steven Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael Cohen. 1996. The Lumigraph. *Proc. of SIGGRAPH 96* 96 (08 1996). https://doi.org/10.1145/237170.237200

Robert T. Held, Emily A. Cooper, James F. O'Brien, and Martin S. Banks. 2010. Using blur to affect perceived distance and size. *ACM Transactions on Graphics* 29, 2 (mar 2010), 1–16. https://doi.org/10.1145/1731047.1731057

G Heron, W.N Charman, and C Schor. 2001. Dynamics of the accommodation response to abrupt changes in target vergence as a function of age. *Vision Research* 41, 4 (2001), 507–519. https://doi.org/10.1016/S0042-6989(00)00282-0

Sebastien Hillaire, Anatole Lecuyer, Remi Cozot, and Gery Casiez. 2008. Using an Eye-Tracking System to Improve Camera Motions and Depth-of-Field Blur Effects in Virtual Environments. In *2008 IEEE Virtual Reality Conference*. IEEE, 47–50. https://doi.org/10.1109/VR.2008.4480749

David M. Hoffman, Ahna R. Girshick, Kurt Akeley, and Martin S. Banks. 2008. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision* 8, 3 (03 2008), 33–33. https://doi.org/10.1167/8.3.33 arXiv:https://arvojournals.org/arvo/content_public/journal/jov/932853/jov-8-3-33.pdf

Fu-Chung Huang, Kevin Chen, and Gordon Wetzstein. 2015. The light field stereoscope: Immersive Computer Graphics via Factored near-Eye Light Field Displays with Focus Cues. *ACM Transactions on Graphics* 34, 4 (jul 2015), 1–12. https://doi.org/10.1145/2766922

Bahram Javidi, Artur Carnicer, Arun Anand, George Barbastathis, Wen Chen, Pietro Ferraro, J. W. Goodman, Ryoichi Horisaki, Kedar Khare, Malgorzata Kujawinska, Rainer A. Leitgeb, Pierre Marquet, Takanori Nomura, Aydogan Ozcan, YongKeun Park, Giancarlo Pedrini, Pascal Picart, Joseph Rosen, Genaro Saavedra, Natan T. Shaked, Adrian Stern, Enrique Tajahuerce, Lei Tian, Gordon Wetzstein, and Masahiro Yamaguchi. 2021. Roadmap on digital holography. *Optics Express* 29, 22 (oct 2021), 35078. https://doi.org/10.1364/OE.435915

Paul V. Johnson, Jared AQ. Parnell, Joohwan Kim, Christopher D. Saunter, Gordon D. Love, and Martin S. Banks. 2016. Dynamic lens and monovision 3D displays to improve viewer comfort. *Optics Express* 24, 11 (may 2016), 11808. https://doi.org/10.1364/OE.24.011808

Petr Kellnhofer, Piotr Didyk, Karol Myszkowski, Mohamed M Hefeeda, Hans-Peter Seidel, and Wojciech Matusik. 2016. GazeStereo3D: Seamless disparity manipulations. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 1–13.

Robert Konrad, Emily A. Cooper, and Gordon Wetzstein. 2016. Novel Optical Configurations for Virtual Reality: Evaluating User Preference and Performance with Focus-Tunable and Monovision Near-Eye Displays *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1211–1220. https://doi.org/10.1145/2858036.2858140

Robert Konrad, Nitish Padmanaban, Keenan Molner, Emily A. Cooper, and Gordon Wetzstein. 2017. Accommodation-invariant computational near-eye displays. *ACM Transactions on Graphics* 36, 4 (jul 2017), 1–12. https://doi.org/10.1145/3072959.3073594

George Alex Koulieris, George Drettakis, Douglas Cunningham, and Katerina Mania. 2016. Gaze prediction using machine learning for dynamic stereo manipulation in games. In *2016 IEEE Virtual Reality (VR)*. IEEE, 113–120. https://doi.org/10.1109/VR.2016.7504694

Pierre-Yves Laffont, Ali Hasnain, Pierre-Yves Guillemet, Samuel Wirajaya, Joe Khoo, Deng Teng, and Jean-Charles Bazin. 2018. Verifocal: A Platform for Vision Correction and Accommodation in Head-Mounted Displays *(SIGGRAPH '18)*. Association for Computing Machinery, New York, NY, USA, Article 21, 2 pages. https://doi.org/10.1145/3214907.3214925

Marc Lambooij, Marten Fortuin, Ingrid Heynderickx, and Wijnand IJsselsteijn. 2009. Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review. *Journal of Imaging Science and Technology* 53, 3 (may 2009), 30201–1–30201–14. https://doi.org/10.2352/J.ImagingSci.Technol.2009.53.3.030201

Douglas Lanman and David Luebke. 2013. Near-eye light field displays. *ACM Transactions on Graphics* 32, 6 (nov 2013), 1–10. https://doi.org/10.1145/2508363.2508366

Kevin J. MacKenzie, Ruth A. Dickson, and Simon J. Watt. 2012. Vergence and accommodation to multiple-image-plane stereoscopic displays: "real world" responses with practical image-plane separations? *Journal of Electronic Imaging* 21, 1 (2012), 1 – 9. https://doi.org/10.1117/1.JEI.21.1.011002

Kevin J MacKenzie, David M Hoffman, and Simon J Watt. 2010. Accommodation to multiple-focal-plane displays: Implications for improving stereoscopic displays and for accommodation control. *Journal of vision* 10, 8 (jan 2010), 22. https://doi.org/10.1167/10.8.22

G. Maiello, M. Chessa, F. Solari, and P. J. Bex. 2014. Simulated disparity and peripheral blur interact during binocular fusion. *Journal of Vision* 14, 8 (jul 2014), 13–13. https://doi.org/10.1167/14.8.13

Radosław Mantiuk, Bartosz Bazyluk, and Anna Tomaszewska. 2011. Gaze-dependent depth-of-field effect rendering in virtual environments. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6944 LNCS (2011), 1–12. https://doi.org/10.1007/978-3-642-23834-5_1

Michael Mauderer, Simone Conte, Miguel A. Nacenta, and Dhanraj Vishwanath. 2014. Depth perception with gaze-contingent depth of field. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 217–226. https://doi.org/10.1145/2556288.2557089

Olivier Mercier, Yusufu Sulai, Kevin Mackenzie, Marina Zannoli, James Hillis, Derek Nowrouzezahrai, and Douglas Lanman. 2017. Fast gaze-contingent optimal decompositions for multifocal displays. *ACM Transactions on Graphics* 36, 6 (nov 2017), 1–15. https://doi.org/10.1145/3130800.3130846

Rahul Narain, Rachel A Albert, Abdullah Bulbul, Gregory J Ward, Martin S Banks, and James F. O'Brien. 2015. Optimal presentation of imagery with focus cues on multi-plane displays. *ACM Transactions on Graphics* 34, 4 (jul 2015), 1–12. https://doi.org/10.1145/2766909

Maria Perez-Ortiz and Rafal K. Mantiuk. 2017. A practical guide and software for analysing pairwise comparison experiments. *arXiv preprint* (dec 2017). arXiv:1712.03686 http://arxiv.org/abs/1712.03686

Kishore Rathinavel, Hanpeng Wang, Alex Blate, and Henry Fuchs. 2018. An Extended Depth-at-Field Volumetric Near-Eye Augmented Reality Display. *IEEE Transactions on Visualization and Computer Graphics* 24, 11 (nov 2018), 2857–2866. https://doi.org/10.1109/TVCG.2018.2868570

Jannick Rolland, Myron Krueger, and Alexei Goon. 2000. Multifocal Planes Head-Mounted Displays. *Applied optics* 39 (08 2000), 3209–15. https://doi.org/10.1364/AO.39.003209

T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks. 2011. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision* 11, 8 (jul 2011), 11–11.

https://doi.org/10.1167/11.8.11

Hans Strasburger, Michael Bach, and Sven P. Heinrich. 2018. Blur Unblurred—A Mini Tutorial. *i-Perception* 9, 2 (2018), 2041669518765850. https://doi.org/10.1177/2041669518765850 PMID: 29770182.

Lech Świrski and Neil A. Dodgson. 2013. A fully-automatic, temporal approach to single camera, glint-free 3D eye model fitting [Abstract]. In *Proceedings of ECEM 2013* (Lund, Sweden). http://www.cl.cam.ac.uk/research/rainbow/projects/eyemodelfit/

D. Vishwanath and E. Blaser. 2010. Retinal blur and the perception of egocentric distance. *Journal of Vision* 10, 10 (aug 2010), 26–26. https://doi.org/10.1167/10.10.26

Dhanraj Vishwanath and Paul B Hibbard. 2013. Seeing in 3-D With Just One Eye. *Psychological Science* 24, 9 (sep 2013), 1673–1685. https://doi.org/10.1177/0956797613477867

Simon J Watt, Kurt Akeley, Marc O Ernst, and Martin S Banks. 2005. Focus cues affect perceived depth. *Journal of vision* 5, 10 (jan 2005), 834–62. https://doi.org/10.1167/5.10.7

Lei Xiao, Anton Kaplanyan, Alexander Fix, Matt Chapman, and Douglas Lanman. 2018. DeepFocus: Learned Image Synthesis for Computational Display *(SIGGRAPH '18)*. Association for Computing Machinery, New York, NY, USA, Article 4, 2 pages. https://doi.org/10.1145/3214745.3214769

Tingting Zhang, Louise O'hare, Paul B. Hibbard, Harold T. Nefs, and Ingrid Heynderickx. 2015. Depth of Field Affects Perceived Depth in Stereographs. *ACM Transactions on Applied Perception* 11, 4 (jan 2015), 1–18. https://doi.org/10.1145/2667227

Fangcheng Zhong, Akshay Jindal, Ali Özgür Yöntem, Param Hanji, Simon J. Watt, and Rafał K. Mantiuk. 2021. Reproducing Reality with a High-Dynamic-Range Multi-Focal Stereo Display. 40, 6, Article 241 (dec 2021), 14 pages. https://doi.org/10.1145/3478513.3480513