Impact of correct and simulated focus cues on perceived realism

Joseph March
Dept. of Computer Science and Technology
University of Cambridge
United Kingdom

Anantha Krishnan
School of Human and Behavioural Sciences
Bangor University
United Kingdom

Simon J. Watt
School of Human and Behavioural Sciences
Bangor University
United Kingdom

Marek Wernikowski
Dept. of Computer Science and Technology
University of Cambridge
United Kingdom

Hongyun Gao
Dept. of Computer Science and Technology
University of Cambridge
United Kingdom

Ali Özgür Yöntem
Dept. of Computer Science and Technology
University of Cambridge
United Kingdom

Rafał K. Mantiuk
Dept. of Computer Science and Technology
University of Cambridge
United Kingdom

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.

ACM Reference Format:

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGGRAPH Asia, December 06–09, 2022, Daegu, South Korea
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-1234-5/22/07...$15.00
https://doi.org/10.1145/8888888.777777

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; Lamm et al. 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; Rathiavel et al. 2018; Rolland et al. 2000] displays. It is well recognized that the lack of correct focus cues results in vergence-accommodation conflicts [Lambooj 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 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; Maurer 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; Mantuik 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 high-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 stereoscopic 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) (e) 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 optimize 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-contingent DoF rendering. Gaze-contingent DoF in
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.
- **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 observed objects.

### 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.
- **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 observed objects.
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 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 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 material. 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.

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 material. 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:

\[
K_d = \frac{180}{\pi} \cdot 10^{-3} P |D_f - D_p|, \tag{1}
\]

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 [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 py3d 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. After pressing a designated key, two vertical bars were shown: one in the display center, the other indicating the gaze direction. The participant was asked to look at the central bar and rotate their head so that both bars overlap. Vertical bars were shown instead of crosses as only horizontal gaze direction was relevant for our experiment. Note, our experiment did not require high accuracy of eye tracking since we only had to distinguish between gaze directed towards the left or right object on the display. We

---

\(^1\)LeoCAD - https://www.leocad.org/
artificially introduce a latency to this transition [Heran et al. 2001] in order to simulate the accommodation time typical in human eyes. We do this by interpolating between two focal depths using a sigmoid function

\[ \alpha = \frac{1}{1 + e^{-30 \cdot (t - t_\text{start})}} \],

where \( \alpha \) 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 plane so that vergence was consistent with NearCorrect.

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 near 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...”

![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).](image)

**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).

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 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 difference 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.

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 "cinematic" 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 greyscale 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 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 (Eye-Code) and Marie Skłodowska-Curie grant agreement N° 76591 (Re-alVision). J.M. acknowledges funding from the Huawei Studentship.

REFERENCES
