Simulated disparity and peripheral blur interact during binocular fusion

Guido Maiello

Department of Ophthalmology, Harvard Medical School, Boston, MA, USA
Department of Informatics, Bioengineering, Robotics and System Engineering, University of Genoa, Genoa, Italy
UCL Institute of Ophthalmology, University College London, London, UK

Manuela Chessa

Department of Informatics, Bioengineering, Robotics and System Engineering, University of Genoa, Genoa, Italy

Fabio Solari

Department of Informatics, Bioengineering, Robotics and System Engineering, University of Genoa, Genoa, Italy

Peter J. Bex

Department of Ophthalmology, Harvard Medical School, Boston, MA, USA

We have developed a low-cost, practical gaze-contingent display in which natural images are presented to the observer with dioptric blur and stereoscopic disparity that are dependent on the three-dimensional structure of natural scenes. Our system simulates a distribution of retinal blur and depth similar to that experienced in real-world viewing conditions by emmetropic observers. We implemented the system using light-field photographs taken with a plenoptic camera which supports digital refocusing anywhere in the images. We coupled this capability with an eye-tracking system and stereoscopic rendering. With this display, we examine how the time course of binocular fusion depends on depth cues from blur and stereoscopic disparity in naturalistic images. Our results show that disparity and peripheral blur interact to modify eye-movement behavior and facilitate binocular fusion, and the greatest benefit was gained by observers who struggled most to achieve fusion. Even though plenoptic images do not replicate an individual’s aberrations, the results demonstrate that a naturalistic distribution of depth-dependent blur may improve 3-D virtual reality, and that interruptions of this pattern (e.g., with intraocular lenses) which flatten the distribution of retinal blur may adversely affect binocular fusion.

Introduction

In natural viewing, humans continuously vary accommodation and eye position to bring into focus on the fovea an image of what is currently being fixated. For an emmetropic or refraction-corrected visual system, all objects in the visual field that are at the accommodative distance will, in first approximation, form sharp images on the retinae. Images of objects that are closer or farther than the accommodative distance will instead be out of focus, and these objects will be imaged with increasing blur with depth away from the plane of fixation (Equation 1, Figure 1). Blur can be defined as the diameter of the circle over which the point at \( Z_1 \) is imaged at the retina when the lens is focused at distance \( Z_0 \). Based on this definition, blur can be expressed as

\[
C = As \left| \frac{1}{Z_0} - \frac{1}{Z_1} \right|
\]

where \( A \) is the pupil diameter and \( s \) is the distance from the lens to the retina, i.e., the posterior nodal distance.
Figure 1. When an eye with pupil diameter $A$ and posterior nodal distance $s$ is focused at distance $Z_0$, objects at that distance will be in focus, while objects at other distances will be blurred on the retina. An object at distance $Z_1$ is blurred over a circular region with diameter $C$.

The visual experience when using manufactured entertainment devices (computers, television, books, newspapers) is very different from natural viewing conditions. While natural conditions contain a broad and continuously varying range of visual depth information, manufactured displays usually contain a narrow range of fixed depth. For example, many artificial images contain little or no blur, because they are designed for a user to be able to extract information from all areas of the image (e.g., when using a desktop application or when playing a videogame). In reading a book or using an e-reader, the pages are sharp all over, and the accommodative distance is much closer to us than in typical natural conditions, especially those outdoors (Geisler, 2008). In directed movies and television, spatial blur is often manipulated to induce the viewer to attend to specific, in-focus portions of the scene (Katz, 1991; Kosara, Miksch, & Hauser, 2001). Such conditions differ from the real world, where the ranges of depths at which our eyes accommodate produce variation in retinal blur across space and over time. Furthermore, stereoscopic 3-D is increasingly used to promote the illusion of depth, but current display technology nearly always presents depth information uncoupled from focus information. Away from fixation, however, blur can be a more precise cue to depth than binocular disparity, and the visual system appears to rely on the more informative cue when both are available (Held, Cooper, & Banks, 2012).

The presence of defocus blur has been shown to diminish visual fatigue during viewing of stereoscopic 3-D stimuli (Hoffman, Girshick, Akeley, & Banks, 2008). We were interested in simulating, via image processing, the changes in peripheral blur due to naturalistic accommodative changes. Our aim was to develop a novel method of presenting both depth and blur information in a way that simulated natural viewing conditions. We implemented a real-time gaze-contingent stereoscopic display with naturalistic distributions of blur and disparity across the retina. At fixation, the display is kept in focus, analogously to foveation and accommodation in the real world. In peripheral vision, images are presented with varying amounts of blur, increasing with distance from the simulated depth plane of fixation. To control dioptric blur, we took advantage of Light Field Rendering photographic technology (Ng, 2006; Ng et al., 2005). We employed light-field photographs of natural scenes taken with a Lytro™ plenoptic camera (Lytro Inc., Mountain View, CA; Figure 2a). Light-field cameras output for a single photographic exposure an “image stack” with images focused at different depths, a depth map of the captured scene (Figure 2b), and a depth lookup table. Although this system does not approximate the optical aberrations of an individual, it provides a general-purpose method to simulate changes in blur and depth that may be implemented in a practical, low-cost system. In this article, we examine whether any functional benefit can be obtained with such a general approach.

While a subject freely viewed such photographs, we used an eye tracker to monitor the point of fixation and we determined the depth of this location from the light-field image’s depth map. With this implementation, we then rendered the appropriate image dioptric blur for all other points in the image in real time by selecting the appropriate image from the light-field image stack. Thus, we allowed for the distribution of blur across the retina to be controlled in real time, which simulates a spatial and temporal distribution of image blur similar to that which occurs under free viewing in natural conditions. We used this display to examine how eye movements and binocular fusion depend on depth cues from blur and stereoscopic disparity in naturalistic images.

Methods

Six subjects (aged from 21 to 33 years) completed the experiment, one of whom (GM) is an author of this article. All subjects had normal or corrected-to-normal acuity and normal stereo vision. The procedure adhered to the tenets of the Declaration of Helsinki and was approved by our departmental Institutional Review Board.

The system was implemented using Psychophysics Toolbox Version 3, a free set of Matlab (MathWorks, Natick, MA) and GNU/Octave functions for vision research (Brainard, 1997; Pelli, 1997). Software was written with Matlab version R2011a interfaced with the EyeLink 1000 Desktop Mount system with a 2000-Hz camera upgrade. Stimuli were presented on a Samsung
SynchMaster 2233 LCD monitor at 120 Hz, run from an NVidia Quadro FX580 graphics processing unit that was positioned 50 cm from the observer. Display dot pitch was 0.282 mm. The spatial resolution of the stimulus images was 1080 × 1080 pixels and subtended 34° of visual angle on the screen. Stereoscopic disparity was presented via the NVIDIA 3D Vision kit. The eye tracker was accessed from the Eyelink Toolbox for Matlab and was calibrated with a supplied five-point calibration. The system latency was measured at 17.5 ± 2 ms using the methods described by Saunders and Woods (2013). The measured latency falls well within the time frame in which visual sensitivity remains reduced following a saccade (Volkmann, Riggs, White, & Moore, 1978; Volkmann, Schick, & Riggs, 1968), and thus system latency is unlikely to affect the presented results.

Plenoptic images of natural scenes

Light-field photographs of natural scenes were taken with a Lytro camera. We collected 30 photographs of natural scenes that were selected by author GM to contain a distribution of objects across a range of depths. This selection process was arbitrary but deliberately avoided panoramic or empty scenes that contained few or distant objects. The photographer positioned the camera at eye level at locations where the closest object was approximately 10 cm from the lens. Examples of 10 images are available at http://projects.iq.harvard.edu/bexlab/Lytro.

The Lytro is a low-cost plenoptic camera that allows the user to refocus on different parts of a picture after the picture has been captured. It was chosen for this work because it allows the extraction of multiple versions of the same image, each focused at a different depth plane. This allowed us to assemble a database of pictures of natural scenes where different objects could be either in focus or at an increasing level of blur with their distance in depth from the current plane of focus (Figure 2). We used open-source software (Patel, 2012) to extract data and images from the Lytro file format. This tool retrieved the component JPEG images at different focal depths, a depth map of the photographed scene, and a depth lookup table, but did not apply any image processing of the files. The depth map is outputted with a resolution of 20 × 20 pixels, which was bicubically interpolated to 1080 × 1080 pixels to match the resolution of the JPEG images. We employed this depth map, which was outputted from the Lytro proprietary software, both for our gaze-contingent implementation and for stereoscopic rendering. To validate the use of the Lytro depth maps for these purposes, we correlated the depth maps of the 30 photographs employed with the corresponding sharpness maps obtained through the focus stacking technique we describe later, and found a mean correlation coefficient of $r = 0.96$, 95% confidence interval [0.92, 0.98] (mean correlation coefficient and confidence intervals were estimated using Fisher’s Z transform). This suggests good agreement between the depth maps outputted by the Lytro proprietary software and the geometric and depth structure of the photographed images.

The image stack extracted from the Lytro data files contained from one to 12 images, depending on the depth structure of the scene. If a scene contains many objects at near depths to the camera, there will be more output images. If instead a scene is panoramic, and everything is focused at infinity, only one image will be generated.

Gaze-contingent blur implementation

For each video frame, the gaze position in screen pixel coordinates was determined from the eye tracker. The depth of the pixel at the gaze coordinate was indexed from the depth map image. The absolute difference between the value of the depth map at the gaze coordinate and the focal depth of each of the plenoptic image planes was computed. The image for which this difference was smallest was selected as the image that was most in focus at the current gaze position, and was displayed to the screen. The viewer thus always viewed an in-focus image at the center of the visual field. Objects not in the center of the viewer’s visual field but at the same focal depth as the fixation coordinates were also in focus, while everything else in

Figure 2. (a) Lytro™ plenoptic camera. After each single photographic exposure, a stack of up to 12 images was generated, each image in focus at a different depth plane. (b) An image and (bii) a registered depth map from a typical scene.
the image was optically blurred, with blur increasing with distance in depth from the plane of fixation. Movie 1 demonstrates the effect.

**Focus stacking**

To evaluate the effects of gaze-contingent blur, we compared the gaze-contingent display with a display where everything was presented in sharp focus. The stack of images extracted from a photograph taken with the Lytro was used for this purpose through focus stacking. Each of the images in the stack is in focus in different regions of the scene. To create an image that is focused everywhere, the in-focus patches in each of the images at different focal depths were located and blended together to generate a final image.

In order to identify the most focused image for each region, we assumed that the focused regions of an image would contain more sharp edges than the same region in other images (Gonzalez & Woods, 2008). A $3 \times 3$ pixel Laplacian filter was used to locate sharp edges for each image. The edge map computed for each image was then smoothed with a $35 \times 35$ pixel moving average filter to indicate the sharpness level of each pixel. Each pixel in the final stacked image was selected from the image with the highest average value at those pixel coordinates. Figure 3 shows an example of the results of focus stacking, which produced an image that is focused all over.

**Stereoscopic images**

The images extracted from the plenoptic camera are not intrinsically stereoscopic, but we used the depth map to create stereoscopic image pairs for each photograph. We created a stereoscopic image for the left and right eye by shifting individual regions of the original image by distances in proportion to the depth value given by the depth map (Figure 4). Each pixel pair from the original image was shifted $d/2$ pixels to the left and $d/2$ pixels to the right in the stereo image pair, where $d$ represents the stereoscopic disparity for the depth value of the original pixel. Occluded pixels in the farther depth plane were completed with the value of the closest unoccluded pixel to generate an abrupt depth step between foreground and background regions. These interpolated pixels were not easily detected, owing to strong crowding effects in natural scenes (Balas, Nakano, & Rosenholtz, 2009; Wallis & Bex, 2012).

Next, we specified the maximum crossed and uncrossed disparity values; for example, specifying the disparity range from 0 to 125 pixels, and given the dot pitch size of 0.282 mm, we obtain a retinal disparity ranging from $0^\circ$ to $4^\circ$ at the 50-cm viewing distance. In
Assessment of the system

To evaluate this system, we examined how eye movements and binocular fusion depend on depth cues from blur and stereoscopic disparity in natural images. We measured the time to perceptual fusion for images in which blur was gaze contingent compared to focused stacked images in which all points were in focus.

Procedure

Subjects were seated with their heads fixed in a chin and forehead rest at a distance of 50 cm from the monitor. All subjects wore NVIDIA 3-D shutter glasses. Observers were asked to freely view the images and to report when the image appeared to fuse into 3-D by pressing a button on a keyboard placed in front of them.

Figure 6 illustrates an experimental trial. Each trial began with subjects looking at a uniformly gray screen. A red fixation cross was presented at a location where the binocular disparity was greatest in the stimulus image to be presented. This ensured that before fusion was experienced, the observer experienced maximal diplopia at the fovea.

As soon as the eye-tracker data confirmed that the subject was fixating within 1.6° of the center of the cross, the stimulus image was presented. Subjects were instructed to report when the portion of the image initially cued was perceived as fused by pressing the space bar on the keyboard in front of them. A complete session consisted of 240 trials. Each of the four levels of disparity was presented 60 times in one session: In random order, 30 stimulus images contained gaze-contingent blur and 30 stimulus images were presented focused all over.

Figure 4. Geometry of disparity rendering. A point at distance \( Z_1 \) is rendered on the monitor with disparity \( d_1 \). A point at a greater distance \( Z_2 \) from the monitor is rendered on the surface of the monitor with greater disparity \( d_2 \). Objects more distant from the surface of the monitor are rendered on the surface of the monitor with greater disparities than objects that are closer in depth to the surface of the monitor.

Figure 5. Illustration of stereoscopic image pairs. (a) Original focused stacked image and depth map. (b) Stereoscopic image pair generated with the method described in the text. Some readers may be able to free fuse these images to experience a 3-D effect.
Results

All data were analyzed using a 2 (blur) × 4 (disparities) repeated-measures ANOVA. There was large intersubject variability in the time to perceptual fusion. Figure 7a shows box plots for the six observers, with mean (○), median (x), interquartile range (box), 95% tails (whiskers), and outliers (○). The raw data were therefore normalized for each subject in order to allow comparisons of the effects of blur and disparity across observers. For each observer, the mean time to perceptual fusion in each condition was divided by the mean time across all conditions. The relative time to fusion across conditions was then averaged across observers.

Figure 7b shows the relative time to perceptual fusion as a function of the disparity presented (x-axis) for gaze-contingent blur (blue) and focused (red) images. ANOVA results showed a significant main effect of disparity, $F(3, 15) = 24.68, p < 0.00001$, no significant main effect of blur, $F(1, 5) = 2.76, p = 0.16$, and a significant interaction between blur and disparity, $F(3, 15) = 3.78, p < 0.05$. Figure 7b shows a reduction in relative time to fusion at high crossed and uncrossed disparities for the gaze-contingent blur condition with respect to the focused images. Post hoc analysis showed a significant reduction in the time necessary to fuse stereoscopic images when informative distributions of retinal blur were present at high uncrossed disparities ($p < 0.05$, Bonferroni corrected t test). The size of the reduction in time to fusion at high uncrossed disparities is highly correlated ($r = 0.97, p < 0.001$) with the time to fuse high uncrossed disparities with no gaze-contingent blur. This suggests that the greatest benefits are gained by the subjects who take the longest to perceptually fuse stereoscopic images. This can be verified by visual inspection of the individual subject data presented in Appendix A.

Analysis of eye movements

Time to perceptual fusion and the number of eye movements in each condition were highly correlated ($r = 0.92, p < 10^{-20}$). As with the time-to-fusion data,
there was large intersubject variability in the number of eye movements made during each trial. Figure 8a shows box plots for the six observers, with mean ( ), median (x), interquartile range (box), 95% tails (whiskers), and outliers ( ○). The raw data were therefore normalized for each subject, as with the time to fusion data. The relative number of fixations across conditions was then averaged across observers.

Figure 8b shows the relative number of fixations that were made before the images appeared to fuse as a function of the disparity presented (x-axis) for gaze-contingent blur (blue) and focused (red) images. ANOVA results showed a significant main effect of disparity, \( F(3, 15) = 15.50, p < 0.0001 \), no significant main effect of blur, \( F(1, 5) = 6.36, p = 0.05 \), and a significant interaction between blur and disparity, \( F(3, 15) = 5.90, p < 0.01 \). Post hoc analysis showed a significant reduction in the number of fixations necessary to fuse stereoscopic images when informative distributions of retinal blur were present at high uncrossed disparities ( \( p < 0.01 \), Bonferroni corrected t test). Individual subject data are presented in Appendix B.

Fixation duration (Figure 9) was longer for gaze-contingently blurred images, \( F(1, 5) = 28.40, p < 0.01 \), but there was no significant effect of disparity, \( F(3, 15) = 0.49, p = 0.69 \), and no significant interaction between blur and disparity, \( F(3, 15) = 0.24, p = 0.86 \). In Appendix C we show that for individual subjects, disparity does seem to have an effect on fixation duration. However, the effect is very different between observers.

When attempting to fuse a stereoscopic image, the eyes need to verge to the correct depth plane. We hypothesized that vergence stability would be elevated along the horizontal direction until stereoscopic fusion is achieved. To test this hypothesis, we computed horizontal vergence stability for each fixation and compared it to vertical vergence stability. Horizontal and vertical vergence stability \( (S_h \text{ and } S_v) \) were defined as the inverse of the standard deviation of horizontal and vertical vergence, which in turn were defined as the

![Figure 8](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933549/)

![Figure 9](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933549/)
difference between the coordinates of each eye along the horizontal and vertical directions:

\[
S_H = \frac{1}{\text{std}(x_L - x_R)}, \quad S_V = \frac{1}{\text{std}(y_L - y_R)}.
\]  

Figure 10 shows median horizontal (red) and vertical (black) vergence stability, averaged across the six subjects, for the first 10 fixations in every trial.

The first fixation in each trial usually shows the lowest horizontal vergence stability (highest standard deviation in horizontal vergence). Along the horizontal direction, stability increases over the first few fixations and plateaus in latter fixations. Vertical vergence stability was substantially greater than horizontal vergence stability and remained approximately constant throughout each trial. We further investigated whether blur and disparity had any effect on horizontal vergence stability. Figure 11 shows horizontal vergence stability of the first fixation as a function of disparity for blurred (blue) or sharp (red) images.

There was a significant main effect of disparity, \( F(3, 15) = 5.45, p < 0.01 \), no significant effect of blur, \( F(1, 5) = 0.01, p = 0.95 \), and no significant interaction between blur and disparity, \( F(3, 15) = 2.76, p = 0.08 \). Horizontal vergence stability therefore increased from crossed to uncrossed disparities. Individual subject data are presented in Appendix D.

**Discussion**

The time course of perceptual fusion was significantly affected by depth away from the initial fixation plane and by the interaction between stereoscopic depth and gaze-contingent blur. Furthermore, there were systematic changes in eye-movement behavior for crossed and uncrossed disparity and the presence of optical blur in the images. The number of fixations per trial increased with either crossed or uncrossed disparity. At high levels of uncrossed disparity, viewers required less time and fewer fixations to perceptually fuse the stereoscopic images when informative distributions of retinal blur were present. This benefit was observed only at high uncrossed disparities, which is consistent with larger fusional ranges for crossed disparities than uncrossed disparities (Yeh & Silverstein, 1990). Furthermore, the largest gains were obtained by people who required the longest time to fuse the images. This indicates that a functional gain can be achieved with the inclusion of gaze-contingent blur that varies naturalistically across the visual field and that people with the greatest difficulty with stereoscopic disparity may benefit most from the addition of blur.

Our gaze-contingent blur implementation necessarily attenuated high-spatial-frequency information in the visual periphery. However, depending on the level of blur and the retinal eccentricity, this blur may not be detected (Hilz & Cavonius, 1974) or may appear to be perceptually sharp (Galvin, O’Shea, Squire, & Govan, 1997). Our results suggest that blur in the visual periphery, which reduces the contrast of geometric information at high spatial frequencies, does not hinder stereoscopic fusion. Instead, stereoscopic depth and peripheral blur interact and facilitate the perceptual fusion of stereoscopic images. These findings are in good agreement with previous research using stimuli with fixed levels of blur across the stimulus (Schor, Wood, & Ogawa, 1984; Siderov & Harwerth, 1995), showing that binocular sensory fusion is greater for low-spatial-frequency gratings. In our display, as in
natural vision, blur increases with depth disparity in the visual periphery, and we find that this facilitates perceptual fusion of high peripheral disparities.

The presence of peripheral blur also lengthens fixation duration in viewing stereoscopic stimuli, even though the foveal image remained sharp at all times. Previous researchers have demonstrated that stereoscopic images that contain natural distributions of blur are associated with reduced levels of visual fatigue (Hoffman et al., 2008). Here we show that naturally blurred images also affect eye-movement behavior, by reducing the number of fixations, without increasing the total time to fusion. Artificially sharp images elicit more fixations that are each shorter, which may be associated with visual discomfort.

Horizontal vergence stability is greater when viewing uncrossed disparities (objects behind the screen) and worsens when viewing crossed disparities (objects in front of the screen). This is consistent with reports of more visual discomfort when viewing crossed disparities (i.e., stimuli perceptually “popping out” of the screen) at near distances (Shibata, Kim, Hoffman, & Banks, 2011). Informative peripheral blur from our system does not seem to affect this trend.

Natural images have sharp depth discontinuities, and owing to microsaccades during fixation, small changes in eye position could introduce large changes in retinal blur when an observer fixates the edge of an object and the gaze drifts between near and far objects in real scenes. When observers fixate depth boundaries in our display, our system may rapidly switch between fixation planes because of eye-tracker variance in gaze estimate in addition to fixational microsaccades. Observers, however, did not report systematically noticing these rapid switches, nor do we notice this effect in natural viewing, and we found no evidence that these switches impacted observers’ experience in the present study.

The gaze-contingent display we have developed does not solve the vergence/accommodation conflict in viewing stereoscopic stimuli. Furthermore, we deliberately introduced significant cue conflicts in order to study fusion under these conditions and examine the benefit of including depth-contingent blur in virtual-reality applications. We monitored pupil size throughout the experiment and found that pupil size varied with presented disparity (the results of this analysis are presented in Appendix E). Since pupil size is linked with accommodative response (Kasthurirangan & Glasser, 2005), it is likely that observers changed accommodation as they changed convergence. Accommodation away from the surface of the display would have induced additional blur due to the eye’s own optics, and this additional source of image blur could have contributed to differences in fusion time. One method to address this issue would be to employ a volumetric display (Akeley, Watt, Girshick, & Banks, 2004; Watt, Akeley, Girshick, & Banks, 2005) to modulate accommodation as well as blur and disparity.

The blur presented in this gaze-contingent paradigm comes from the optics and image processing of the Lytro camera, which does not perfectly correspond to the blur due to the optics of a human eye. An ecological gaze-contingent model of blur can be envisioned. Assuming that an image with infinite depth of field and a calibrated depth map of the visual scene were available as starting points, the mechanism would be as follows: Portions of the scene at the focal distance would be displayed in focus. Portions of the scene away from the focal plane would be displayed filtered through a Gaussian filter with standard deviation equal to the diameter of the blur circle given by Equation 1, or with a sinc function that introduces periodic phase reversals similar to those in optically blurred images (Murray & Bex, 2010). This solution would employ a blur model which is closer to the blur introduced by the optics of the human eye than blur introduced by the optics of the Lytro camera. To implement such a system, the gaze-contingent algorithm proposed by Geisler and Perry (2002) could be used as a starting point.

Multifocal intraocular and contact lenses use diffractive optics or zones with differing refractive power to produce retinal images that are in focus for near and distance vision, thus affecting the distribution of blur across the retina (for reviews, see Bellucci, 2005; Leyland & Zinicola, 2003; Stein, 1990). The present results suggest that some observers may experience problems with binocular fusion or changes in oculomotor behavior when blur cues to depth are removed in this way, which is consistent with reported worsening of stereo acuity with multifocal contact lenses (Ferrer-Blasco & Madrid-Costa, 2010). The primary endpoint for multifocal lenses is near and distance acuity; however, changes in fusion or oculomotor behavior are not currently assessed, although they may contribute to subjective estimates of visual discomfort on quality-of-life instruments (Hays et al., 2003).

Although our display implementation does not replicate the optics of any individual observer, one of our research goals was to examine the benefit of standardized implementations of naturalistic blur in virtual-reality displays. It is not currently feasible to reproduce the individual aberrations of each observer in a consumer device. Nevertheless, we demonstrate how to approximate some of the changes in the distribution of blur and depth in a simple and inexpensive system that could be implemented in standard electronics. We show that a practical implementation of generalized dioptric blur provides measurable benefits in terms of time to fusion in a stereoscopic display. Additionally, the system we
propose enables control over multiple cues to depth, which can be carefully manipulated in real images of natural scenes.

**Conclusions**

We developed a method of presenting gaze-contingent blur in a low-cost and practical system that can be used to study vision under more realistic conditions. We used this system to study fixation stability and time to fusion. Even though the image at the fovea was always sharp as in real-world viewing and the individual aberrations of the observer were not replicated, we found that the presence of naturalistic image blur in the visual periphery facilitated fusion and modified eye movement and fixation behavior. Blur is therefore a useful cue to depth that may be used in conjunction with stereoscopic disparity to benefit depth perception in virtual-reality applications, whereas its removal with multifocal systems may impair binocular fusion.

**Keywords:** blur, disparity, binocular fusion, natural images, multifocal lenses

**Acknowledgments**

This research was supported by NIH grant R01EY019281 to PB.


**References**


Appendix A

Data on time to perceptual fusion for the six participants are shown in Figure A1. Note that y-axes have different scales.

Appendix B

Data on number of fixations for the six participants are shown in Figure B1. Note that y-axes have different scales.

Appendix C

Fixation-duration data for the six participants are shown in Figure C1. Note that y-axes have different scales. Fixation durations are longer for the gaze-contingent blur condition (blue curves) across subjects. ANOVA analysis could not identify a significant effect of disparity on fixation duration because disparity seems to have different effects on different subjects. For subjects GM, WH, AM, PC, and MT, fixation durations were longer at high disparities, although individual curves differ. For subject MD, fixation duration decreased at high crossed and uncrossed disparities.

Appendix D

Data on horizontal vergence stability for the six participants are shown in Figure D1. Note that y-axes have different scales.

Appendix E

Pupil size as a function of disparity is shown in Figure E1. Pupil size increases from crossed to uncrossed disparities. ANOVA analysis showed a significant main effect of disparity, $F(3, 15) = 16.47, p < 0.0001$, no significant effect of blur, $F(1, 5) = 0.03, p = 0.88$, and no significant interaction between blur and disparity, $F(3, 15) = 0.17, p = 0.92$. Since pupil size varies with accommodation, this is evidence that subjects were accommodating away from the surface of the display in conjunction with presented disparity.
Figure B1. Number of fixations required to fuse the stimuli as a function of disparity for all six observers. The data are the means for each condition; error bars represent ±1 SEM. The blue curves show the gaze-contingent blur condition, and the red curves show the focused-everywhere condition.

Figure C1. Fixation duration while viewing the stimuli as a function of disparity for all six observers. The data are the means for each condition; error bars represent ±1 SEM. The blue curves show the gaze-contingent blur condition, and the red curves show the focused-everywhere condition.
Figure D1. Horizontal vergence stability of the first fixation of every trial as a function of disparity for all six observers. The data are the medians for each condition; error bars represent ±1 SEM. The blue curves show the gaze-contingent blur condition, and the red curves show the focused-everywhere condition.

Figure E1. Pupil size while viewing the stimuli as a function of disparity collapsed across six observers. The data are the means for the six observers; error bars represent ±1 SEM. The blue curves show the gaze-contingent blur condition, and the red curves show the focused-everywhere condition.