High-speed switchable lens enables the development of a volumetric stereoscopic display

High-speed switchable lens enables the development of a volumetric stereoscopic display

Gordon D. Love[1,*], David M. Hoffman[2], Philip J.W. Hands[1,3], James Gao[2], Andrew K. Kirby[1] and Martin S. Banks[2]

[1]*Durham University, Dept. of Physics & Biophysical Sciences Institute, South Road, Durham, DH1 3LE, UK*
[2]*University of California at Berkeley, School of Optometry, Vision Science Program, Berkeley CA, USA*
[3]*Current address University of Cambridge, Dept. of Engineering, 9 JJ Thomson Ave., Cambridge, CB3 0FA, UK*
*g.d.love@durham.ac.uk

Abstract: Stereoscopic displays present different images to the two eyes and thereby create a compelling three-dimensional (3D) sensation. They are being developed for numerous applications including cinema, television, virtual prototyping, and medical imaging. However, stereoscopic displays cause perceptual distortions, performance decrements, and visual fatigue. These problems occur because some of the presented depth cues (i.e., perspective and binocular disparity) specify the intended 3D scene while focus cues (blur and accommodation) specify the fixed distance of the display itself. We have developed a stereoscopic display that circumvents these problems. It consists of a fast switchable lens synchronized to the display such that focus cues are nearly correct. The system has great potential for both basic vision research and display applications.

©2009 Optical Society of America

OCIS codes: (120.2040) Displays; (330.1400) Vision - binocular and stereopsis; (230.5440) Polarization-selective devices

References and links

1. M. Matsuki, H. Kani, F. Tatsugami, S. Yoshikawa, I. Narabayashi, S.-W. Lee, H. Shinohara, E. Nomura, and N. Tanigawa, "Preoperative assessment of vascular anatomy around the stomach by 3D imaging using MDCT before laparoscopy-assisted gastrectomy," AJR Am. J. Roentgenol. **183**(1), 145–151 (2004).
2. Y. Hu, and R. A. Malthaner, "The feasibility of three-dimensional displays of the thorax for preoperative planning in the surgical treatment of lung cancer," Eur. J. Cardiothorac. Surg. **31**(3), 506–511 (2007).
3. B. Mendiburu, "*3d Movie Making: Stereoscopic Digital Cinema from Script to Screen"*. Focal Press, Oxford, (2009)
4. D. C. Hutchinson, and H. W. Neal, "The design and implementation of a stereoscopic microdisplay television," IEEE Trans. Consum. Electron. **54**(2), 254–261 (2008).
5. J. P. Wann, and M. Mon-Williams, "Measurement of visual after effects following virtual environment exposure". In K.M. Stanney (Ed.), Handbook of virtual environments: Design, implementation, and applications (pp. 731–749). Hillsdale, NJ: Lawrence Erlbaum Associates (2002).
6. L. M. J. Meesters, W. A. Ijsselsteijn, and P. J. H. Seunties, "A survey of perceptual evaluations and requirements of three-dimensional TV," IEEE Trans. Circ. Syst. Video Tech. **14**(3), 381–391 (2004).
7. E. F. Fincham, and J. Walton, "The reciprocal actions of accommodation and convergence," J. Physiol. **137**(3), 488–508 (1957).
8. B. G. Cumming, and S. J. Judge, "Disparity-induced and blur-induced convergence eye movement and accommodation in the monkey," J. Neurophysiol. **55**(5), 896–914 (1986).
9. S. J. Watt, K. Akeley, M. O. Ernst, and M. S. Banks, "Focus cues affect perceived depth," J. Vis. **5**(10), 834–862 (2005).
10. D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," J. Vis. **8**(3), 1–30 (2008).
11. M. Emoto, T. Niida, and F. Okano, "Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television," Journal of Display Technology **1**(2), 328–340 (2005).
12. A. S. Percival, "The Prescribing of Spectacles". Bristol: J. Wright & Sons. (1920)
13. K. N. Ogle, T. G. Martens, and J. A. Dyer, "Oculomotor Imbalance in Binocular Vision and Fixation Disparity," London: Henry Kingdom (1967)
14. T. A. Nwodoth, and S. A. Benton, "Chidi holographic video system. In *SPIE Proceedings on Practical Holography*, **3956** (2000).
15. G. E. Favalora, J. Napoli, D. M. Hall, R. K. Dorval, M. G. Giovinco, M. J. Richmond, *et al.*, "100 million-voxel volumetric display," Proc. SPIE **712**, 300–312 (2002).
16. A. Sullivan, "DepthCube solid-state 3D volumetric display," Proc. SPIE **5291**, 279 (2004).

17. K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks, "A stereo display prototype with multiple focal distances," ACM Trans. Graph. **23**(3), 804–813 (2004).
18. F. W. Campbell, and J. G. Robson, "Application of Fourier analysis to the visibility of gratings," J. Physiol. **197**(3), 551–566 (1968).
19. F. W. Campbell, "The depth of field of the human eye," J. Mod. Opt. **4**(4), 157–164 (1957).
20. W. N. Charman, and H. Whitefoot, "Pupil diameter and depth-of-field of human eye as measured by laser speckle," Opt. Acta (Lond.) **24**, 1211–1216 (1977).
21. B. T. Schowengerdt, and E. J. Seibel, "True 3-D scanned voxel displays using single or multiple light sources," J. Soc. Inf. Disp. **14**(2), 135–143 (2006).
22. T. Shibata, T. Kawai, K. Ohta, M. Otsuki, N. Miyake, Y. Yoshihara, and T. Iwasaki, "Stereoscopic 3-D display with optical correction for the reduction of the discrepancy between accommodation and convergence," J. Soc. Inf. Disp. **13**(8), 665–671 (2005).
23. A. Shiwa, K. Omura, and F. Kishino, "Proposal for a 3-D display with accommodative compensation: 3DDAC," J. Soc. Inf. Disp. **4**(4), 255–261 (1996).
24. S. Liu, and H. Hua, "Time-multiplexed dual-focal plane head-mounted display with a liquid lens," Opt. Lett. **34**(11), 1642–1644 (2009).
25. S. Suyama, M. Date, and H. Takada, "Three-dimensional display system with dual frequency liquid crystal varifocal lens," Jpn. J. Appl. Phys. **39**(Part 1, No. 2A), 480–484 (2000).
26. Displaytech Inc, Model LV2500. *www.displaytech.com*
27. Y. Nishimoto, "Variable Focal Length Lens". US Patent. 4,783,152, Nov 8th (1988).
28. A. K. Kirby, P. J. W. Hands, and G. D. Love, "Adaptive lenses based on polarization modulation. *Proceedings of SPIE,* **6018**-14 (2005).
29. VRLogic, http://www.vrlogic.com/html/stereographics/stereographics.html
30. E. F. Canon, 50mm f/1.8 lens. http://www.usa.canon.com/consumer/controller?act=ModelInfoAct&fcategoryid=152&modelid=7306
31. A. Bradley, and I. Ohzawa, "A comparison of contrast detection and discrimination," Vision Res. **26**(6), 991–997 (1986).
32. S. Mathews, and P. B. Kruger, "Spatiotemporal transfer function of human accommodation," Vision Res. **34**(15), 1965–1980 (1994).
33. D. A. Owens, "A comparison of accommodative responsiveness and contrast sensitivity for sinusoidal gratings," Vision Res. **20**(2), 159–167 (1980).
34. E. M. Granger, and K. N. Cupery, "Optical merit function (SQF), which correlates with subjective image judgments," Photographic Science and Engineering **16**, 221–230 (1972).
35. D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," J. Opt. Soc. Am. A **4**(12), 2379–2394 (1987).

1. Introduction

Pictorial displays of three-dimensional (3D) information have widespread use in our society. Adding stereoscopic information (i.e., presenting slightly different images to the two eyes) to such displays yields a compelling 3D sensation and this has proven useful for medical imaging [1,2], cinema [3], television [4], and many other applications.

Despite the clear advantages of stereoscopic displays, there are some well-known problems [5,6]. Figure 1 illustrates the differences between viewing the real world and viewing a conventional stereoscopic display. In natural viewing, images arrive at the eyes with varying binocular disparity, so as the viewer looks from one point to another they must adjust the eyes' vergence (the angle between the lines of sight; Fig. 1a). The distance at which the lines of sight intersect is the *vergence distance*. The viewer also adjusts the focal power of the lens in each eye (i.e., accommodates) appropriately for the fixated part of the scene (i.e., where the eyes are looking). The distance to which the eye must be focused to create a sharp retinal image is the *focal distance*. Variations in focal distance create differences in image sharpness (Fig. 1c). Vergence and accommodation responses are neurally coupled: that is, changes in vergence drive changes in accommodation (vergence accommodation) and changes in accommodation drive changes in vergence (accommodative vergence) [7,8]· Vergence-accommodation coupling is advantageous in natural viewing because vergence and focal distances are nearly always identical. In conventional stereoscopic displays, images have binocular disparity thereby stimulating changes in vergence as happens in natural viewing, but the focal distance remains fixed at the display distance. Thus, the natural correlation between vergence and focal distance is disrupted (Fig. 1b,d) and this causes several problems. 1) Perceptual distortions occur due to the conflicting disparity and focus information [9]. 2) Difficulties in simultaneously fusing and focusing a stimulus occur because the viewer must now adjust vergence and accommodation to different distances [10];

if accommodation is accurate, he/she will see the object clearly, but may see double images; if vergence is accurate, the viewer will see one fused object, but it may be blurred. 3) Visual discomfort and fatigue occur as the viewer attempts to adjust vergence and accommodation appropriately [5,10,11]; Fig. 1e shows the range of vergence-accommodation conflicts that can be handled without discomfort: conflicts large enough to cause discomfort are commonplace with near viewing [12,13].
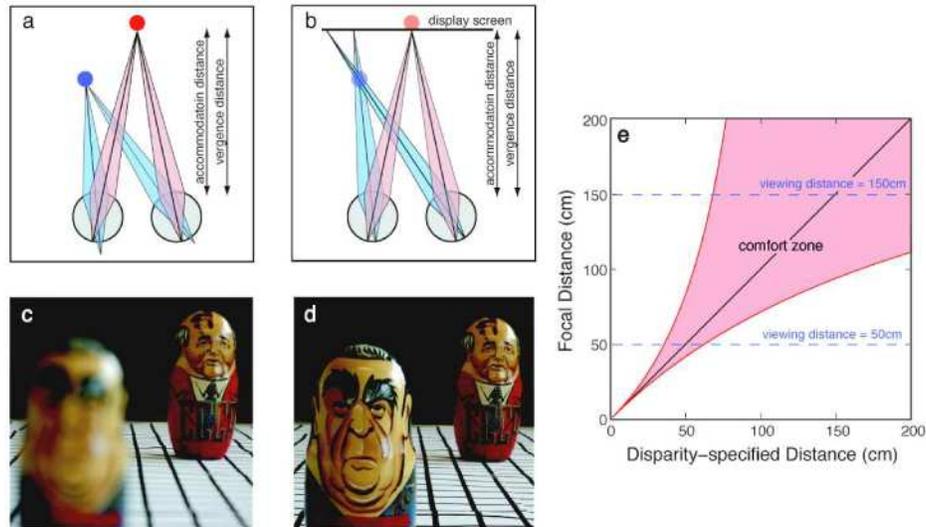

Fig. 1. Vergence distance and accommodation distance in natural viewing and with conventional stereoscopic displays. **a**) Plan view of a viewer and two objects in the natural environment. The viewer is fixating the far object and not the near object. The lines of sight to the far object intersect at the object. The distance to the intersection is the vergence distance. The distance to which the eye must be focused to form a sharp retinal image of the object is the focal distance; the distance to which the eyes are focused is the accommodation distance. **b**) Simulation of a conventional stereoscopic display of the same pair of objects. The display screen is at the same distance as the simulated far object so the vergence and focal distance of the image of the far object are the same as in **a**. However, the near object is presented on the display screen so its focal distance is no longer equal to the vergence distance, resulting in vergence-accommodation conflict and incorrect blur. **c**) Photograph of two objects like the ones depicted in **a** with the camera focused on the far object. Note the blurred image of the near object and the nearer parts of the ground plane. **d**) Photograph of two objects in which the focal distance is effectively the same as in a conventional stereoscopic display. Note the sharp image of the near object and the ground plane. **e**) Plot showing range of comfortable vergence-accommodative stimuli. The abscissa represents the simulated distance, and the ordinate represents the focal distance. Stimuli that fall within the red zone will be comfortable to fuse and focus [10,12]

Because of these problems, there has been increasing interest in creating stereoscopic displays that minimize the conflict between simulated distance cues and focus cues. Several approaches have been taken to constructing such displays, but they fall into two categories: 1) wave-front reconstructing displays and 2) volumetric displays. To date, none of these approaches are widely used due to some significant limitations. Wave-front reconstructing displays, such as holograms, present correct focus information but require extraordinary resolution, computation, and optics that make them currently impractical [14]. Volumetric displays present scene illumination as a volume of light sources and have been implemented as a swept-volume display by projecting images on to rotating display screen [15], and with a stack of liquid-crystal panels [16]. Each illumination point naturally provides correct disparity and focus cues, so the displays do not require knowledge of the viewer's gaze direction or accommodative state. However, they prevent correct handling of view-dependent lighting effects such as specular highlights and occlusions for more than a single point. Furthermore, these displays require a huge number of addressable voxels, which limits their spatial and

temporal resolution and they have a restricted workspace. By restricting the viewing position, these displays become *fixed-viewpoint volumetric displays*, which have several distinct advantages over multi-viewpoint, volumetric displays [17]. By fixing the viewpoint, the graphics engineer can separate the simulated 3D scene into a 2D projection and a depth associated with each pixel. The 2D resolution of the human visual system is approximately 50 cpd [18]; by industry standards the 2D resolution of an adequate display system is about half that value. The focal-depth resolution of the visual system is not nearly so great: viewers can under optimal conditions discriminate changes in focal distance of ~1/3D [19,20], so the focal-depth resolution of an adequate display can be relatively coarse whereas a multi-viewpoint display requires high resolution in all three dimensions. Thus the number of voxels that must be computed for a fixed-viewpoint display is a small fraction of that needed for a multiple-viewpoint display.

Presenting the light sources at different focal distances in fixed-viewpoint, volumetric displays has been done in various ways: using a deformable mirror to change the focal distance of parts of the image [21], a set of three displays combined at the viewer's eyes via beam splitters [17], a translating micro-display [22], a translating lens between the viewer and display [23], and a non-translating lens that changes focal power [24,25]. The deformable mirror is an interesting solution, but a solution based on transmissive optics is more desirable if the device is to be miniaturized to be made wearable. The translating micro-display, deformable mirror, and translating lens require mechanical movements that greatly limit the size of the workspace and the speed of changes in focal distance. In all of these designs, it would be very challenging, if not impossible, to miniaturize them sufficiently to produce a practical, wearable device.

Here we describe a fixed-viewpoint, volumetric display that represents a significant advancement. The display presents the standard 3D depth cues including disparity, occlusion and perspective in the fashion that conventional displays do, but it also presents correct or nearly correct focus cues. A stationary, switchable lens is placed in front of the eye and is synchronized to the graphic display such that each depth region in the simulated scene is presented when the lens is in the appropriate state. In this way, we construct a temporally multiplexed image with correct or nearly correct focus cues. Liu *et al.* [24] and Suyama *et al.* [25] both employed a similar approach using variable focal length lenses, but both these displays were limited by the time response of the lens. The highest frequency either of these lenses could achieve is ~50-60Hz, so with N focal states, the frame rate becomes ~50/N Hz. As we will show, a useful system requires at least four focal states, which with the liquid-lens system yields a frame rate of 12.5Hz, and this would produce quite unacceptable flicker and motion artifacts. Our system is intrinsically much faster and thereby enables the construction of a more useful, compact, flexible, and flicker-free display. Our system has the drawbacks that the user must wear active glasses (but all other systems that we are aware of require active optics), and also has the drawback that it requires the use of polarized light, so some light is wasted. However, the results in the following sections (including the video) demonstrate useful real-time stereo display that provides nearly correct focus cues. We also present data on the required number of depth planes in this type of display; the data are relevant to both our technology and other ways (referenced above) of implementing focus-correct stereo displays.

2. System information

The key technical innovation is the high-speed, switchable lens schematized in Fig. 2. The refracting element is a fixed birefringent lens. Birefringent materials have two refractive indices (ordinary and extraordinary) depending on the polarization of the incident light, so the lens has two focal lengths that are selected with a polarization modulator. If the lens is arranged such that the extraordinary axis is vertical and the ordinary axis is horizontal, incoming vertically polarized light is focused at a distance corresponding to the extraordinary refractive index.
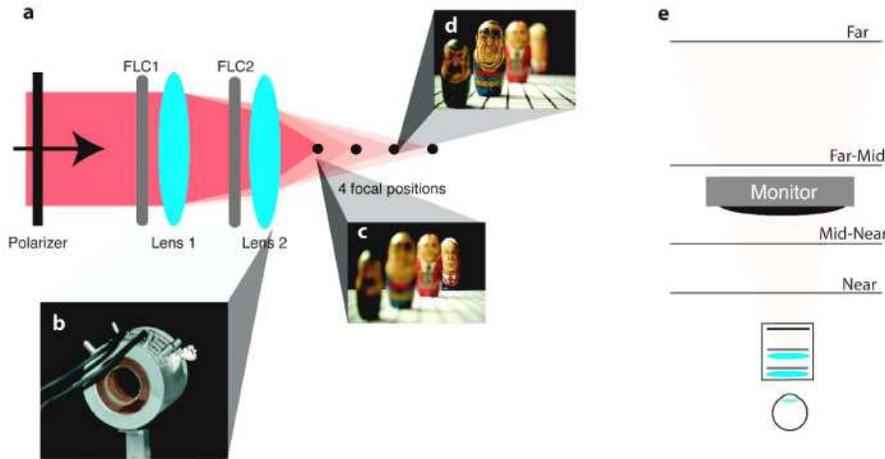
Fig. 2. The high-speed switchable lens. a) Schematic of the lens. The first polarizer produces vertically polarized light that is then either rotated, or not, through 90° by the first ferroelectric liquid-crystal polarization (FLC1) switch. The first lens focuses the two polarization states differently. A second FLC and lens produces two more possible focal lengths for each of the first polarization states creating four focal states in all. b) Photograph of the lens assembly. c and d) Images of four dolls at different distances with the lens in two different focal states. e) Overhead view of the lens, eye, and CRT. The viewing frustum is indicated by the tan shading. The focal distances associated with the four lens states are indicated by the horizontal lines.

If the light's polarization axis is rotated to horizontal before the lens, it is focused at the distance corresponding to the ordinary index. We use ferroelectric liquid-crystal modulators (FLCs) [26] to switch the polarization orientation. They act like half-wave plates whose optical axis can be rotated through ~45°. The incident polarization is therefore either aligned with or at 45° to the optical axis and hence the output polarization is rotated by either 0° or 90°. The switching between focal lengths can occur very quickly (<1ms). More focal lengths are achievable by stacking lenses and polarization modulators. With N devices, the system produces 2^N focal lengths. We have constructed a system with stacks of two devices thereby achieving four focal states. The concept of a birefringent lens was described in the patent literature [27], and realised as a single lens with two states [28]. Our novel contribution is to use a four-state system to create a stereoscopic display with nearly correct focus cues.

The lens material is calcite, which has the advantages of transparency, high birefringence (0.172) in the visible light range, and machinability. The lenses are plano-convex with a diameter of 20mm. The convex surfaces have radii of curvature of 143.3 and 286.7mm, so the four focal powers are 5.09, 5.69, 6.29, and 6.89 diopters (D), and the separations are 0.6D. A fixed glass lens (not shown) allows adjustment of the whole focal range. The number and dioptric separation of the focal states are important design features. With N focal states and average separations of Δ, the workspace is:

$$(2^N - 1)\Delta \tag{1}$$

which in our current system is 1.8D. The separations Δ can be unequal, which might be advantageous for some applications. The degree to which the system produces retinal images similar to those created in natural viewing depends critically on the dioptric separation between the focal states, which we discuss in Section 3.
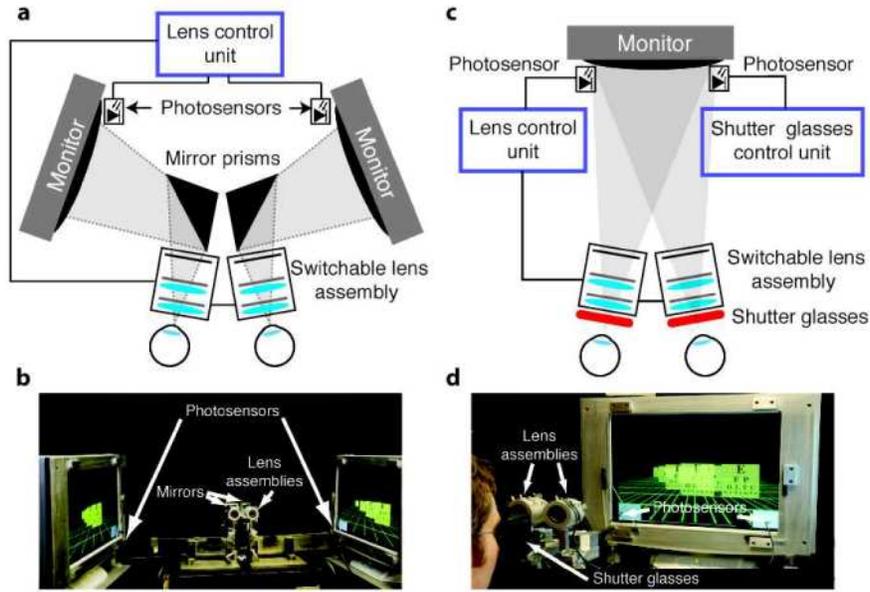
Fig. 3. Two display configurations. **a**) Schematic of the optical system with two CRTs, one for each eye, viewed through mirrors. Separate lens assemblies for each eye. With the CRT running at 180Hz, the refresh rate is 45Hz for each eye. Photosensors detect light on the CRT and through control electronics synchronize the lens assemblies to the CRTs. **b**) Photograph of the system in the two-display configuration. **c**) Schematic of the system with one CRT and two lens assemblies. Images are presented time sequentially to the two eyes by using liquid-crystal shutter glasses that alternately block and pass light to the eyes. Photosensors again detect light from the CRT to enable synchronization of the lens assemblies and shutter glasses [29] with the CRT. Eight sub-frames are presented for each volumetric frame, four per eye. With the CRT running at 180Hz, the refresh rate is therefore 22.5Hz per eye. **d**) Photograph of the display in the one-display configuration. A video of the display in action can be viewed in Fig. 5.

We have constructed two display systems: one uses two lens assemblies and one CRT and presents separate images to the two eyes in a time-sequential fashion (Fig. 3c,d); the other uses two CRTs and lens assemblies, a pair for each eye (Fig. 3a,b), and presents images simultaneously to the two eyes. Both systems are able to present all the standard cues in modern computer-graphic images, including view-dependent lighting and binocular disparity. The lens assemblies change focal state with each refresh of the CRT(s). Each image to be displayed is split into four depth zones corresponding to different ranges of distances in the simulated scene (Fig. 2e). The presentation of each of the four zones is synchronized with the lens system. Thus, when the most distant parts of the scene are displayed, the lens system is switched to its shortest focal length so that the eyes have to accommodate far to create sharp retinal images. When nearer parts are displayed, the lens system is switched to longer focal lengths so that the eye must accommodate to closer distances to create sharp images. The system thereby creates a digital approximation to the light field the eyes normally encounter in viewing 3D scenes. We do not have to know where the viewer's eye is focused to create appropriate focus cues. If the viewer accommodates far, the distant parts of the displayed scene are sharply focused on the retinas and the near parts are blurred. If the viewer accommodates near, distant parts are blurred and near parts are sharp. In this way, focus cues—blur in the retinal image and accommodation—are nearly correct.

For all but the very unlikely case that the distance of a point in the simulated scene coincides exactly with the distance of one of the focal planes, a rule is required to assign image intensities to focal planes. We use depth-weighted blending [17] in which the image intensity at each focal plane is weighted according to the dioptric distance of the point from

that plane. For an object with simulated intensity I_s at dioptric distance D_s, the image intensities I_n and I_f at the nearer and farther planes that bracket D_s are:

$$I_n = \left[1 - \frac{(D_n - D_s)}{(D_n - D_f)} \right] I_s \quad I_f = \left[\frac{(D_n - D_s)}{(D_n - D_f)} \right] I_s \tag{2}$$

where D_n and D_f are the dioptric distances of the nearer and farther planes. Thus, pixels representing an object at the dioptric midpoint between two focal planes are illuminated at half intensity on the two planes. The corresponding pixels on the two planes lie along a line of sight, so they sum in the retinal image to form an approximation of the image that would occur when viewing a real object at that distance. The sum of intensities is constant for all simulated distances (i.e., $I_s = I_n + I_f$). The depth-weighted blending algorithm is crucial to simulating continuous 3D scenes without visible discontinuities between focal planes. We evaluate the display's and algorithm's effectiveness in creating natural retinal images in Section 3.

As mentioned earlier, we have constructed two stereoscopic display systems. In both cases, the speed limitation is the display, a cathode-ray tube (CRT) running at 180Hz with 800x600 resolution. One system, shown in Fig. 3a,b, contains two CRTs and lens systems, one for each eye. Images are delivered to the eyes via front-surface mirrors. With the CRT frame rate at 180Hz, each of the four focal states is presented at 45Hz per eye. Flicker is barely visible. The other system, shown in Fig. 3c,d, uses one CRT that presents separate images to the two eyes in a time-sequential fashion. Liquid-crystal shutter glasses [29] alternatively open and block the light path to the left and right eyes in synchrony with images on the CRT. At the 180Hz frame rate, focal states are presented at 22.5Hz per eye, producing fairly noticeable flicker. Because the speed limitation is the CRT, faster display technologies, such as DLPs and OLEDs, will eventually allow higher presentation rates and more focal states without visible flicker.

3. Performance evaluation

Two important considerations are the optical quality of the switchable lens assembly and how well the assembly simulates stimuli in-between focal planes. As a rough measure of optical quality, we took still photographs and videos through the system. Figure 4 shows still photographs of Russian dolls (from near to far, they are respectively Stalin, Brezhnev, Gorbachev, and Yeltsin). The lens assembly was focused successively to each of its four focal lengths in the four pictures. The optical quality of the images is subjectively good and the blur patterns are qualitatively correct for the various focal states. A video demonstration of conventional displays and the switchable display is shown in Fig. 5.

To examine optical quality quantitatively, we measured the modulation transfer function (MTF) of the birefringent lens system. Figure 6 shows the results for the four focal states. For on-axis imaging, the MTFs of our lens are excellent: for example, transfer at 26 cpd is ~0.6-0.8 depending on focal state. The MTFs are similar to that of a high-quality commercial lens [30] that we also assessed. Our lens system has not yet been optimized to minimize field or chromatic aberration, so even better quality is attainable.

The retinal images created by a volumetric display are certainly a much closer approximation to the images created by the real world than are the images created by conventional stereoscopic displays. Indeed, a volumetric display can in principle produce retinal images that are perceptually indistinguishable from the images generated by real scenes. To see how close our display comes to achieving that, we compared the retinal images formed by viewing objects at different simulated distances in the display with the retinal images formed by viewing real objects at different distances.
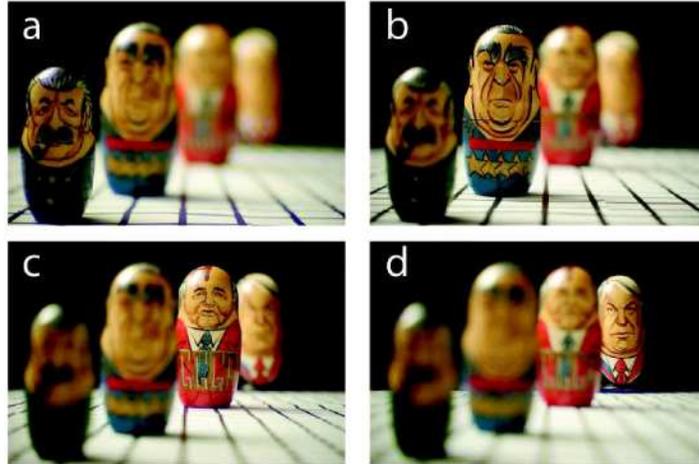
Fig. 4. Images of a real scene recorded through the switchable lens assembly. The four focal states are shown. The distances of each of the focused objects from the lens were (a) 285 mm for Stalin, (b) 375mm for Brezhnev, (c) 590mm for Gorbachev, and (d) 970mm for Yeltsin.
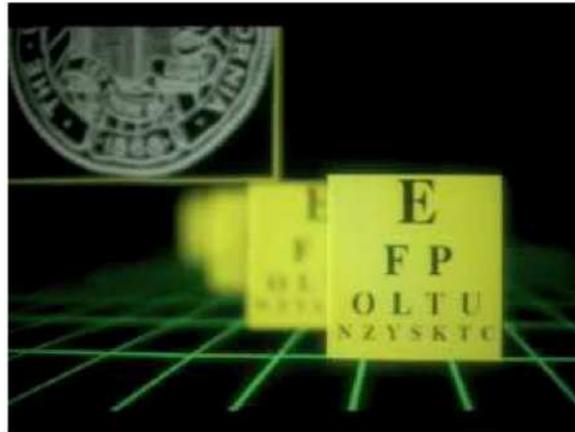

Fig. 5. (Media 1) The video shows images captured through the system with the lens assembly in each of its four focal states. The simulated scene consists of four letter-acuity charts placed on a ground plane at different distances from the viewer. The first segment of the video shows the effect of refocusing the camera used to record the video when the switchable lens is inactive and the display becomes a conventional display; in this case, all the charts are a fixed focal distance from the camera, so they go in and out of focus simultaneously. The second segment of the video shows the effect of refocusing the camera when the switchable lens is activated; the various charts go in and out of focus separately. The third and fourth segments show an object moving in depth. In the third, the switchable lens is inactivated so the object remains equally focused as it moves in depth. In the fourth, the lens is activated, so the object goes in and out of focus appropriately.
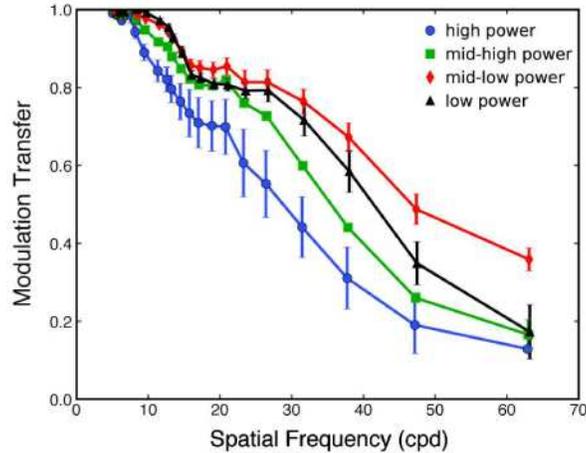
Fig. 6. Modulation transfer of the switchable lens system. Modulation transfer is plotted as a function of spatial frequency for the four focal states of the system. We printed test patterns of high-contrast square-wave gratings that ranged from 5 to 63 cpd. To measure the MTFs, we used a Canon 20D dSLR camera with a Canon 50mm, f/1.8 prime lens (set at f/4.0). The direct measurements determined the MTF of the camera plus any attenuation in the printing process. Then with the same camera, we photographed the same test patterns at the same magnification through the switchable lens system. We set the system in one of its four states and made the measurements, and then repeated this for the other focal states. The plotted MTFs are the MTFs for imaging through the switchable lens system divided by the MTFs for imaging with the camera alone. Error bars represent standard deviations. There is some variation in modulation transfer across the four focal states, but the MTFs are similar to that of a high-quality digital camera.

We implemented depth-weighted blending for these calculations. Figure 7 plots modulation transfer (retinal contrast/incident contrast) for a wide variety of situations. The dioptric separation D between focal planes is plotted on the abscissa of each panel. The simulated distance D_s of the object is plotted on the ordinate as a proportion of the distance between focal planes. The left, middle, and right columns represent the results for sinusoidal gratings of 3, 6, and 18 cpd, respectively. The upper and lower rows represent the results respectively for an aberration-free, diffraction-limited eye and for a typical human eye. We show results for both types of optics to make the point that the higher-order aberrations of the typical human eye make it easier to construct a display that produces retinal images indistinguishable from those produced by natural viewing. In each panel, color represents modulation transfer, red representing a transfer of 1 and dark blue a transfer of 0. Modulation transfer is maximized when object position is 0 or 1 because those distances correspond to cases in which the image is on one plane only and our simulation is perfect. When object position is 0.5, image intensity is distributed equally between the bracketing near and far planes, so the retinal images are approximations of the images formed in real-world viewing. The results for the typical eye show that small separation of ~0.4D are required to produced retinal-image contrast at 18 cpd that are within 30% of the retinal contrast produced in natural viewing, a value that would be perceptually indistinguishable [31]. With such small separation, the workspace would be quite constrained. For example, with two lenses (and therefore four focal states), it would be only 1.2D (Eq. (1). Fortunately, the perception of blur and control of accommodation are driven primarily by medium spatial frequencies (4-8 cpd) [32–34]. Furthermore, the contrasts of natural scenes are proportional to the reciprocal of spatial frequency [35], so such scenes contain little contrast above 4-8 cpd. Thus, the effectiveness of a volumetric display should be evaluated by the modulation transfer at 4-8 cpd. The lower middle panel in Fig. 7 shows that a plane separation of ~0.77D produces retinal-image contrasts at 6 cpd that are within 30% of real-world images and would therefore be indistinguishable [31]. With two lenses (producing four focal states), a volumetric display

with depth-weighted blending should produce an excellent approximation to the real world within a workspace of 2.3D. With three lenses (eight focal states), such a display should produce the same excellent approximation within a workspace of 5.4D. This would suffice for producing a workspace extending from 18 cm to infinity.
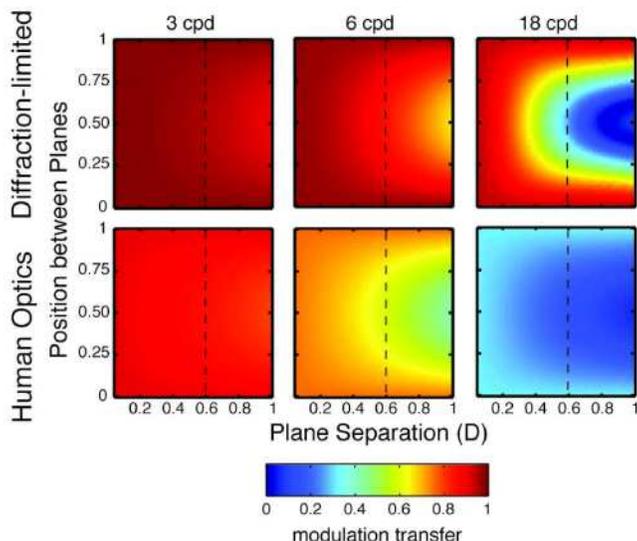

Fig. 7. Retinal-image contrast with the multi-focal display for different types of optics, spatial frequencies, and separations of the focal planes. Each panel plots plane separation in diopters on the abscissa and the position of the simulated object on the ordinate. Color (see color bar) represents modulation transfer (contrast in the retinal image divided by incident contrast). In constructing the figure, we assumed that the eye has accommodated precisely to the simulated distance. The upper row shows the calculations for diffraction-limited optics with a 4-mm aperture. The lower row shows the calculations for typical human optics (left eye of author DMH) with a 4-mm pupil; the transfer properties were determined by wave-front measurements with a Shack-Hartmann sensor. The left, middle, and right columns show the results for spatial frequencies of 3, 6, and 18 cpd, respectively.

4. Conclusions

We have developed a new stereoscopic display system that produces realistic blur to drive accommodation while also producing the appropriate cues to drive vergence and generate high-quality 3D percepts without discomfort and fatigue. The key technical development is the high-speed, switchable lens integrated with the computer display.

The system provides opportunities to conduct basic vision research while maintaining correct or nearly correct blur and accommodation. The technology also has several potentially important applications for situations like diagnostic medical imaging and surgery in which correct depth perception is critical. To realize those applications, the system would have to be miniaturized so that it could be worn like spectacles. This can be achieved by combining the birefringent lens and liquid crystals into a single unit. With head-tracking, the viewer would be free to move with respect to the display.

Acknowledgements