3D Displays

Martin S. Banks,[1] David M. Hoffman,[2] Joohwan Kim,[3] and Gordon Wetzstein[4]

[1]University of California, Berkeley, California 94720; email: martybanks@berkeley.edu

[2]Samsung, San Jose, California 95134

[3]Nvidia, Santa Clara, California 95051

[4]Stanford University, Stanford, California 94305

Annu. Rev. Vis. Sci. 2016. 2:397–435

First published online as a Review in Advance on August 15, 2016

The *Annual Review of Vision Science* is online at vision.annualreviews.org

This article's doi: 10.1146/annurev-vision-082114-035800

Keywords

display, stereo vision, visual acuity, visual resolution, depth perception, motion perception, eye movements, accommodation, light field

Abstract

Creating realistic three-dimensional (3D) experiences has been a very active area of research and development, and this article describes progress and what remains to be solved. A very active area of technical development has been to build displays that create the correct relationship between viewing parameters and triangulation depth cues: stereo, motion, and focus. Several disciplines are involved in the design, construction, evaluation, and use of 3D displays, but an understanding of human vision is crucial to this enterprise because in the end, the goal is to provide the desired perceptual experience for the viewer. In this article, we review research and development concerning displays that create 3D experiences. And we highlight areas in which further research and development is needed.

1. INTRODUCTION

Imagine a Turing test for displays. A person views input that comes either from a direct view of the real world or from a simulated view of that world presented on a display. He or she has to decide: real or display? Today's displays would fail the test miserably because everyone would be able to correctly make the distinction. Many current displays would fail because of limitations in spatial and temporal resolution. More would fail because of limitations in color reproduction and the range of displayable intensities. We focus on another critical property in which current displays fall well short: How realistic a three-dimensional (3D) experience do so-called 3D displays create? Creating realistic 3D experiences has been a very active area of research and development, and this article describes progress and what remains to be solved.

The problem of how we see in three dimensions is interesting in its own right because one dimension is lost in the projection of the natural world onto the retina. Vision scientists conceive of the 3D experience as a construction based on a variety of so-called depth cues, properties of the retinal images that signify variations in the depth dimension (Palmer 1999). We can categorize depth cues according to their cause: (*a*) depth cues based on triangulation, (*b*) cues based on perspective projection, and (*c*) cues based on light transport and reflection. The second category includes linear perspective, the texture gradient, relative size, and such. The third category includes shading, occlusion, atmospheric effects, and so forth. These cues have been well studied, and methods for presenting them are now quite advanced in computer graphics and display technology. Thus, we do not devote much attention to those two categories.

The first category of triangulation is based on viewing the world from different vantage points. Binocular disparity is the spatial differences in the images seen by the two eyes. Motion parallax is differences in images seen over time when the eye translates. Blur is defocus in images caused by light rays passing through different parts of the pupil (Held et al. 2010, 2012). Disparity, motion parallax, and blur are measureable in the retinal images, but they are also associated with specific extraretinal signals. The extraretinal component for disparity is vergence, the degree to which the lines of sight are converged. For motion parallax, the extraretinal signals include proprioceptors associated with head and body movements, kinesthesia, and vestibular signals. The extraretinal component for blur is accommodation, the change in focal power of the eye's crystalline lens. To make clear that retinal and extraretinal signals are involved, we refer to disparity and vergence collectively as stereo cues, to motion parallax and sensing self-motion as parallax cues, and to blur and accommodation as focus cues. A very active area of technical development has been to build displays that create the correct relationship between viewing parameters and these triangulation cues. The head tracking and image updating required to create high-fidelity parallax has improved greatly, but we concentrate on stereo and focus cues because our understanding of how they affect viewer experience has grown substantially in the past decade and because a number of new technologies are emerging that support those cues better than ever before.

Several disciplines are involved in the design, construction, evaluation, and use of 3D displays including materials science, electrical engineering, computer graphics, and human-factors engineering. But an understanding of human vision is also crucial to this enterprise because in the end, the goal is to provide the desired perceptual experience for the viewer.

Before categorizing different types of 3D displays, we need to get something straight. All modern displays—even conventional televisions—are 3D displays because they all employ numerous depth cues to create an impression of three dimensionality. Display types are more usefully categorized by the kind of depth cues they support. Thus, 3D displays can be broadly categorized as nonstereoscopic displays when they support only perspective-based and light-transport-based cues; as stereoscopic displays when they also support stereo cues; and as advanced displays when

they also support focus cues and in some cases parallax cues. Most of the recent work has been on stereoscopic displays, so we start there.

2. STEREOSCOPIC DISPLAYS

Stereoscopic displays are becoming increasingly important for many applications including operation of remote devices, scientific visualization, virtual prototyping, entertainment, medical imaging, surgical training, and more. A review of these applications is beyond the scope of this article, but we mention briefly some findings in medicine (Held & Hui 2011, McIntire et al. 2014a, van Beurden et al. 2012). The added depth of information afforded by stereoscopic displays has been shown to aid the detection of diagnostically relevant shapes, orientations, and positions of anatomical features, especially when monocular cues are absent or unreliable. For example, stereoscopic viewing of medical imagery significantly improves tumor detection in breast imaging and the visualization of internal structures in ultrasound. Stereoscopic displays also help surgeons orient themselves in the surgical landscape to better perform complicated tasks. In minimally invasive surgery, for instance, stereoscopic imagery decreases surgery time and increases procedure accuracy.

2.1. Techniques for Image Separation

There are two general types of stereoscopic display: those with nonoverlapping optical paths to the two eyes and those with overlapping paths (**Figure 1**).

The nonoverlapping solutions use two separate displays or two different regions on one display to present distinct images to the two eyes. Examples are head-mounted, near-eye displays (Cakmakci & Rolland 2006) and mirror haploscopes (Wheatstone 1838) (**Figure 1*a***). Because they have nonoverlapping paths, these displays guarantee complete separation of the images delivered to the two eyes. But to achieve a large enough binocular field, the displays or mirrors must be close to the eyes. The nonoverlapping solution is viable for virtual-reality and augmented-reality applications (Cakmakci & Rolland 2006) and for scientific apparatus (Backus et al. 1999), but it is not a suitable replacement for a conventional display because the solution works only for one user and requires a fixed relationship between the display and the viewer's head, which can be quite cumbersome.

The overlapping-path solution is much more common but requires a means to encode photons leaving the display so that some are visible to only the left eye and some to only the right eye. An old approach is the color-anaglyph method (Borel & Doyen 2013, Woods & Harris 2010) (subpanel *i* in **Figure 1*b***). In this method, different wavelength bands are used to encode light for the left and right eyes. The left and right images are presented on the screen in red for one eye and blue-green for the other using the display primaries. The viewer wears a red analyzer over one eye and a cyan analyzer over the other, and hence, the two eyes receive their intended images. There are numerous problems with the color-anaglyph approach. First and most important, the colors of the two delivered images are generally very different from the color of the original content, so color appearance is significantly altered. Second, by presenting different colors to the two eyes, binocular rivalry can occur, often creating a lustrous percept (Formankiewicz & Mollon 2009). Third, the spectral bandwidth of the display primaries and the acceptance band of the analyzers are generally broad, so crosstalk (incomplete image separation) often occurs (Woods et al. 2007).

Another anaglyph approach uses two sets of narrowband primaries [red, green, and blue (RGB)] and two matched notch-filter analyzers. The three bands (primaries and analyzers) are shifted in wavelength for one eye relative to the other (Jorke et al. 2009, Simon & Jorke 2011). With this

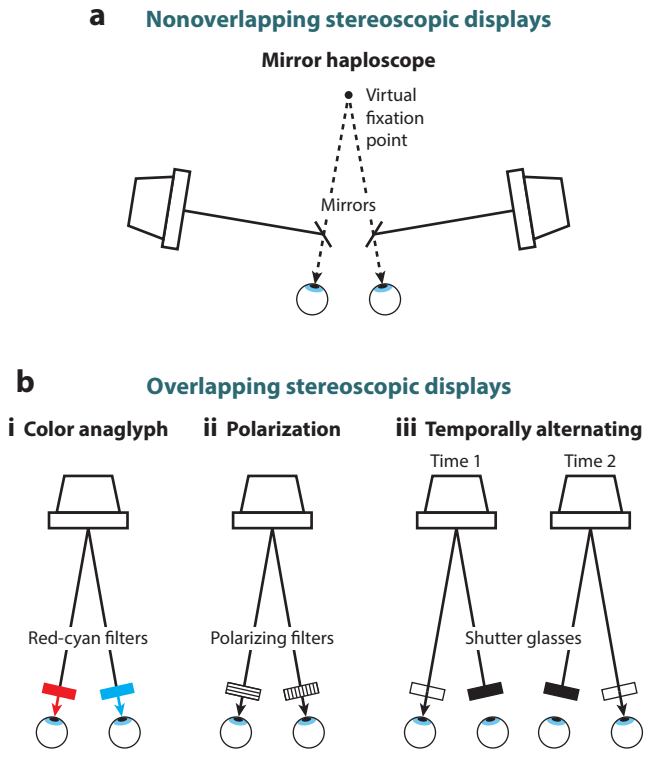

Figure 1

Techniques for image separation in stereoscopic displays. (*a*) Nonoverlapping displays have two separate optical paths, one for the left eye and one for the right. The depicted display is a mirror haploscope. The left eye sees the image on the left display screen once reflected off a mirror. The right eye sees the image on the right screen reflected off another mirror. Points illuminated in the middle of the left and right screens create a virtual point indicated by the intersection of the dashed lines. (*b*, *i*) An overlapping display in which the optical paths for the left and right eyes come from one display screen. Light rays of all three color primaries approach the left eye, but a red-transmitting filter should deliver only red rays to that eye. Likewise, a cyan-transmitting filter delivers green and blue rays to the right eye. (*b*, *ii*) An overlapping display in which image separation is created by polarization. The left eye's image is produced with one orientation of polarization and the right eye's with the orthogonal orientation. Polarized filters in front of the eyes, one orientation in front of the left eye and the orthogonal one in front of the right, deliver the desired images to the intended eyes. (*b*, *iii*) An overlapping display in which image separation is created by temporal multiplexing. Shutter glasses in front of the eyes switch between transmitting to the left eye and not the right and then transmitting to the right eye and not the left. Images intended for the left and right eyes alternate on the screen, synchronized to the switching of the glasses.

approach, both eyes see a full color image, and crosstalk is minimal. Apparent color differences can occur, but they are generally small and do not produce binocular rivalry. The disadvantages are the cost associated with making three-band notch filters, the light loss due to narrowband transmission, and the fact that color percepts across individual viewers tend to vary more with narrowband than with broadband primaries (Fairchild & Wyble 2007).

Another approach for encoding the left and right imagery uses polarization (Borel & Doyen 2013) (subpanel *ii* in **Figure 1*b***). The light from the display is orthogonally polarized for the left and right eyes, and the viewer wears analyzers such that one polarization is delivered to one eye and the orthogonal polarization to the other. Two types of polarization are used: linear and circular.

Linear polarization has the advantage that image separation is excellent, and the impact on color is minimal. It has the disadvantage that image separation depends critically on the orientation of the analyzers relative to the polarization filters at the display. If the viewer's head tilts to the side, significant crosstalk can ensue, which of course reduces image quality (Seuntiëns et al. 2005). Circular polarization eliminates this problem because orthogonality is unaffected by head tilt. But circular polarization has poorer image separation than properly aligned linear polarization and also has more effect on color.

Another approach for differentiating photons from the display involves controlling the angle at which they exit the display. Such displays are called autostereoscopic displays because they do not require glasses (Dodgson 2006). The exit angle can be manipulated by using lenticular sheets placed on the display surface or by using parallax barriers in front of or behind the display surface (Son et al. 2003). The most common complaints with autostereoscopic displays are that they have degraded resolution, have a limited sweet spot (the region in which the left and right eyes see the appropriate views), and are usable by only one viewer. But autostereoscopic displays can be constructed to support additional viewers by creating more than two viewing zones. They can also be constructed with larger sweet spots at the expense of spatial resolution. An alternative approach for expanding the sweet spot is to use a camera to track viewer head position and adjust the images on the display to move the sweet spot with the viewer (Kim et al. 2015). To minimize resolution loss, some implementations use liquid-crystal barriers that allow switching to native panel resolution for nonstereoscopic imagery (Lee & Park 2010).

2.2. Temporal- and Spatial-Interlacing Displays

Nearly all conventional stereoscopic 3D displays use temporal or spatial interlacing to present different images to the left and right eyes. Temporal interlacing delivers the left- and right-eye views alternately in time (subpanel *iii* in **Figure 1*b***). This is often accomplished either by switching views with an active element at the viewer's eye (e.g., liquid-crystal shutter glasses; Turner & Hellbaum 1986, Edwards 2009) or by switching views at the source (e.g., switching polarization at the output of a projector; Cowan 2008). The alternation must of course be synchronized with the display to ensure that images intended for the left eye are indeed seen by that eye and likewise for the right eye. Thus, in temporal-interlacing displays, only one eye receives light at a given time, but it receives all the pixels. Spatial interlacing delivers one set of pixels to the left eye and another to the right eye simultaneously. In most implementations, even pixel rows go to one eye and odd rows to the other (Dawson 2012). This is typically done using a film-patterned retarder on the display that polarizes the emitted light in opposite directions row by row. The viewer wears passive eyewear that transmits alternate rows to the two eyes. In this way, both eyes receive light at any given moment, but each receives only half the pixels.

Temporal interlacing and spatial interlacing have different shortcomings from a perceptual standpoint. Temporal interlacing is prone to temporal artifacts such as flicker, unsmooth motion appearance, and distortions in the perceived depth of moving objects. Spatial interlacing results in lower spatial resolution and some distortions of perceived depth.

2.3. Crosstalk and Ghosting

Crosstalk is the incomplete segregation of the two eyes' images. It is defined quantitatively as the proportion of the light delivered to the intended eye that leaks into the unintended eye. Nearly all commercial stereoscopic display systems have some crosstalk (Blondé et al. 2011, Woods 2012).

Ghosting is perceived crosstalk. It can occur when as little as 1% of the light leaks into the unintended eye (Nojiri et al. 2004, Pastoor 1993).

The contents of the image clearly affect ghosting: It becomes more apparent with higher contrast, sharper edges, and larger disparities. Ghosting affects perceived image quality: Reported quality decreases in rough proportion to the amount of crosstalk (Seuntiëns et al. 2005, Wilcox & Stewart 2003). Crosstalk also affects depth percepts: The amount of perceived depth decreases monotonically with increasing crosstalk (Tsirlin et al. 2011).

2.4. Capture-Display-Viewing Pipeline

For the viewer of a stereoscopic display to accurately perceive the geometry of the 3D scene from the binocular images, the pipeline from the generation of the binocular images (captured via cameras or created via computer graphics) to the viewing of the displayed images must be appropriate; in other words, the capture, display, and viewing parameters must be compatible. Errors in the pipeline generally lead to distortions of perceived depth relative to the original scene and to reduced visual quality.

The simplest approach for producing 3D imagery is to capture a scene with a stereo camera (a camera system with two positions corresponding to the two eye positions) and then display the images from the left and right cameras to the viewer's left and right eyes respectively on one or two displays (**Figure 1**). We ignore the camera distortions by assuming pinhole apertures (although we explore the importance of camera bokeh in Section 3.1). Camera and display pipelines include several other factors that influence image quality, but we ignore these for now because they have little unique bearing on 3D displays. For more information on such issues, see Brainard et al. (2002) and Wandell & Silverstein (2003).

In any application based on perspective projection (e.g., photographs, cinematography, computer graphics, realistic paintings), all light rays from a captured or generated image pass through the center of projection (COP) (Kubovy 1986). Correct stereoscopic imagery has two COPs, one for the left eye and one for the right. If care is taken in the camera configuration, the display setup, and the viewer's position, the viewer's eyes will be located at the respective COPs and have the correct alignment so he or she will receive the same retinal images and eye-position signals that would have been received from the original scene. If these constraints are obeyed, the percepts should be quite similar. In practice, creating the appropriate pipeline to reproduce veridical binocular retinal images and eye-position signals is challenging. When the pipeline is altered, there may not be consistent COPs for the left- and right-eye images; there may be consistent COPs but the viewer's eyes are not located at those points; and/or the horizontal vergence required to fixate points in the scene may be inappropriate. In each case, predictable distortions in perceived depth ensue, some more important than others.

All angular relationships captured by the camera system should be faithfully reproduced by the display system. Camera adjustments such as zooming and cropping that are acceptable in 2D imagery can be very disruptive in 3D imagery if not done in a consistent manner. The two main parameters of a camera are focal length and dimensions of the sensor plane. The size of the image in the sensor plane is determined by focal length, and the field of view is determined by the tangent of focal length/sensor dimension, which can be changed with cropping. In the display, there are two analogous parameters of viewing distance and screen size. A veridical percept requires matching the focal length/sensor size to the viewing distance/display size. Additionally, one must ensure that the optical axes of the cameras are appropriate for the manner in which the images are displayed. For example, parallel camera axes (coplanar sensors) are appropriate for overlapping stereoscopic displays (subpanels *i–iii* in **Figure 1*b***), whereas converging axes are

appropriate for nonoverlapping stereoscopic displays arranged as in **Figure 1*a***. When the camera-display parameters are inappropriate (e.g., converging axes with an overlapping display), geometric perspective correction is required to guarantee that correct disparities are presented to the viewer's eyes (Held & Banks 2008).

When there are multiple viewers, a veridical pipeline is not possible with conventional technology (although with some high–frame-rate displays, time multiplexing for multiple viewers is possible; Hoffman et al. 2014). When there are incompatibilities in the pipeline, a ray-intersection model is typically used to determine the perceptual distortions that are likely to occur. In the model, two rays propagate from the eyes through corresponding points in the stereoscopic display and beyond. The rays intersect at a location in space that should correspond with the location of the object that created the two disparate image points. When the rays do not intersect (because of violations of epipolar geometry), the model finds the point of closest passage and predicts perceived depth based on the 3D location of such points. With this model, one can calculate the retinal disparities for a wide range of capture, display, and viewing parameters (Held & Banks 2008, Pollock et al. 2012, Woods et al. 1993). When the disparities are not the same as those that would be generated by the original scene, the perceptual experience is usually a distortion of the 3D structure of that scene because viewers are generally unable to compensate for an incorrect viewing position when viewing stereoscopic imagery (Banks et al. 2009, Bereby-Meyer et al. 1999, Hands et al. 2015). This contrasts with the viewing of nonstereoscopic imagery whereby viewers are able to compensate impressively for incorrect viewing position (Bereby-Meyer et al. 1999, Hands et al. 2015, Vishwanath et al. 2005).

Figure 2 summarizes some of the distortions that occur when stereo imagery is viewed from various positions. In panel *c*, the capture-display-viewing parameters are all appropriate, so the original scene (an open-book hinge) should be perceived veridically. In panels *b* and *d*, the viewer is too far to the left and too far to the right, respectively, and the scene is perceived as skewed respectively to the left and right. In panels *a* and *e*, the viewer is too close and too far, respectively, and the scene is seen respectively as compressed and expanded in depth. Other errors in the pipeline (converged or so-called toed-in cameras, larger or smaller separation between the cameras, etc.) also lead to predictable distortions of perceived depth, such as unintended curvature and changes in apparent size (Held & Banks 2008, Woods et al. 1993).

An additional problem in the capture-display-viewing pipeline is optical distortion in the stereo cameras. With simple lenses, the focal distance for axial rays can be different than for oblique rays. This leads to barrel distortion (center of image swells) or pincushion distortion (corners are stretched from image center). These distortions often affect corresponding parts of the left- and right-camera images differently, so they can create unintended disparities, and therefore unintended distortions in the perceived 3D scene (Woods et al. 1993). Distortions also occur in the lenses used in near-eye displays. This problem can be serious because such displays require a wide field of view. The optical distortions can, however, be corrected in software provided that the lens properties and positions of the eyes relative to the screens are known with sufficient accuracy.

2.5. Head Roll

In properly constructed stereoscopic media for presentation on one screen, all disparities are horizontal on the screen. To generate the same retinal images as would be received when viewing the original scene, the viewer's eyes must be placed at the respective COPs, separated by the interocular distance. When the eyes are so positioned, the horizontal and vertical disparities at the retinas are identical to those that would be created by looking at the original 3D scene. But

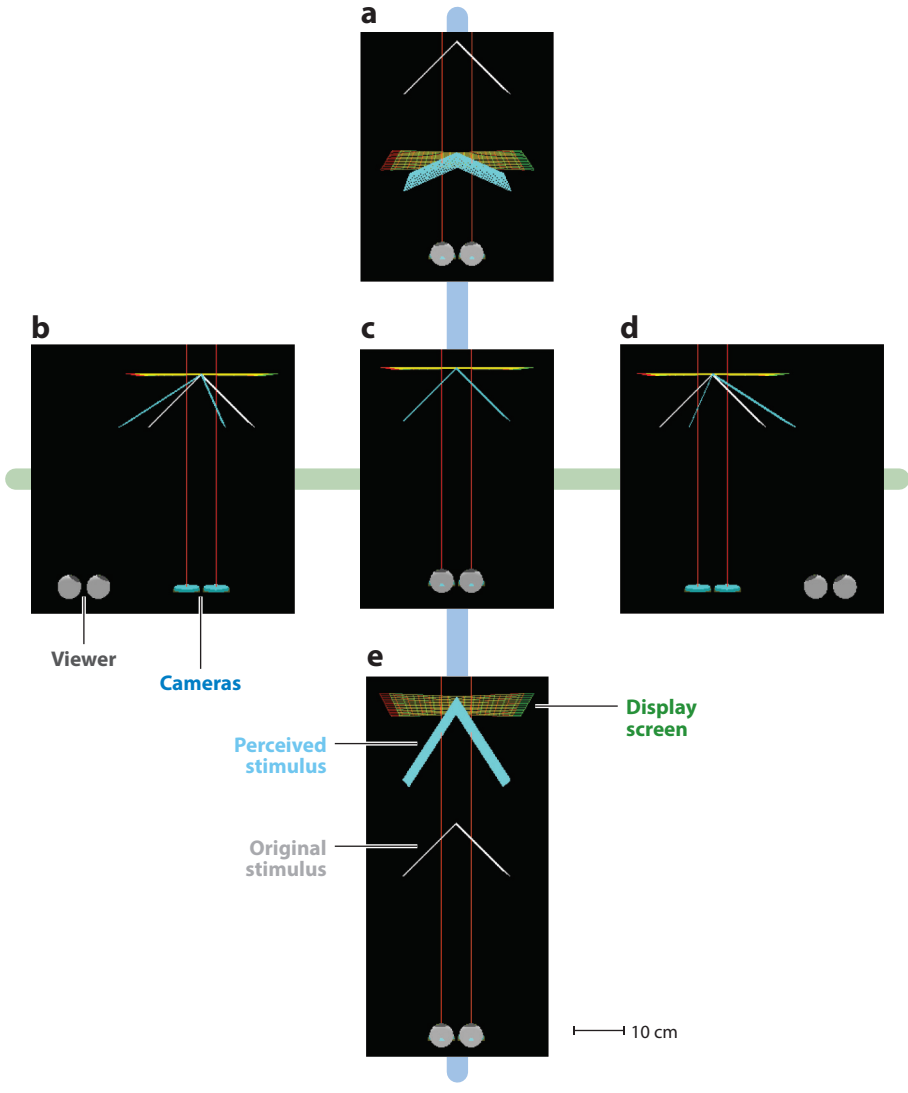

Figure 2

Predicted 3D percepts for different capture-display-viewing situations. Each panel shows an overhead view of the viewer (*dark gray*), stereo cameras (*blue*), display screen (*yellow*), original stimulus (*light gray*), and predicted percept (*light blue*). The parameters are the following. Capture: parallel orientation of optical axes, intercamera separation of 6.2 cm, focal length of 6.5 mm. Display: one display screen, picture magnification of 69×. Viewing: viewing distance of 45 cm, interocular distance of 6.2 cm, viewer positioned such that midpoint of interocular axis is on central surface normal of display screen, viewer oriented with face parallel to display surface. The stimulus is a vertical hinge with a hinge angle of 90°. (*c*) With display and viewing parameters set correctly for the capture parameters, the original and predicted perceived stimuli are identical. (*a*) Viewer is too close to the display. Predicted perceived hinge angle is greater than 90°. (*e*) Viewer is too far from the display. Perceived angle is now less than 90°. (*b*) Viewer is translated to left of display. Predicted hinge rotates toward viewer, and predicted angle is less than 90°. (*d*) Viewer is translated to right of display. Predicted hinge rotates toward viewer, and predicted angle is less than 90°. Figure adapted from Held & Banks (2008).

when the viewer's head is tilted to the side, the strictly horizontal on-screen disparities become partly vertical at the viewer's eyes.

Consider a stereoscopic image with on-screen horizontal disparity H_d that varies sinusoidally over time and has a mean of zero in the equation

$$H_d = A\sin(t), \tag{1}$$

where A is disparity amplitude. The on-screen vertical disparity is zero. Assuming that the on-screen horizontal disparity does not vary with position in the stimulus, the temporal change in disparity specifies an approaching and receding frontoparallel plane. Now consider rolling the head by angle φ while viewing the same stereoscopic image. The interocular axis is no longer horizontal, so the orientation of the disparities at the eyes changes. In particular, horizontal on-screen disparities are now horizontal and vertical disparities relative to the head. Ignoring the small torsional eye movements made with head roll relative to gravity (Collewijn et al. 1985), the retinal disparities are

$$\begin{aligned} \tilde{H}_d &= H_d(t)\cos(-\varphi), \\ \tilde{V}_d &= H_d(t)\sin(-\varphi). \end{aligned} \tag{2}$$

There are two likely consequences.

1. The amount of perceived depth is reduced because we know that perceived depth variation is determined almost entirely by variations in horizontal, not vertical, disparity. Indeed, if the viewer's head were rolled 90°, no depth variation should be seen because horizontal disparity would be effectively zero. Kane et al. (2012) measured reductions in perceived depth with head roll that are well predicted by Equation 2.
2. When a portion of horizontal disparity is converted into vertical disparity, the viewer must make both horizontal vergence eye movements (one eye leftward and the other rightward) and vertical vergence movements (upward and downward) to fuse the stimulus. Vertical vergence occurs naturally, but it is quite small and slow relative to horizontal vergence (Howard et al. 1997, Krishnan et al. 1973). **Figure 3** shows that amount of required vertical vergence as a function of head roll and on-screen horizontal disparity. The need to make vertical vergence movements leads to two undesirable effects: Binocular fusion becomes more difficult and visual discomfort ensues (Kane et al. 2012).

As we noted earlier, circular polarization is much better than linear polarization is at retaining separation of the two eyes' images when the head is rolled. Ironically, this apparent advantage of circular polarization might in the end be a disadvantage because viewers might be more likely to keep their heads upright with linear polarization, thereby avoiding the perceptual and discomfort effects that occur with a tilted head.

2.6. Perceptual Artifacts with Temporal- and Spatial-Interlacing Displays

We next consider spatiotemporal artifacts and distortions of perceived depth that occur with different types of stereoscopic displays.

2.6.1. Spatiotemporal artifacts. It is generally desirable for visual displays to create faithful impressions of the real world, but of course they provide only spatiotemporal approximations. For example, a smoothly moving object is represented by a sequence of static views, which may or may not create the impression of smooth motion. Likewise, a complex image with fine detail is represented by the pattern of illumination from discrete points on the screen—pixels—and the density of pixels may or may not be sufficient to yield a convincing impression of the detail. Here, we first

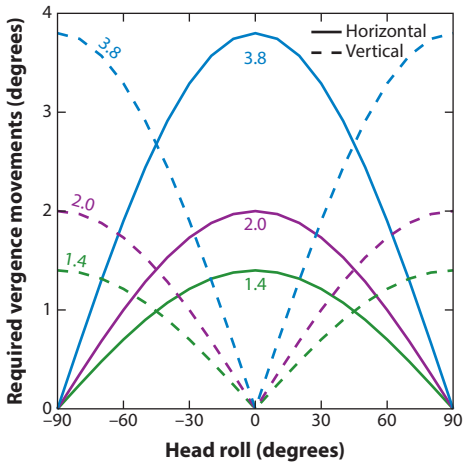

Figure 3

Horizontal and vertical vergence eye movements required to track and fuse a stereoscopic image when the head is tilted to the side. The on-screen horizontal disparity of the stimulus oscillates by 1.4°, 2°, or 3.8° from peak to trough. The required horizontal and vertical vergence movements are indicated by the solid and dashed curves, respectively. Figure adapted from Kane et al. (2012).

discuss how to evaluate the temporal properties of stereoscopic displays and summarize research on the perception of those properties. We then discuss how to evaluate the spatial properties of stereoscopic displays and summarize research on perceiving those properties.

Nonstereoscopic digital displays have been widely adopted, so the requirements for creating acceptable motion and spatial resolution have been well researched and described (Sugawara et al. 2008, Watson 2013, Watson et al. 1986). But the introduction of stereoscopic displays requires a reevaluation because the prevailing techniques for creating separate images to the two eyes introduce noticeable temporal and spatial artifacts that are not observed with nonstereoscopic displays. We take advantage of previous work with nonstereoscopic displays to better understand problems that arise with stereoscopic displays.

The impressions created by digital displays can be understood by relating the display's spatial and temporal characteristics to the spatiotemporal properties of the visual system. The analysis can be done in space and time or spatial and temporal frequency, but one gains more insight from analysis in the frequency domain, so we focus there. We initially consider the images produced for one eye because many of the perceptual artifacts that occur with stereoscopic displays are best understood by considering monocular inputs.

The display of video content involves three dimensions (two in space, one in time), but we show the analysis for two dimensions (one in space, one in time) for ease of visualization. The pipeline from image data to presentation to the eye is schematized in **Figure 4*a***. The stimulus in this example is a horizontal line moving vertically across the screen at constant speed. Typically, image data $i(y,t)$ are antialiased before being sent to the display, which we do by convolving with an interpolation function, $a(y,t)$. We then calculate how intensity varies over space and time when the image data are presented on the display. To do so, we sample the antialiased image data with a comb function representing the display's spatiotemporal sampling, where the samples are separated spatially by y_0 (pixel spacing) and temporally by t_0 (frame time). The encoding method can affect the periodicity of the sampling function. Temporally interlacing displays double the temporal separation, whereas spatially interlaced displays double the spatial separation.

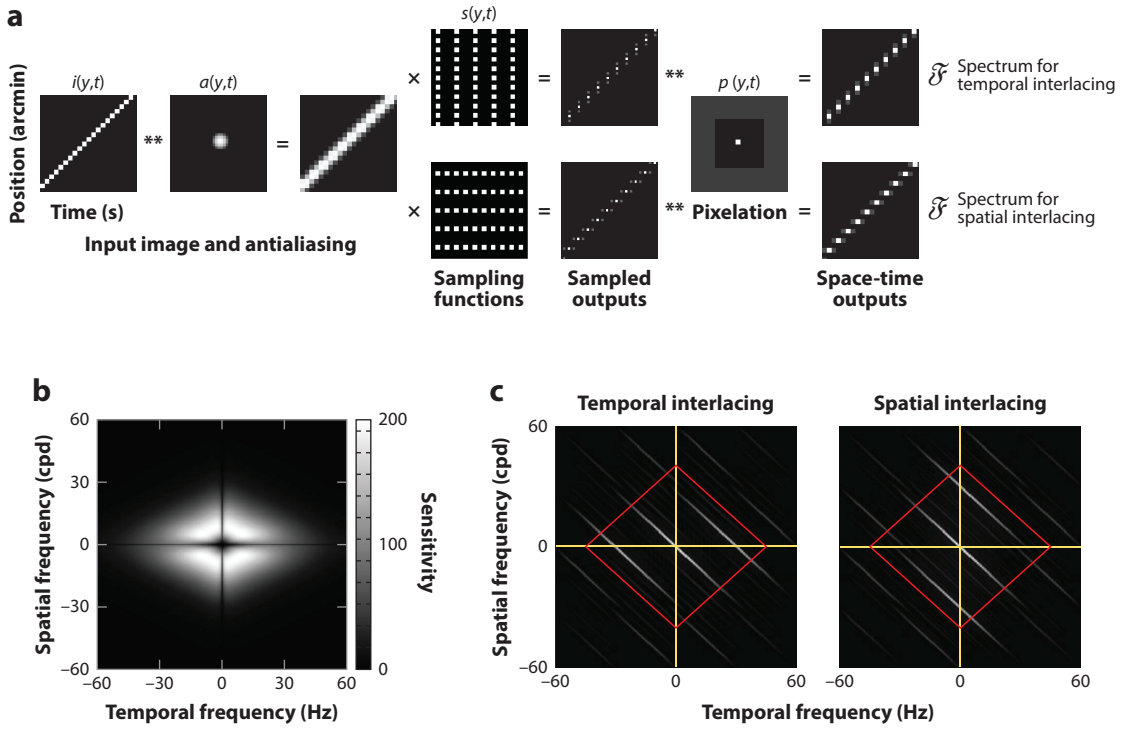

Figure 4

Pipeline of image generation, display, and viewing. (*a*) The pipeline proceeds from left to right. Where there are two rows, the upper one represents the pipeline for temporal interlacing, and the lower one represents the pipeline for spatial interlacing. The image data $i(y,t)$ represent an object moving vertically at a constant speed of 1 degree/s. The image data are antialiased by convolution with a cubic-interpolation function $a(y,t)$. Then the antialiased data are sampled with spatiotemporal comb function $s(y,t)$ with samples separated by y_0 and t_0 (1 arcmin and 16.7 ms). Different encoding methods for stereoscopic displays result in different periodicities of the sampling function (upper and lower show temporally and spatially interlaced displays, respectively). The displayed intensities have finite spatial and temporal extent (pixelization) represented by spatiotemporal aperture function $p(y,t)$. The resultant is a space-time plot of the sampled and pixelated imagery being presented on the screen: upper one for temporal interlacing and lower for spatial interlacing. Those space-time plots are then subjected to Fourier transformation, represented by $\mathfrak{F}$. (*b*) Human spatiotemporal contrast-sensitivity function. Sensitivity (reciprocal of contrast required for detection) is plotted as a function of temporal and spatial frequency. Panel based on data from Kelly (1979). (*c*) Amplitude spectra for temporally and spatially interlacing displays. The amplitude spectra were derived from the pipeline in panel *a*. Amplitude is plotted as a function of temporal and spatial frequency. A diagonal line through the center of each panel (not shown) would be the spectrum of the smoothly moving stimulus (the signal). The diagonal through the center that is shown is a filtered version of the signal. The other lines are aliases created by sampling. Red diamonds represent the spatiotemporal contrast sensitivity function in panel *b*. Components within the diamonds will be visible, whereas components outside the diamonds will not be. Abbreviation: cpd, cycles per degree.

The displayed intensities have finite spatial and temporal extent, which we represent with spatiotemporal aperture function $p(y,t)$. The aperture function is set such that pixel width is equal to pixel spacing (i.e., fill factor is 1), and duration of illumination in each frame is equal to frame period (i.e., duty cycle is 1). In real displays, the fill factor would be a bit less than 1 and duty cycle could be much less than 1 depending on the technology being used, illustrated in the equations

$$[[i(y,t) **a(y,t)]s(y,t)] **p(y,t) \qquad (3)$$

and

$$\left[[i(y,t) **a(y,t)]\, comb\left(\frac{y}{y_0}, \frac{t}{t_0}\right)\right] **rect\left(\frac{y}{y_0}, \frac{t}{t_0}\right), \quad (4)$$

where *rect* is a scaled rectangle function with widths y_0 in space and t_0 in time, and y_0 and t_0 also represent the spatial and temporal separations of samples in the comb function. In the Fourier domain, Equation 4 becomes

$$[[I(f_y, f_t)A(f_y, f_t)] **comb(y_0 f_y, t_0 f_t)]sinc(y_0 f_y, t_0 f_t), \quad (5)$$

where f_y and f_t are spatial and temporal frequency and the *sinc* functions have zeros at $f_y = 1/y_0$, $2/y_0$, etc., and $f_t = 1/t_0$, $2/t_0$, etc.

The rightmost panels in **Figure 4*a*** are space-time intensity distributions created by temporally and spatially interlacing stereoscopic displays. They have been calculated for one eye's image for reasons we make clear later. The amplitude spectra associated with those distributions are shown in **Figure 4*c***. An extended diagonal line through the center of each panel (not shown) would be the spectrum of the original smoothly moving stimulus (the signal). The diagonal through the center that is shown is a filtered version of that signal. The other lines are aliases introduced by the spatiotemporal sampling. The separation of the aliases along the temporal-frequency axis is $1/t_0$ (the frame rate) and along the spatial-frequency axis is $1/y_0$ (pixel spacing).

The human visual system is not equally sensitive to contrast variation at all spatial frequencies: Variations above a particular frequency are invisible because they exceed the acuity limit. Likewise, humans are not equally sensitive to contrast modulations at all temporal frequencies: Variations above a specific frequency are invisible because they exceed the critical flicker frequency. The variation in sensitivity to different combinations of spatial and temporal frequency is not separable (i.e., cannot be represented by the product of a function in spatial frequency and a function in temporal frequency) (Kelly 1979). The sensitivity to different combinations is represented by the spatiotemporal contrast sensitivity function (the window of visibility) (**Figure 4*b***). To first approximation, spatiotemporal frequencies falling within the red diamonds in **Figure 4*c*** will be visible and those falling outside the diamonds will be invisible. The amplitude spectra of a smoothly moving stimulus and the digital approximation to that stimulus differ primarily at higher spatiotemporal frequencies, so those differences may not be visible, in which case we predict that the digital stimulus will appear to move smoothly just like an object moving in the real world.

The viewer will see various artifacts when the aliases are low enough in frequency and high enough in amplitude to be visible. Two general spatiotemporal artifacts have been examined: flicker and motion artifacts. Visible flicker is defined as perceived fluctuation in the brightness of a (usually stationary) stimulus. Motion artifacts are defined as object motions that appear unsmooth or otherwise distorted.

2.6.1.1. Flicker. The relevant part of the amplitude spectrum for predicting flicker visibility is along the temporal-frequency axis. When aliases on the axis fall within the window of visibility, we expect flicker to be visible. Visibility is influenced by several factors, including the temporal frequency and amplitude of luminance modulation, overall luminance (Kelly 1972), stimulus area, and retinal eccentricity (Rovamo & Raninen 1988). Conditions producing the most noticeable flicker are large bright areas in peripheral vision with high-amplitude modulation at lower temporal frequencies. The modulation amplitude differs greatly for temporally and spatially interlacing displays. In spatial displays, the duty cycle of illumination to an eye can be nearly 1, whereas in temporal displays, duty cycle can be no greater than 0.5 because the two eyes' images are shown in alternation. The reduction in duty cycle makes flicker much more likely to be visible. For a spatial- or temporal-interlacing display alternating between luminances of L_{max} and L_{min} at a particular

temporal frequency, the contrast of the lowest frequency component is

$$\left[\frac{2\sin(\pi d)}{\pi d}\right]\left[\frac{d(L_{\max}-L_{\min})}{d(L_{\max}-L_{\min})+L_{\min}}\right], \tag{6}$$

and the average luminance is

$$d(L_{\max}-L_{\min})+L_{\min}, \tag{7}$$

where d is duty cycle (Campbell & Robson 1968). These equations show that shorter duty cycles generally create greater amplitudes of the fundamental frequency and lower average luminances. Greater amplitudes would yield more visible flicker, whereas lower luminance would yield less visible flicker. The amplitude effect is generally much greater than the luminance effect, so shortening the duty cycle typically makes flicker more visible. Thus temporal-interlacing displays should produce more visible flicker than spatially interlacing displays.

Hoffman et al. (2011) measured flicker visibility with various presentation protocols. For each protocol, they found the highest frame rate at which flicker became visible. When the presentations alternated between eyes (as in temporal interlacing), flicker was just visible at a presentation rate of ~40 Hz. When the images were delivered simultaneously to the two eyes, flicker was slightly more visible. This means that some cancellation of flicker occurs in combining the two eyes' images binocularly (Cavonius 1979), but the effect is small. Thus, the primary determinant of flicker visibility is the fundamental temporal frequency presented to one eye. To avoid visible flicker, temporally interlacing stereoscopic displays require higher frame rates than nonstereoscopic displays and spatially interlacing stereoscopic displays because of the need to alternate illuminated and nonilluminated frames, which by reducing duty cycle increases the amplitude of the fundamental frequency.

The content for film-based cinema was almost universally captured at 24 Hz. The duty cycle of presentation was much less than 1 to provide sufficient time for each film frame to advance before being illuminated. A mechanical shutter ensured that the film was not illuminated while it was being advanced and was illuminated while the film frame was stationary. When a sequence of still images was projected at that rate, flicker was very noticeable, so film engineers developed a technique for presenting each frame twice before advancing to the next frame (Elkins 2013). This double-shuttering technique greatly reduced flicker visibility by presenting 24-Hz data at 48 Hz.

Because of the need to insert dark frames in stereoscopic temporal-interlacing displays, double shuttering and even triple shuttering has been used to reduce flicker visibility (Cowan 2008). Hoffman et al. (2011) and Johnson et al. (2015a,b) verified that multi-shuttering does indeed reduce flicker visibility. Wilcox et al. (2015) reported that viewers of stereoscopic 3D displays have a general preference for higher frame rates in part because they reduce flicker visibility.

2.6.1.2. Motion artifacts. Motion artifacts include judder (jerky or unsmooth motion appearance), motion blur (apparent smearing in the direction of stimulus motion), and banding (appearance of multiple edges in the direction of stimulus motion). We next describe the conditions that make these artifacts visible in stereoscopic displays. The theoretical analysis was originally developed for nonstereoscopic displays, but the same principles apply.

As with flicker, one can best understand the visibility of motion artifacts by analysis in the frequency domain. In that domain, one finds that the spatiotemporal frequencies of the aliases are affected only by capture rate and speed of object motion (Hoffman et al. 2011, Klompenhouwer 2006, Watson 2013). Changes in duty cycle and single, double, or triple shuttering do not change the alias frequencies, but they affect the magnitude of aliases at different temporal frequencies.

A smoothly moving real-world stimulus does not generate aliases, but a digitally presented rendition does (**Figure 4*c***). Thus, one predicts that smooth, artifact-free motion will be perceived if no aliases fall within the window of visibility and if the amplitude of the central spectral component does not deviate noticeably from the amplitude of that component for a smoothly moving stimulus. Increasing the speed of motion s and decreasing the capture rate τ_c move the aliases toward the window of visibility, making them more likely to be seen. Thus, artifact visibility should be roughly constant when the ratio s/τ_c is constant.

Viewers often track a real moving object with smooth-pursuit eye movements, thereby keeping the object on the fovea. Assuming the pursuit movement is accurate, the stimulus becomes stationary on the retina, so its retinal speed s_{retinal} equals 0. When a viewer tracks a digitally displayed moving object, different sorts of motion artifacts become visible (Watson 2010). With tracking, the artifacts tend to be motion blur and edge banding. Without tracking (i.e., stationary fixation), the artifacts tend to be judder. The cause of the change in the type of artifact seen is schematized in **Figure 5**. The top row shows the retinal position of the stimulus over time for stroboscopic (duty cycle $d \approx 0$) and sample-and-hold presentation ($d > 0$). Each image presentation of duration

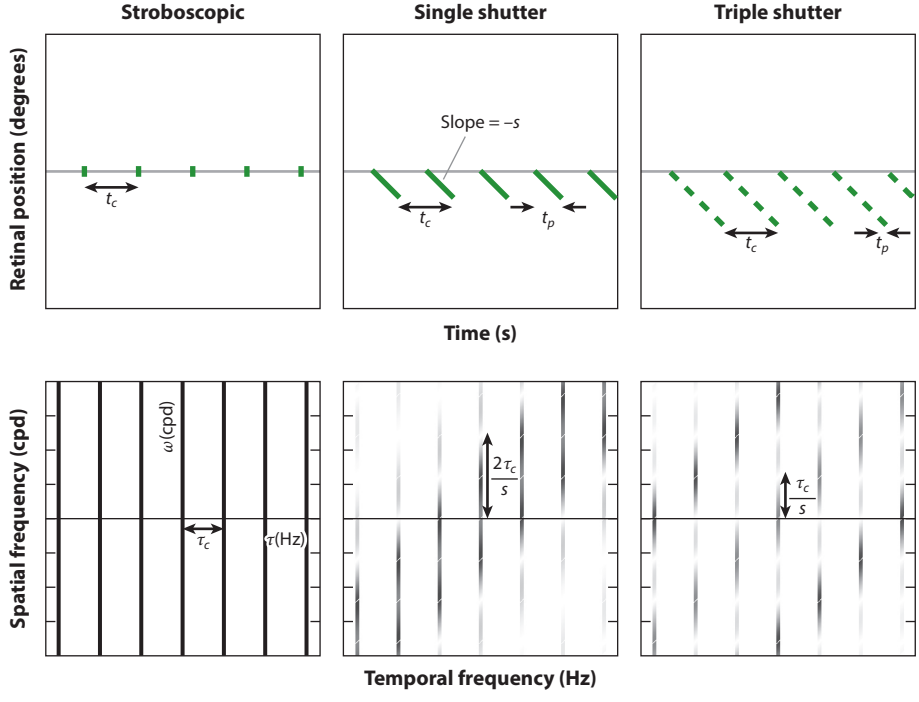

Figure 5

Motion artifacts with tracking eye movements. Columns depict stroboscopic, single-, and triple-shutter protocols. In each case, the viewer makes a smooth eye movement so eye speed matches object speed on the screen. The top row shows the retinal position of the stimulus as a function of time for different protocols. Frame time is t_c. Presentation time is t_p. Gray horizontal lines represent a smoothly moving stimulus. Green dots and line segments represent discrete stimuli moving at the same speed. Each sample of the discrete stimulus shifts across the retina by $\Delta x = -st_p$. The bottom row shows amplitude spectra for each stimulus. The abscissa is temporal frequency (τ), and the ordinate is spatial frequency in retinal coordinates (ω). The origin ($\tau = 0, \omega = 0$) is in the middle of each panel. Capture rate is τ_c. Darker grays represent greater amplitude. Abbreviation: cpd, cycles per degree.

t_p (where $t_p = d/\tau$ for duty cycle d and frame rate τ) displaces across the retina by $\Delta x = -st_p$ or $\Delta x = -sd/\tau$. Thus, significant displacement can occur with high stimulus speeds, low frame rates, and long duty cycles, which blur the stimulus on the retina, yielding visible motion blur. Multi-shuttering (**Figure 5**, *right column*) should decrease the blur but increase edge banding because of repeated presentation on slightly different parts of the retina. The amplitude spectra in retinal coordinates for the three presentation protocols are displayed in the bottom row of **Figure 5**. The signal and aliases are sheared parallel to the temporal-frequency axis such that they have a slope of $-1/s_{\text{retinal}}$; in other words, they become vertical. The zero crossings of the aliases are unchanged because eye movements do not affect the rate at which images are delivered. The envelope by which the signal and aliases are attenuated is sheared in the same fashion as the signal and aliases. The amplitude spectra in the bottom row are the sheared signal and aliases multiplied by the sheared envelope. Imagine that the component along the spatial-frequency axis is the only visible component because of filtering by the window of visibility. The stroboscopic stimulus ($d = 0$) has a uniform spectrum along the spatial-frequency axis, so it should look like a vertical line that is stationary on the retina and would therefore be perceived to move smoothly as the eye rotates. The more realistic sample-and-hold stimuli in the middle and right columns have duty cycles greater than 0, and this generates amplitude spectra that are attenuated along the spatial-frequency axis, more attenuation with greater speeds, larger duty cycles, and lower frame rates. The attenuation along that axis produces motion blur.

From this analysis, one expects the following: (*a*) Combinations of speed and capture rate that yield a constant ratio (s/τ_c) should have equivalent motion artifacts. (*b*) Multi-shuttering to minimize flicker for a given capture rate ($\tau_p = f\tau_c$) should not reduce visibility of motion artifacts. (*c*) Edge banding should be determined by the number of repetitions in multi-shutter protocols: two bands being perceived with double shutter, three with triple shutter, etc. (*d*) Sensitivity in the disparity domain is restricted to a much smaller range of spatial and temporal frequencies than in the luminance domain (Banks et al. 2004, Kane et al. 2014, Norcia & Tyler 1984, Tyler 1974). Thus, one predicts that motion artifacts will be no more visible with binocular, disparity-varying stimuli than with monocular stimuli.

Hoffman et al. (2011) and Johnson et al. (2015a,b) measured the probability of observing motion artifacts with stereoscopic displays for a variety of conditions and compared their findings to the above predictions. Consistent with the first prediction, the ratio of object speed divided by capture rate (s/τ_c) was a good predictor of artifact visibility. Thus, to maintain the appearance of smooth motion, an increase in object speed must be accompanied by an increase in capture rate. Consistent with the second prediction, the results were similar across different protocols except that at a given capture rate, the multi-shutter protocols produced artifacts at slightly slower speeds than the corresponding single-shutter protocol did. Consistent with the fourth prediction, Hoffman et al. observed only a very small increase in the likelihood of observing motion artifacts with binocular as opposed to monocular viewing.

2.6.1.3. Spatial resolution. In spatial-interlacing displays, the left- and right-eye views are presented simultaneously, but one eye receives the odd rows on the display while the other eye receives the even rows. Because each eye receives half the pixels, it seems likely that the effective spatial resolution will be lower than with temporal-interlacing displays. However, some researchers have claimed that the visual system can construct a full-resolution binocular view from the two monocular views (Kelley 2011, Soneira 2012). The claim in effect is that the visual system composites the bright rows while rejecting the dark rows in forming the binocular image. This claim is implausible, given what we know about binocular fusion. Binocular matching is made between monocular features with similar properties—in other words, similar orientation (Mitchell & O'Hagan 1972),

motion (van Ee & Anderson 2001), color (Jordan et al. 1990), and contrast polarity (Krol & van de Grind 1983). The seen and unseen rows in each eye form a pattern of bright and dark rows. It is very unlikely that the viewer would match bright rows in one eye with dark ones in the other eye because such matches would violate the contrast-polarity constraint (Krol & van de Grind 1983). Instead, the viewer is likely to make small vertical vergence eye movements (one eye rotating upward or downward relative to the other) to align the bright and dark rows in the retinas to match bright to bright and dark to dark (Hakala et al. 2015). As a consequence, there should be a loss in effective resolution in binocular viewing.

To examine this, Kim & Banks (2012) measured letter acuity in spatial- and temporal-interlacing stereo displays. At the industry-recommended viewing distance (where a pixel subtends 1 arcmin), acuity was significantly worse with spatial than with temporal interlacing. At longer viewing distances (where pixel rows were too small to be visible), acuity became the same in the two types of displays. Thus, at the recommended distance, the effective resolution of spatial-interlacing displays is indeed lower than that of temporal-interlacing displays. Kim & Banks also measured letter acuity in spatially and temporally interlacing displays with binocular and monocular viewing. According to Kelley (2011) and Soneira (2012), effective resolution should be much lower with monocular viewing of spatial-interlacing displays because the second eye's data would not be available to fill in the missing data from every other row in the first eye's data. Kim & Banks observed only slightly lower resolution with monocular viewing, a reduction that is consistent with the well-known improvement in binocular over monocular acuity (Blake & Fox 1973, Blake et al. 1981, Campbell & Green 1965). Most significantly, the monocular-binocular difference was the same for spatial and temporal interlacing displays. This is quite inconsistent with Kelley's (2011) and Soneira's (2012) claim because they have to predict a larger decrease with spatial interlacing, as the second eye's data are not delivered with monocular viewing. Yun et al. (2013) also observed lower effective resolution with spatial- as opposed to temporal-interlaced stereo displays.

2.6.2. Distortions of perceived depth. We next consider the distortions of perceived depth that occur with temporal and spatial interlacing.

2.6.2.1. Depth distortions in temporal interlacing. When a continuous moving stimulus is presented binocularly but the neural response to one eye's image is delayed relative to the other's by reducing the luminance in one eye, the Pulfrich effect occurs: Sinusoidal motion in the frontoparallel plane appears elliptical in depth (Julesz & White 1969, Pulfrich 1922, Ross & Hogben 1975). This alteration of perceived depth is readily explained. Suppose that the image being delivered to the right eye is delayed by Δi relative to those delivered to the left eye. An object moving rightward with speed s is at position x when it is first seen by the left eye. By the time this same image reaches the right eye, the left eye's image is at a new position of $x + s\,\Delta i$. At this instant, the two eyes' images are at different positions, creating a spatial disparity of $s\,\Delta i$, which leads naturally to a change in perceived depth.

Similar distortions in the perceived depth of moving stimuli occur in temporally interlacing displays. **Figure 6** illustrates the sequence of images delivered to the two eyes with a typical temporally interlacing display. The left panel is a space-time plot of a stimulus moving horizontally at constant speed. It has a spatial disparity of zero, so it should be seen as moving in the plane of the display screen. But it often appears to be in front of or behind the screen, depending on the direction of motion and order of eye stimulation. These distortions of perceived depth are a serious problem because they lead to situations in which one depth cue, binocular disparity, is in conflict with many others. Imagine, for example, a woman moving from left to right and that the error in the disparity estimate causes her to appear closer. If she passes behind a stationary object,

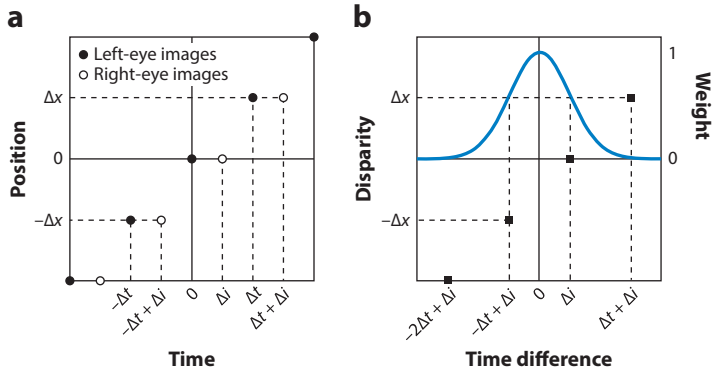

Figure 6

Disparity computation with temporal interlacing. (*a*) Space-time plot of a stimulus moving horizontally at constant speed on a temporally interlacing display. The stimulus has a spatial disparity of zero and a speed of $\Delta x/\Delta t$. Left- and right-eye presentations are represented by filled and unfilled symbols, respectively. In each frame, right-eye images are delayed by Δi relative to left-eye images. (*b*) Disparity estimation with weighted averaging over time. The abscissa represents the arrival time of each candidate match from the right eye relative to the reference image from the left eye. The left ordinate represents the disparity of each potential match. The black squares represent the disparities and time differences for candidate matches. The right ordinate represents the weight given to each match. The estimated disparity is a weighted average of the disparities of potential matches. In the example, the stimulus is moving rightward and the left image leads the right, so the erroneous disparity leads to the object being seen nearer than intended.

it can be startling to see a person, who initially appeared closer than the object, to suddenly be occluded by that object.

To estimate spatial disparity with any sequence of images, the brain has to solve the binocular-matching problem: Which image feature in one eye should be matched with a given feature in the other eye? With temporal-interlacing displays, no features are presented simultaneously, so the neural mechanisms that perform the matching have to determine whether a given image in the left eye should be matched with a later or earlier image in the right eye. Both right-eye images are offset by the same interocular delay relative to the left-eye image, so there is presumably no way to know which match to make. Several observations indicate that the brain uses a time-weighted average to solve the problem (Hoffman et al. 2011; Johnson et al. 2015a,b; Kuroki 2012; Read & Cumming 2005). The averaging process is depicted in **Figure 6*b***. The weighting function gives the highest weight to images that are delivered simultaneously and successively lower weights to images that arrive at increasingly different times.

When the interocular delay is less than ~50 ms, as it always is with practical temporal-interlacing displays, the behavior of the weighted-averaging model is very consistent with observed depth distortions (Burr & Ross 1979; Hoffman et al. 2011; Johnson et al. 2015a,b; J. Kim et al. 2014a; Morgan 1979; Read & Cumming 2005).

There are two obvious ways to minimize or eliminate the distortion of perceived depth. (*a*) Eliminate the interocular delay; in other words, set Δi to 0. Johnson et al. (2015a,b) found, as predicted, that the distortion was eliminated. (*b*) Adjust the spatial disparity to be consistent with the interocular delay; in other words, for a stimulus speed of $\Delta x/\Delta t$, shift the position of the right-eye image by $\Delta x/2$. Hoffman et al. (2011) found that this did indeed eliminate the distortion. One can also reduce the magnitude of the depth distortion by increasing the capture rate (number of unique video frames per unit time). In that case, Δt is reduced, yielding a time-average disparity estimate closer to the intended value.

The visual system estimates disparity primarily from luminance information as opposed to chromatic information (Lu & Fender 1972). One can exploit this to reduce depth distortions in temporal-interlacing stereoscopic displays that use narrowband-wavelength filtering for image separation (Jorke et al. 2009). Specifically, one can present the green primary to the left eye at the same time as presenting the red and blue primaries to the right eye, and then present the green to the right eye at the same time as red and blue to the left eye (J. Kim et al. 2014a; Simon & Jorke 2011). This color-interlacing technique greatly reduces the magnitude of perceived depth distortions due to temporal interlacing provided that the stimulus is not highly saturated (J. Kim et al. 2014a). When the stimulus is saturated (e.g., all green), color interlacing becomes equivalent to temporal interlacing and the expected depth distortions are seen. Most stimuli are, however, not highly saturated so temporal-interlacing displays can benefit from multiplexing primaries.

2.6.2.2. Depth distortions in spatial interlacing. The slight vertical misalignment between the left- and right-eye images in row-by-row spatial interlacing presents an interesting problem for the interpretation of depth. An illuminated row in the left eye's image is at the same elevation as an unilluminated (i.e., dark) row in the right eye's image. When the rows are visible, the visual system will match bright rows to bright rows and dark to dark (Krol & van de Grind 1983) by making a small vertical vergence eye movement. The vertical movement can change horizontal disparity at the retina. When displayed contours are neither vertical nor horizontal, the alteration of retinal horizontal disparity should affect the depth interpretation of the binocular image (Hakala et al. 2015). **Figure 7** illustrates this. Two intersecting diagonal lines are presented with zero horizontal disparity on the display, so both limbs of the X should be seen in the screen plane. If one eye rotates vertically to align illuminated and unilluminated rows retinally, the binocular image acquires unintended horizontal disparity. In **Figure 7*c***, the right eye has rotated downward relative to the left, introducing uncrossed disparity for the limb sloping up and to the left and crossed disparity for the other limb. Thus, in the fused image, the lines should appear at different depths and not to intersect. Hakala et al. (2015) presented such stimuli on a spatial-interlacing display and measured the magnitude of the depth alteration. They found that perceived depth was indeed consistent with the geometric analysis in **Figure 7**. This alteration of perceived depth can be mitigated by spatially averaging the image data across two rows (Kelley 2011). Hakala et al. (2015) found that such vertical averaging does indeed eliminate the alteration of perceived depth but that the averaging causes some loss of spatial detail.

2.7. Vergence-Accommodation Conflict

Binocular fixation involves two oculomotor functions: vergence and accommodation. The former is the rotation of the eyes in opposite directions to obtain a single fused image of the fixated object. The latter is the adjustment of the power of the crystalline lens to obtain a sharp image. Thus, accurate vergence and accurate accommodation are both required to achieve a single, clear image of a fixated object. The primary stimulus that drives vergence is, of course, binocular disparity (disparity-driven vergence), and the primary stimulus that drives accommodation is retinal-image blur (blur-driven accommodation).

But vergence and accommodation are also coupled. Specifically, accommodative responses evoke vergence responses (blur-driven vergence or accommodative vergence), and vergence responses evoke accommodative responses (disparity-driven accommodation or vergence accommodation) (Fincham & Walton 1957, Martens & Ogle 1959, Schor 1992). Vergence-accommodation coupling is helpful in the real world because vergence and focal distances are

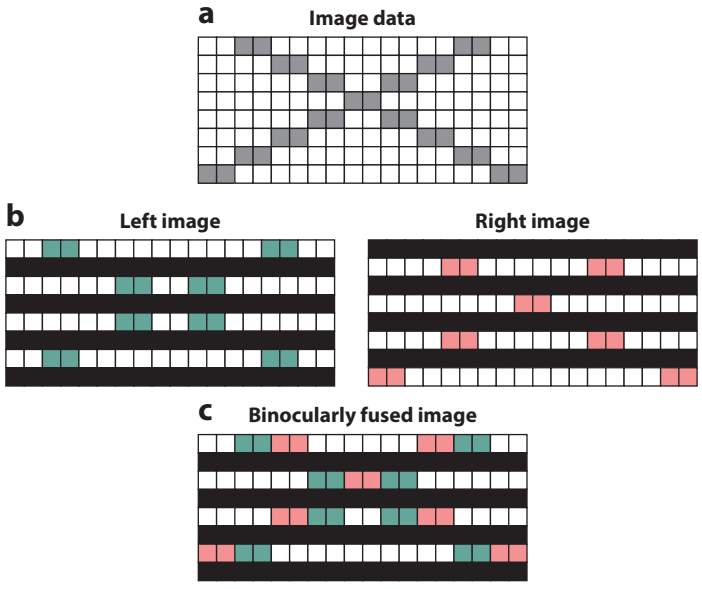

Figure 7

Disparity errors in spatial-interlacing displays. (*a*) The image data to be presented on the display. Each square represents a pixel. The image has zero disparity and should be seen in the plane of the screen. (*b*) The data displayed to the left and right eyes. The dark rows represent unilluminated rows (even rows to the left eye, odd rows to the right) and bright rows represent illuminated rows (odd rows to the left, even rows to the right). (*c*) The presumed binocularly fused image. In binocularly fusing the two eyes' images, the visual system matches illuminated rows in the left eye with illuminated rows in the right and likewise for the unilluminated rows. To do this, the right eye rotates upward or downward by one pixel row relative to the left eye. In the bottom panel, the right eye has rotated downward. This vertical vergence eye movement creates horizontal disparity at the retina. Uncrossed disparity is created for edges that are tilted counterclockwise from vertical and crossed disparity for edges rotated clockwise.

almost always the same no matter where one looks. The coupling increases speed of response. Specifically, accommodation is faster with binocular viewing—where blur and disparity signals specify the same change in distance—than it is with monocular viewing where only blur provides a useful signal (Cumming & Judge 1986, Krishnan et al. 1977). Similarly, vergence is faster when disparity and blur signals specify the same change in distance than when only disparity specifies a change (Cumming & Judge 1986, Semmlow & Wetzel 1979).

When the distances to which the eyes must converge and accommodate differ, the visual system must override the vergence-accommodation coupling. This produces the vergence-accommodation conflict. The conflict is commonplace in conventional stereoscopic 3D displays and is frequently cited as a cause of visual discomfort (Howarth 2011, Kooi & Toet 2004, Lambooij et al. 2009, Sheedy et al. 2003, Urvoy et al. 2013).

We first discuss this conflict in the context of optically correcting a patient because several useful concepts have arisen from that situation. We then discuss the viewing of stereoscopic displays and whether the optometric/ophthalmic concepts are useful to the design, assessment, and use of such displays.

When the relationship between the accommodative stimulus and vergence stimulus is altered by a new optical correction, the patient may be unable to maintain clear and single vision

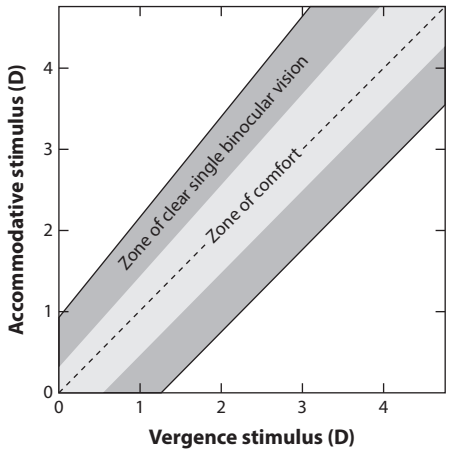

Figure 8

Zone of clear single binocular vision (ZCSBV) and Percival's zone of comfort. ZCSBV is represented by the whole gray region. It represents the combinations of vergence and accommodative stimuli [here expressed in diopters (D)] for which a viewer can maintain sharp and fused binocular percepts by accommodating and converging sufficiently accurately. Percival's zone of comfort is represented by the light gray. It represents the combinations of vergence and accommodative stimuli for which people do not experience discomfort.

simultaneously. The zone of clear single binocular vision (ZCSBV) summarizes the vergence-accommodation conflicts that allow maintenance of fused, sharp imagery (Fry 1939, Hofstetter 1945) (**Figure 8**).

Optometrists and ophthalmologists found that patients often experience visual discomfort or asthenopia (Sheedy et al. 2003) when attempting to manage the vergence-accommodation conflict induced by new optical correction. From clinical experience, Percival (1928) and Sheard (1930) proposed zones of comfort. Percival's zone is the middle third of ZCSBV at each accommodative distance (light gray in **Figure 8**). Sheard's has the same width as Percival's but is also determined by the person's phoria.

Vergence-accommodation conflicts also arise in viewing stereoscopic displays because the vergence stimulus can be at various distances relative to the display screen, whereas the accommodative stimulus remains fixed at the screen. The conflicts that arise are different than those that occur with optical correction. Optical correction introduces a constant offset in diopters (D) relative to natural viewing at all distances. Adding −1D (concave lens), for example, increases the accommodative stimulus by 1D, displacing the diagonal line in **Figure 8** upward. Stimuli presented on stereoscopic displays enable variation in vergence stimuli, but the accommodative stimulus remains at the distance of the screen, which would be a horizontal line in the figure. Thus, the ability to adapt to the vergence-accommodation conflicts induced by viewing stereoscopic displays might be quite different from the ability to adapt to a new optical correction.

Despite the widespread belief that vergence-accommodation conflict is a significant source of discomfort, there are few studies that actually tested the hypothesis directly. Many studies compared symptoms after viewing content stereoscopically to symptoms after viewing similar content nonstereoscopically. Most found that viewers reported more severe symptoms after stereoscopic viewing (Emoto et al. 2005; Häkkinen et al. 2006; Nojiri et al. 2003; Peli 1998; Wöpking 1995; Yang & Sheedy 2011; Yano et al. 2002, 2004). These results have practical importance in assessing visual comfort for the two types of viewing experience, but they cannot

inform us about what the causes of discomfort are because they did not isolate candidate causes. That is, the source of discomfort could be the vergence-accommodation conflict, crosstalk (Kooi & Toet 2004), increased immersion, increased sense of self-motion (Palmisano 1996, 2002), decreased brightness, more salient motion artifacts (Hoffman et al. 2011), and many other things that differ between stereoscopic and nonstereoscopic viewing experiences. Without knowing the cause(s) of discomfort, one cannot know how to minimize them.

To persuasively determine whether a particular property causes discomfort, one has to manipulate that property while holding other potential causes constant. Furthermore, natural viewing in which vergence and accommodation are consistent with one another is the most appropriate baseline for assessing discomfort when they are in conflict. The development of multi-plane stereoscopic displays (Section 3.2) has made this possible. Such displays create digital approximations to a 3D volume with nearly correct focus cues. They are stereoscopic, so they enable independent manipulation of the stimulus to vergence and the stimulus to accommodation.

Recent studies used multi-plane displays or displays coupled with tunable lenses to directly test the hypothesis that vergence-accommodation conflict is a source of visual discomfort (Hoffman et al. 2008, Johnson et al. 2016, J. Kim et al. 2014b, Konrad et al. 2016, Shibata et al. 2011). They presented two main conditions: (*a*) natural viewing in which the distance of the vergence stimulus and the distance of the accommodative stimulus changed in unison as they do in the natural environment; and (*b*) stereo viewing in which the distance of the vergence stimulus changed in the same fashion as the natural-viewing condition whereas the distance of the accommodative stimulus remained constant. Hoffman et al. (2008) found that the vergence-accommodation inconsistencies in stereoscopic viewing induced more discomfort than the consistent vergence and accommodation stimuli in natural viewing. This was direct evidence that vergence-accommodation conflict in stereoscopic displays is a cause of visual discomfort.

Shibata et al. (2011) extended this to estimate the zone of comfort for viewing stereoscopic displays. They too observed that vergence-accommodation conflict causes visual discomfort. They also found that subjects are more susceptible to conflict at long viewing distances and that positive conflict induces greater discomfort at far viewing distance, whereas negative conflict induces greater discomfort at closer viewing distance. From their data, they made a rough estimate of the zone of comfort for young adult viewers of stereoscopic displays (**Figure 9**). Panel *b* shows that discomfort is less likely to occur at typical viewing distances for television and cinema. Consistent with this, Read & Bohr's (2014) study found minimal discomfort with viewing of stereoscopic television.

Although the accommodative stimulus of stereoscopic displays is fixed, accommodative responses can still occur when the simulated distance of an object changes (Inoue & Ohzu 1997, Torii et al. 2008). The change in simulated distance is signaled by binocular disparity and other ancillary depth cues such as perspective and motion parallax, and they can drive accommodation via the vergence-accommodation cross-coupling and the proximal component (Heath 1956). Simultaneously, the blur-driven component to accommodation remains fixed at the screen distance. Therefore, the components that drive accommodation are in conflict: two attempting to drive accommodation according to the video content and one attempting to hold it fixed at the screen. It is therefore not surprising that accommodative responses are more unstable when viewing stereoscopic displays than in natural viewing (Fukushima et al. 2009, Torii et al. 2008).

Visual discomfort is a subjective phenomenon, so most assessments have relied on self-reports of subjective experience. There have, however, been many recent attempts to develop objective indices in the hope that more reliable and rapid assessment can be achieved. These indices have included measurements of performance in visually demanding tasks (Akeley et al. 2004, Hoffman et al. 2008), electroencephalography (EEG) (Mun et al. 2012), functional magnetic resonance imaging (fMRI) (D. Kim et al. 2014), and heart rate (Park et al. 2014). Our reading of this work is that none have proven to be an effective index.

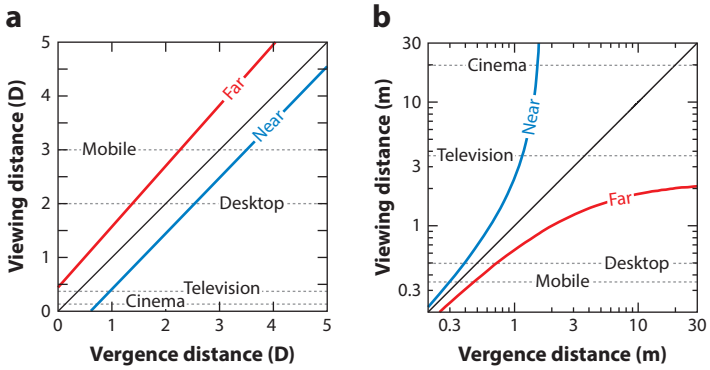

Figure 9

Zone of comfort for stereoscopic viewing. (*a*) Comfort zone plotted in diopters. The abscissa is the distance of the vergence stimulus, and the ordinate is viewing distance, which corresponds to the focal stimulus. The black diagonal line represents natural viewing. The red and blue lines represent estimates from the data of Shibata and colleagues (2011). The dashed horizontal lines represent typical viewing distances for mobile devices, desktop displays, television, and cinema. (*b*) Comfort zone plotted in meters. The abscissa and ordinate are the distance of the vergence stimulus and viewing distance. The black diagonal line represents natural viewing. Red and blue lines again represent the boundaries of the comfort zone. Dashed horizontal lines represent viewing distances for the same devices as in the panel *a*. Abbreviation: D, diopter.

It has been frequently observed that some individuals are more susceptible than others to visual discomfort when viewing stereoscopic displays. It would be very useful to know what makes one person susceptible and another not. Recent research has investigated whether age and properties of binocular visual function are predictive.

The range of distances over which an individual can accommodate declines steadily with age (Glasser & Campbell 1998). For people in their forties, the restricted range becomes noticeable. By their fifties and sixties, the accommodative range is effectively nil. Restricted range of accommodation is presbyopia. Presbyopes experience vergence-accommodation conflicts all the time in natural viewing because they can converge appropriately but cannot accommodate. This leads to the expectation that vergence-accommodation conflicts associated with stereoscopic viewing will not cause visual discomfort in presbyopes. Yang et al. (2012) reported data consistent with this expectation. They exposed subjects to prolonged viewing of similar content presented stereoscopically and nonstereoscopically. Young adults reported more severe symptoms after stereoscopic than nonstereoscopic viewing. But subjects 46 years of age or older reported more symptoms after nonstereoscopic viewing. This reversal of the usual outcome makes sense because stereoscopic viewing allows consistency between disparity-driven vergence and proximal vergence in presbyopes, whereas nonstereoscopic viewing puts the two in conflict.

Optometric aspects of binocular visual function have also been investigated as potential predictors of susceptibility to discomfort. Results are inconclusive. Shibata et al. (2011) examined whether gains of the vergence-accommodation cross-links [i.e., the accommodative-convergence/accommodation (AC/A) ratio and convergence-accommodation/convergence (CA/C) ratio] are predictive and observed only a weak relationship. They found that an individual's phoria was reasonably predictive of the conditions in which he/she would experience discomfort. That finding was confirmed by Shibata et al. (2013) but not by Häkkinen et al. (2006) and McIntire et al. (2014b). Others reported that the width of Panum's fusion zone is predictive (narrower zones predicting great susceptibility) (Chen et al. 2012). Lambooij et al. (2011) found that a combination of optometric parameters yielded reasonable predictions for susceptibility.

Read et al. (2016) found no evidence in children or adults of long-term effects on balance, coordination, and vision following two months of watching stereoscopic 3D television every day.

3. ADVANCED DISPLAYS

3.1. Focus Cues

It has often been stated that focus cues (blur and accommodation) do not have noteworthy effects on seeing three dimensionally. For example, Mather (2006, p. 276) stated that blur provides only "coarse ordinal (depth) information," and Mather & Smith (2000, p. 3504) stated that "blur is always treated as a relatively weak depth cue by the visual system." The claim is based on the fact that defocus blur is unsigned; the same blur can be observed when an object is nearer or farther than the distance to which the eye is focused. Because of this relatively common view, we spend some time describing evidence to the contrary: specifically, evidence that focus cues affect both 3D shape perception and the apparent scale of the scene.

Buckley & Frisby (1993) and Frisby et al. (1995) observed striking differences in 3D shape percepts when stimuli were presented as real objects versus on a conventional stereoscopic display. Their stimuli were raised ridges. When presented on a display, the depths specified by disparity and perspective were manipulated independently in the conventional way. The real objects were patterned cards wrapped onto wooden forms. They manipulated disparity by changing the shape of the form and perspective by warping the texture pattern on the card. When stimuli were presented on a display, disparity and perspective both affected the amount of perceived depth, but disparity dominated when the perspective-specified depth was large and perspective dominated when the perspective-specified depth was small. The results differed dramatically when the stimuli were presented as real objects. The disparity-specified depth now dominated the percept in all cases, and judgments were more accurate. Buckley and Frisby reasoned that focus cues played a critical role in generating the 3D percepts. When the stimuli were presented on a display, focus cues signaled flatness because the light all came from the same distance. With the real objects, focus and stereo cues both signaled the true shape. To test whether focus cues were indeed the determining factor, they redid the experiment with pinholes in front of the two eyes. Interestingly, the display data were unaffected by the pinholes, but the real-ridge data became similar to the display data. This result can be explained only by assuming that focus cues played an important role in perceiving shape. Viewing through a pinhole increased depth of field substantially. Greater depth of field does not change the retinal images arising from stimuli on a flat display, so percepts were in that case understandably not affected. Increased depth of focus does cause a change in the images from real 3D objects: blur no longer varies across the object. And that is presumably why the perception of real objects changed when viewing through pinholes. Overall, Buckley and Frisby's observations provide convincing evidence for a noteworthy effect of focus cues on 3D shape perception. Similar results have been reported elsewhere (Hoffman et al. 2008, Pentland 1987, Watt et al. 2005).

The pattern of blur in an image can also strongly influence the perceived size of scenes. For example, cinematographers make miniature models look larger by using small camera apertures, which increase depth of field (Fielding 1985). The opposite effect is created in a photographic manipulation called the tilt-shift effect: A full-size scene is made to look smaller by adding blur with either a special lens or postprocessing software tools (Held et al. 2010, Laforet 2007, Vishwanath & Blaser 2010). Scenes can also be made to look smaller by increasing the size of the effective camera aperture as demonstrated in **Figure 10**. The left image has been rendered sharply with a small aperture and looks like a life-size urban scene. The right image has been rendered with a blur pattern consistent with a much larger aperture, and it looks like a miniature model. We can

Figure 10

Effect of depth-of-field blur on apparent scale. The left image was rendered with a pinhole camera creating a long depth of field. The right image was rendered with a 60-m aperture creating a short depth of field. The scene on the right looks much smaller than the one on the left. Figure reprinted from Held et al. (2010).

understand why aperture size has such a dramatic influence on perceived scale by examining the determinants of blur in images. Consider a camera (or eye) focused on an object at distance z_0. A point at another distance z_1 creates a blurred image. The diameter of the blurred image is:

$$b = \frac{As_0}{z_0}\left|1 - \frac{z_0}{z_1}\right|, \tag{8}$$

where A is the diameter of the aperture (or pupil) and s_0 is a focal length (or eye length) term. Converting the blur circle to radians and simplifying results in

$$\beta \approx A\left|\frac{1}{z_0} - \frac{1}{z_1}\right| = A|\Delta D|, \tag{9}$$

where β is blur-circle diameter in radians and ΔD is the difference between the focused distance and object distance in diopters. The absolute value is used because defocus blur is unsigned. Thus, the amount of blur variation from a given scene is proportional to aperture size and to the difference in object distances in diopters. With an aperture of normal size (~4 mm in humans), large variations in blur can happen only if the objects are close to the eye. Recent experiments have shown that depth-of-field blur can indeed strongly affect perceived scale (Held et al. 2010, Nefs 2012, Trentacoste et al. 2011, Zhang et al. 2014).

Binocular disparity has the same basic geometry as blur because both cues are based on viewing the world from different vantage points (Held et al. 2010, Schechner & Kiryati 2000). Disparity is created by the differing vantage points of two cameras or eyes; blur is created by differing vantage points across the camera's or eye's aperture. The disparity in radians between the point on which the cameras (or eyes) are converged and another point is

$$\delta \approx I(\Delta D), \tag{10}$$

where I is interocular separation. Thus, the magnitudes of blur and disparity caused by a point in a 3D scene should be proportional to one another. In humans, pupil diameter is ~1/12 the distance between the eyes, so blur-circle diameters are generally 1/12 the magnitudes of disparities. Because

the geometries underlying disparity and blur are similar, the basic relationship holds for the viewing of all real scenes. This natural disparity-blur relationship should be obeyed if one wants the scene to appear natural in scale.

Visual performance can also suffer when focus cues are not presented correctly on a stereoscopic display. For example, many viewers find it difficult to fuse a binocular stimulus when vergence and accommodative distances differ substantially. Specifically, the time required to fuse a stimulus decreases monotonically when the conflict between vergence and accommodative distances decreases (Akeley et al. 2004, Hoffman et al. 2008). The minimum time occurs with zero conflict. This effect is surely a consequence of the vergence-accommodation coupling. The coupling is beneficial when there is no conflict between vergence and accommodative distances because disparity-driven and blur-driven vergence are in agreement, so vergence responds rapidly and accurately. The coupling is not beneficial when there is a conflict because the disparity- and blur-driven components of vergence attempt to drive vergence to different distances. Other performance decrements due to vergence-accommodation conflict have been reported (Lambooij et al. 2012).

One approach for minimizing the conflict is to remap the disparities of the stereoscopic content so that they do not deviate greatly from zero at the screen in the region of interest (Didyk et al. 2014, Lang et al. 2010). Doing this, however, is very likely to distort 3D percepts.

In summary, there is now substantial evidence that incorrect focus cues (blur and accommodation) can alter 3D shape and scene perception, limit visual performance, and cause discomfort. For these reasons, there has been considerable effort recently toward developing and evaluating displays that support correct focus cues.

3.2. Volumetric and Multi-Plane Displays

Volumetric displays place light sources in a 3D volume by, for example, using rotating display screens (Favalora et al. 2002) or stacks of switchable diffusers (Sullivan 2004). These allow correct stereo, parallax, and focus cues, but the displayed scene is confined to the display volume. Furthermore, a very large number of addressable voxels is required, which places practical limits on effective resolution at the viewer's eye. Importantly, these displays cannot reproduce occlusions and viewpoint-dependent effects such as reflections. Instead, they show a world of glowing, transparent voxels. Recent techniques (Cossairt et al. 2007, Jones et al. 2007) have used anisotropic diffusers to overcome this limitation, but focus cues then become incorrect.

Multi-plane displays are a variant of volumetric displays in which the viewpoint is fixed. Such displays are promising because they can in principle provide correct depth cues, including focus cues, with conventional display hardware. In multi-plane displays, images are drawn on presentation planes at different focal distances. Most multi-plane displays have separate optical paths for the two eyes, so stereo cues and focus cues are both enabled. These displays have been constructed using a system of beam splitters (Akeley et al. 2004, MacKenzie et al. 2010) and by time multiplexing with high-speed switchable lenses (Liu et al. 2008, Love et al. 2009) to superimpose multiple presentation planes additively on the viewer's retina. Most current implementations support high-resolution imagery by presenting the full resolution of a conventional monitor at each focal distance. A multi-plane head-mounted version has been described (Hu & Hua 2013, 2014).

With multi-plane displays, a simulated object at one of the presentation planes is displayed by illuminating the pixels on that plane. A rule is, however, required to determine how to illuminate pixels in the more likely case that the simulated object is not at the distance of a presentation plane. This is accomplished by distributing intensities on the pixels of adjacent planes by linear interpolation (Akeley et al. 2004) or somewhat more complicated rules (Hu & Hua 2014, Liu & Hua 2010). This approach works well when the presentation planes are not too far apart (less

than ~0.7 D), the simulated stimuli are diffuse surfaces, and depth varies slowly across the image (Ravikumar et al. 2011).

But the approach produces objectionable haloing artifacts around occlusions. Reflections, refractions, and other non-Lambertian phenomena also produce image features that cannot be assigned a consistent focal distance and therefore cannot be handled by these interpolation rules. A novel optimized blending algorithm was recently developed that creates much more realistic results for occlusions and non-Lambertian effects (Narain et al. 2015). Using a model of image formation in the typical human eye, the researchers obtained for each desired viewpoint the stack of retinal images that would be experienced by a viewer when accommodating to different distances. They then optimized the assignment of light intensities to presentation planes so that the resultant retinal images were as close as possible to the retinal images that would be produced by the original scene. The outcome looks much more similar to the real world.

The presentation of nearly correct focus cues using multi-plane displays has several benefits. First, multi-plane displays support accommodation. With monocular viewing, subjects accommodate to simulated objects even when they are positioned between presentation planes. Accommodation is quite accurate when the planes are 0.67 D apart and reasonably accurate when the planes are 1 D apart (MacKenzie et al. 2010, 2012). Second, the time to fuse a binocular stimulus is reduced in such displays when the vergence and accommodation distances are the same (Akeley et al. 2004, Hoffman et al. 2008, Maiello et al. 2014). Third, reducing the vergence-accommodation conflict leads to a reduction in visual discomfort. Finally, the presentation of nearly correct focus cues improves the accuracy of 3D percepts. An interesting case is the perception of depth ordering at an occlusion boundary. In earlier studies using conventional displays, viewers were presented two abutting surfaces, one with a blurred texture and one with a sharp one (Marshall et al. 1996, Mather & Smith 2002, Palmer & Brooks 2008). Viewers indicated which of the two surfaces was nearer—in other words, which was the occluder and which was the background. The blur of the boundary between the surfaces is a completely informative cue. When the boundary is blurred, the blurrier surface is the occluder; when the boundary is sharp, the sharper surface is the occluder. Despite this informative cue, viewers were quite poor in making depth-order judgments. The stimuli in those experiments were rendered for and displayed on a single plane. Zannoli et al. (2016) repeated those studies with single- and multi-plane stimuli (**Figure 11**). Subjects viewed the stimuli monocularly. With single-plane stimuli, subjects performed barely above chance, consistent with the previous findings. When the same occlusion relationship was displayed on the multi-plane display with optimized blending, performance improved markedly. Subjects generally perceived the physically nearer surface as nearer even though they viewed the stimuli monocularly with no motion parallax and at durations too brief for an accommodative change to occur (**Figure 11**). Thus, multi-plane displaying with optimized rendering significantly improves the perception of depth order at occlusion boundaries.

3.3. Light-Field Displays

The natural environment contains light rays of diverse intensities and colors running in various directions. The light field is a function that describes the amount and color of light flowing in every direction through every point in space (Gershun 1939). A binocular viewer moving through the light-field experiences the triangulation cues to which we referred earlier. Stereo cues are created because the two eyes receive different light rays. Motion parallax cues are created because the eyes receive different rays as the viewer's head moves. And focus cues are created because different parts of the viewer's pupil receive different light rays. Light-field displays are intended

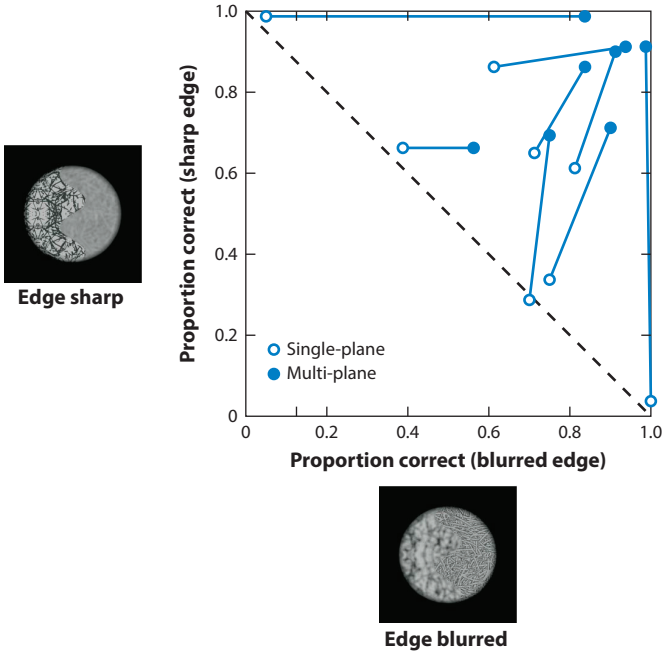

Figure 11

Judgments of depth order at an occlusion boundary. Two surfaces were shown, one with a blurred texture and one with a sharp texture. They were presented in two ways: single-plane display in which the blur was rendered and multi-plane display in which the blur was created in the viewer's eye. The single-plane results are represented by the open symbols and multi-plane results by the filled symbols. The lines connect the results for each subject. The horizontal axis is the proportion of correct responses when the boundary was blurred, and the vertical axis is the proportion correct when the boundary was sharp. The dashed diagonal line represents chance performance with different biases.

to recreate those cues with sufficient accuracy to enable a high-fidelity, comfortable 3D viewing experience.

A light-field display emits a four-dimensional distribution of light rays, which varies over the two dimensions of a display surface but also over the horizontal and vertical viewing angle of each pixel. The display primitives of conventional displays are 2D pixels (picture elements), those of volumetric or multi-plane displays are 3D voxels (volume elements), and those of 4D light-field displays are light rays, each carrying radiance at some location into a specific direction. **Figure 12** illustrates the common two-plane parameterization of a light field: A plane x is located on the physical display screen, and another plane u coincides with the pupils of the viewer. To pass our Turing test for displays—that is, to create a sufficiently persuasive 3D experience—a 4D light-field display would have to provide appropriate stereo, motion parallax, and focus cues. No such display exists today, but different tradeoffs can be made to create reasonable approximations.

Over a century ago, Ives (1903) conceived of parallax barriers. A barrier mask consisting of an array of pinholes or slits would be mounted at a slight offset in front of a display such that a viewer would perceive only a subset of the display pixels from any given perspective. The display would render an image that contains the corresponding, interlaced perspectives of the light field. Soon after, Lippmann (1908) built the first integrated light-field camera and display using integral imaging. Instead of pinhole arrays, he mounted microlens arrays on photographic plates and

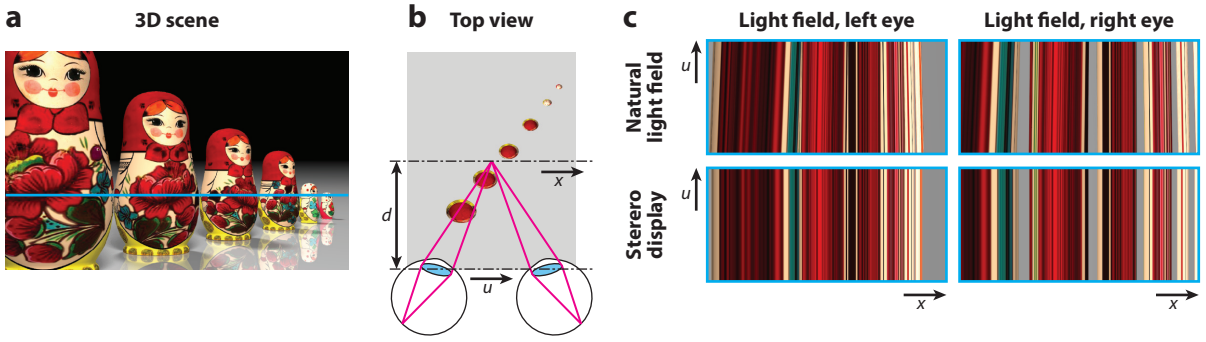

Figure 12

(*a*) The light field of a natural scene (*b*) is a collection of rays parameterized by their coordinates of intersection with two planes *x* and *u*. (*c*) All rays on a horizontal scanline (*cyan*) observed in the centers of the viewer's two pupils are shown on the lower right (stereo display) and all rays on the same scanline across the viewer's pupils are shown on the upper right (natural light field). The two eyes observe the scene from different vantage points so the left- and right-eye rays differ. Conventional stereoscopic displays do not provide parallax across the pupils and therefore do not support focus cues. The natural light field does provide parallax across the pupils and thereby provides focus cues.

exposed and developed these plates with the lens arrays in place, such that they could be viewed as a light-field or glasses-free 3D image after the fact. The main drawback of parallax barriers and integral imaging is the spatio-angular resolution tradeoff: Adding more light-field viewing zones comes at the cost of reduced spatial display resolution. Additionally, parallax barriers are usually dim because most of the emitted light is blocked. To overcome these limitations, many alternative technologies have emerged over the last 100 years to deliver high-resolution, glasses-free 3D experiences, as discussed in the previous sections. Yet, none can deliver experiences that come close to satisfying our Turing test for displays.

With an ever-increasing demand on image resolution, one of the major bottlenecks in the light-field display pipeline is the computation. Consider the example of a high-quality, light-field display with 100 × 100 views, each having high-definition resolution, streamed at 60 Hz. More than one trillion light rays have to be rendered per second requiring more than 100 terabytes of floating point RGB ray data to be stored and processed. Furthermore, with conventional integral imaging or parallax barriers, a display panel that has a resolution 10,000 times higher than available panels would be needed. To tackle the big data problem and relax requirements on display hardware, compressive light-field displays have recently been introduced. They rely on two insights: light fields of natural imagery are highly redundant, high-dimensional visual signals; and the human visual system has limitations that can be exploited for visual signal compression.

Multiplexing methods (e.g., temporal, spatial, polarization) can be adopted to optimize the tradeoff between spatial and angular resolution, brightness, etc., dynamically in a content-adaptive manner. For example, the refresh rate of modern displays often exceeds the flicker threshold of human vision. A parallax-barrier display implemented with fast liquid-crystal display (LCD) would allow for the optimal layout of time-multiplexed pinholes to be determined for each target light field. Further relaxing the requirement that the barrier mask is constrained to showing only pinholes leads to the concept of content-adaptive parallax barriers that optimize the time-multiplexed patterns for both display and barrier mask (Lanman et al. 2010). Such a content-adaptive optimization not only allows adaptive tradeoffs between spatial and angular resolution to be made, but it also allows for display brightness to be optimized with respect to pinhole-based barriers. The light field generated by a time-multiplexed parallax barrier displays with two LCDs

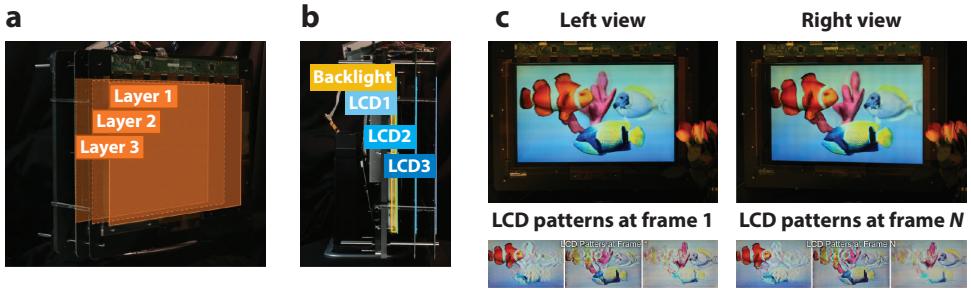

Figure 13

Compressive light-field prototype. (*a*) The prototype uses three stacked layers of liquid-crystal displays (LCDs) that are rear-illuminated by one backlight. (*b*) Top-down view of prototype display showing the LCD layers and backlight. (*c*) A light-field factorization algorithm computes time-multiplexed patterns for all LCD layers. The bottom subpanels show the three layers for the left and right views. They are displayed at a speed exceeding the flicker threshold of the visual system. Perceptually, these patterns fuse into a consistent, high-resolution light field that supports stereo cues and parallax without needing glasses.

is given by

$$\tilde{l}(x,u) = \frac{1}{M}\sum_{m=1}^{M} f_m^{(1)}\left(x + \frac{d_1}{d}(u-x)\right) f_m^{(2)}\left(x + \frac{d_2}{d}(u-x)\right) f_m^{(1)}, \tag{11}$$

where M is number of time-multiplexed images that the visual system perceptually averages and d_1 and d_2 are the distances between LCD layers 1 and 2 and the x-plane, respectively. Given a target light field $l(x,u)$, a least-squared error approximation of it can be physically reproduced with the parallax-barrier display by computing the best set of time-multiplexed patterns $f_m^{(1)}$ and $f_m^{(2)}$ and displaying them on the respective LCD panels. The corresponding inverse problem is usually formulated as a numerical optimization problem,

$$\min\left\{f_m^{(1)}, f_m^{(2)}\right\} \left\| l(x,u) - \tilde{l}(x,u)\right\|_F^2, \text{ subject to } 0 \le f_m^{(1)}, f_m^{(2)} \le 1, \tag{12}$$

which can be efficiently solved with nonnegative matrix factorization approaches (Lanman et al. 2010).

Compressive light-field displays generalize the idea of content-adaptive parallax barriers to a variety of display architectures, including multiple stacked layers of LCDs (**Figure 13**), a thin "sandwich" of two LCDs enclosing a microlens array or, in general, any combination of stacked, programmable light modulators and refractive optical elements (Wetzstein et al. 2012). Similar to parallax barriers, cascading LCDs usually have a multiplicative effect on the incident light that can selectively attenuate light in some directions (Huang et al. 2015; Lanman et al. 2010; Maimone et al. 2013; Wetzstein et al. 2011, 2012). The light-field factorization outlined above generalizes to all of these display architectures. Their nonlinear, multiplicative image formation is fundamentally different from the linear, additive image formation provided by multi-focal plane displays, volumetric displays, and many other time-multiplexed displays. In general, a nonlinear image formation has the potential to provide more degrees of freedom for the image generation algorithm than an additive, linear image formation (Huang et al. 2015, Maimone et al. 2013, Wetzstein et al. 2012).

A useful approach for characterizing the degrees of freedom of light-field displays is a frequency analysis. For this, we make one simplifying assumption: Instead of using an absolute two-plane parameterization on the x-u planes, we use a relative two-plane parameterization x-v, such that $l(x,u) = l(x, x + dv)$.

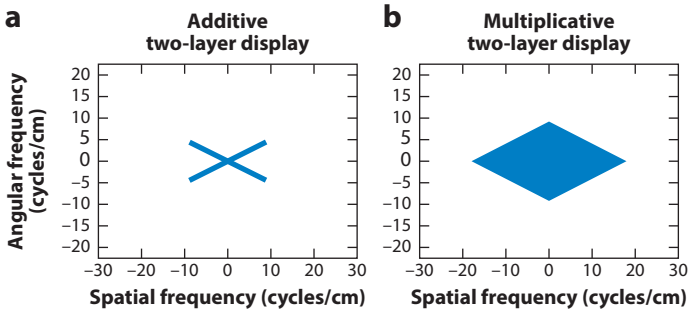

Figure 14

Spatio-angular frequency support of (*a*) an additive two-layer display and (*b*) a corresponding multiplicative two-layer display. The specific display parameters, including layer spacing and panel resolution, are adopted from Wetzstein et al. (2012). The spatio-angular frequency support of a conventional display is a slanted line; the frequency support of two additive displays is their combined support (*a*). If the same display layers interact in a multiplicative manner, their frequency support is convolved and covers a significantly larger area (*b*). The support shown for the multiplicative display is a theoretical upper bound that is not easily achieved in practice.

Real-world light fields contain all spatio-angular frequencies. Conventional 2D displays are a poor approximation: From each pixel, light is emitted in all directions, so such displays do not provide any angular variation. The Fourier transform of the light field emitted by a conventional display is therefore a line: $\hat{f}_m(\omega_x, \omega_v) = \delta\left(\omega_v - \frac{d_0}{d}\omega_x\right)$ (Wetzstein et al. 2011, 2012), which falls well short of representing the span of spatio-angular frequencies in natural light fields. The frequency support of additive, multi-plane displays is the sum of the support provided by each plane. Thus, display layers at different distances support a larger range of frequencies than a conventional display. This is illustrated for two layers in **Figure 14*a***. The frequency support for two layers is sparse but can be improved simply by adding more closely spaced layers. Adding more layers approximates a volumetric display and can provide adequate coverage of the required span of spatio-angular frequencies. Before the design specifications can be understood, however, we need better models of human sensitivity to angular resolution.

The spatio-angular frequencies created by the multiplication of two display layers are given by the convolution of the Fourier transforms of the individual layers:

$$\hat{I}(\omega_x, \omega_v) = \frac{1}{M}\sum_{m=1}^{M} \hat{f}_m^{(1)}(\omega_x)\delta(\omega_v - \frac{d_1}{d}\omega_x) * \sum_{m=1}^{M} \hat{f}_m^{(2)}(\omega_x)\delta(\omega_v - \frac{d_2}{d}\omega_x). \quad (13)$$

Figure 14*b* shows the theoretical upper bounds on the spatio-angular frequencies that can be created by a two-layer, multiplicative display. The multiplicative approach theoretically provides greater frequency support than the additive approach does. The estimated frequency support for multiplicative displays is, however, a theoretical upper bound that may be difficult to achieve in practice. One practical limitation is nonnegativity constraints on pixel states. Another is time-multiplexing constraints due to the noninfinite speed of display panels. Finally, the resolution of multiplicative displays is fundamentally limited by diffraction. Observing one display panel through another one that has small pixels makes the farther panel appear blurred due to diffraction. The smaller the pixels become, the more obvious this becomes. If pixel size approached the wavelength of light, the blur due to diffraction would be quite problematic. **Figure 15** illustrates the diffraction problem for a range of panel resolutions and distances between the panels. Here, we model monochromatic light with a wavelength of 555 nm. We assume viewing zones that are at most as

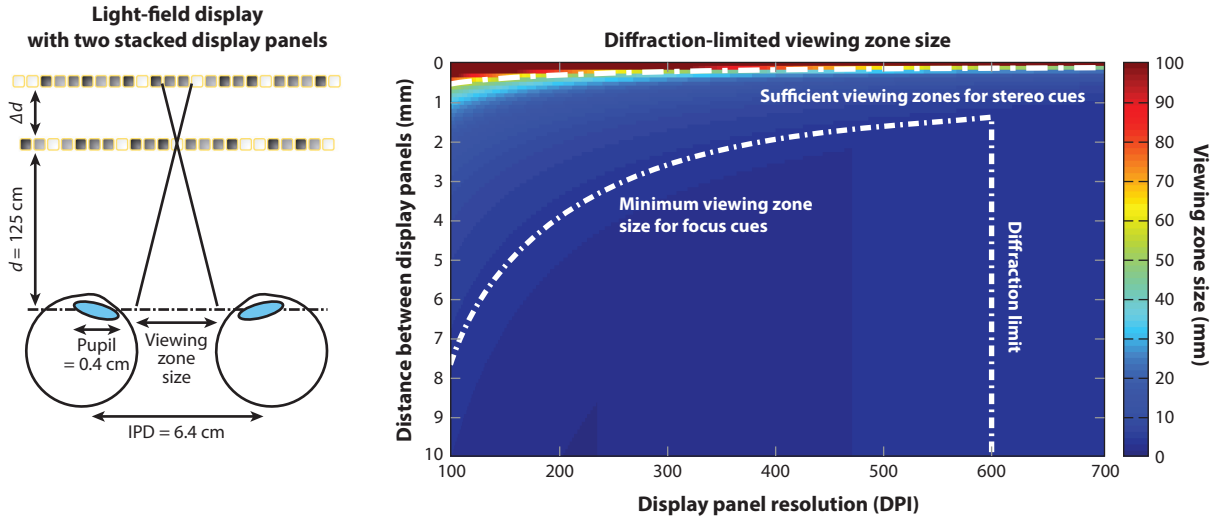

Figure 15

Diffraction-limited viewing zone size for two stacked liquid-crystal display layers. (*Left*) We simulate a generalized parallax barrier display setup with the front panel being 125 cm from the viewer. The distance between panels as well as their resolution is varied for the simulation on the left. For an appropriate choice of screen resolution and screen distance, stereo cues are easily achieved when the size of the viewing zone is larger than the interpupillary distance. When more than two views enter the same pupil (here, simulated with a diameter of 4 mm), accommodation is theoretically possible. However, the viewing zone size of parallax barriers and other multilayer displays is fundamentally limited by diffraction as shown in the right panel. Abbreviations: DPI, dots per inch; IPD, interpupillary distance.

large as the interpupillary distance to support stereo cues. To support focus cues, several viewing zones must enter the same pupil, which in the simulation has a diameter of 4 mm. To provide sufficiently dense viewing zones, either the parallax barriers can be spaced farther apart or panels with higher pixel density can be employed. For a given observer distance, we can compute the viewing zone size with a geometric optics approximation. Diffraction effects can be accounted for by simulating the panel closest to the viewer as an array of apertures with feature size p given by display pixel pitch. The diffraction-limited spot size on the rear panels can be approximated as the first minimum of the Airy disk resulting from diffraction shown in the equation

$$2\Delta d \tan\left(\sin^{-1}\left(1.22\frac{\lambda}{p}\right)\right), \tag{14}$$

where Δd is the distance between the panels. We observe that the viewing zones designed to support stereo cues are usually not affected much by diffraction, but for those targeting focus cues, diffraction limits must be carefully considered.

The frequency support for additive displays is more readily expanded in practice. For example, higher panel resolution extends the frequency lines outward; adding more densely spaced layers fills in the span, thereby covering more frequencies. The blurring effect due to diffraction is generally not observed with additive displays. In summary, the multiplicative approach can theoretically support a larger set of spatio-angular frequencies than the additive approach. But practical considerations currently limit the performance of multiplicative displays.

Compressive light-field displays have proven quite promising. Their primary benefit is that the same display hardware configuration allows content-adaptive dynamic tradeoffs between spatial

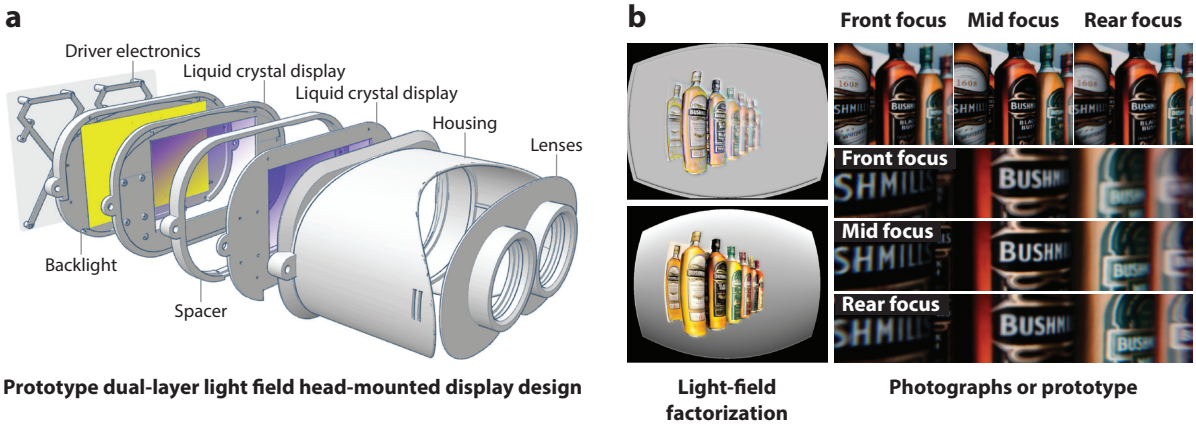

Figure 16

Near-eye light-field display with support for focus cues. (*a*) The design is based on the stereoscope but uses two stacked liquid-crystal displays inside the device to generate a separate light field for each eye. (*b*) Light-field factorization is employed to generate patterns shown on the front and rear panels. (*c*) The light field provides sufficient angular resolution for focus cues to be supported, so the viewer can accommodate within the virtual scene.

and angular resolution as well as image brightness, sweet-spot location, and crosstalk. A main challenge for a compressive display, however, is the need for advanced computational processing (i.e., real-time light-field factorization), which adds to the system complexity.

With increasing understanding of the benefits and limitations of generalized parallax barriers, other display applications are being explored. For example, light-field projection systems supporting parallax and stereo cues have emerged (Balogh 2006, Jones et al. 2014). These display systems are most suitable for collaboration among viewers and provide impressive image quality over large depths of field. Unfortunately, dozens of projectors have to be employed, making such display systems expensive, difficult to calibrate, power hungry, and bulky. The compressive light-field methodology can also be applied to projection systems (Hirsch et al. 2014). In this case, the goal is to compress the number of required devices. Hirsch et al. demonstrated that this is possible by generating a light field inside a single projection device, via content-adaptive parallax barriers, and then optically amplifying the limited field of view of the emitted light field using a screen composed of an array of microscopic Keplerian telescopes, one in each screen pixel.

Supporting focus cues have recently attracted a lot of attention with the renewed interest in near-eye displays for virtual reality (VR) and augmented reality (AR). Most near-eye displays provide stereo cues because they have either two separate microdisplays, one for each eye, or a single screen that is optically split with two lenses. Lanman & Luebke (2013) demonstrated an integral-imaging-type near-eye display that allowed for a limited accommodation but reduced display resolution. More recently, Huang et al. (2015) investigated high-resolution compressive light-field synthesis via two stacked LCDs (**Figure 16**). The device design is inspired by conventional stereoscopes, but employs two LCD panels with 1-cm spacing. Using factorization algorithms similar to those employed by the content-adaptive parallax barriers described above, a 4D light field is emitted independently to each eye, providing parallax over the eye box. Without eye tracking, the pupil can freely move within the eye box and, as long as multiple different perspectives enter that pupil simultaneously, focus cues can be produced. The frequency and diffraction analyses described above apply to near-eye displays.

Another application of light-field displays is correction of visual aberrations for a human observer (Huang et al. 2014). Instead of correcting vision with eyeglasses or contact lenses, the same can potentially be done directly in the screen, allowing for myopia, hyperopia, astigmatism, and even higher-order aberrations to be corrected. For such an application, the light-field display presents a distorted light field to the eyes of the viewer such that their aberrations optically undistort them, resulting in the desired image.

4. CONCLUSIONS

Significant advances have been made in developing displays that create a compelling experience of three dimensionality. We may be nearing the time when a display could pass a 3D version of the Turing test: Can you distinguish the 3D scene geometry you perceive from an advanced display from the geometry you perceive when viewing the real world?

SUMMARY POINTS

1. Different types of displays use different depth cues to create 3D experiences.
2. Significant advances have been made in how to incorporate stereoscopic and motion-parallax information in such displays. But ergonomic and perceptual problems persist.
3. An understanding of the human visual system is crucial to designing future displays. Such an understanding guides the appropriate generation of visual inputs that affect perception while not generating inputs that do not affect perception.

FUTURE ISSUES

The most significant issue that remains to be solved is how to generate appropriate focus cues (blur and accommodation) in practical 3D displays.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Akeley K, Watt SJ, Girshick AR, Banks MS. 2004. A stereo display prototype with multiple focal distances. *ACM Trans. Graph.* 23:I804–13

Backus BT, Banks MS, van Ee R, Crowell JA. 1999. Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vis. Res.* 39:1143–70

Balogh T. 2006. The HoloVizio system. *Proc. SPIE* 6055:60550U

Banks MS, Gepshtein S, Landy MS. 2004. Why is spatial stereoresolution so low? *J. Neurosci.* 24(9):2077–89

Banks MS, Held RT, Girshick AR. 2009. Perception of 3-D layout in stereo displays. *Inf. Disp.* 25:12–16

Bereby-Meyer Y, Leiser D, Meyer J. 1999. Perception of artificial stereoscopic stimuli from an incorrect viewing point. *Percept. Psychophys.* 61:1555–63

Blake R, Fox R. 1973. The psychophysical inquiry into binocular summation. *Percept. Psychophys.* 14:161–85

Blake R, Sloane M, Fox R. 1981. Further developments in binocular summation. *Percept. Psychophys.* 30:266–76

Blondé L, Sacré J-J, Doyen D, Huynh-Thu Q, Thébault C. 2011. Diversity and coherenece of 3D crosstalk measurements. *SID Symp. Digest Tech. Pap.* 42:804–7

Borel T, Doyen D. 2013. 3D display technologies. In *Emerging Technologies for 3D Video: Creation, Coding, Transmission, and Rendering*, ed. F Dufaux, B Pesquet-Popescu, M Cagnazzo. New York: Wiley

Brainard DH, Pelli DG, Robson T. 2002. Display characterization. In *Encylopedia of Imaging Science and Technology*, ed. J Hornak. New York: Wiley

Buckley D, Frisby JP. 1993. Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vis. Res.* 33(7):919–33

Burr DC, Ross J. 1979. How does binocular delay give information about depth? *Vis. Res.* 19:523–32

Cakmakci O, Rolland J. 2006. Head-worn displays: a review. *J. Disp. Technol.* 2:199–216

Campbell FW, Green DG. 1965. Monocular versus binocular visual acuity. *Nature* 208:191–92

Campbell FW, Robson JG. 1968. Application of Fourier analysis to the visibility of gratings. *J. Physiol.* 197:551–66

Cavonius CR. 1979. Binocular interaction in flicker. *Q. J. Exp. Psychol.* 31:273–80

Chen Z, Shi J, Tai Y. 2012. An experimental study on the relationship between maximum disparity and comfort disparity in stereoscopic video. *Proc. SPIE* 8556:855608

Collewijn H, Van der Steen J, Ferman L, Jansen TC. 1985. Human ocular counterroll: assessment of static and dynamic properties from electromagnetic scleral coil recordings. *Exp. Brain Res.* 59:185–96

Cossairt OS, Napoli J, Hill SL, Dorval RK, Favalora GE. 2007. Occlusion-capable multiview volumetric three-dimensional display. *Appl. Opt.* 46(8):1244–50

Cowan M. 2008. *Real D 3D theatrical system: a technical overview*. European Digital Cinema Forum, April 24. **http://www.edcf.net/articles.html**

Cumming BG, Judge SJ. 1986. Disparity-induced and blur-induced convergence eye movement and accommodation in the monkey. *J. Neurophysiol.* 55:896–914

Dawson S. 2012. Passive 3D from the beginning. *HiFi Writer Blog*, June 3. **http://hifi-writer.com/wpblog/?p=3797**

Didyk P, Ritschel T, Eisemann E, Myszkowski K, Seidel H-P. 2014. A perceptual model for disparity. *ACM Trans. Graph.* 30:96

Dodgson NA. 2006. On the number of viewing zones required for head-tracked autostereoscopic display. *Proc. SPIE* 6055:60550Q

Edwards L. 2009. *Active shutter 3D technology for HDTV*. Phys.org, Sept. 25. **http://phys.org/news173082582.html**

Elkins DE. 2013. *The Camera Assistant's Manual*, p. 21. Burlington, MA: Focal Press

Emoto M, Niida T, Okano F. 2005. Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television. *J. Disp. Technol.* 1(2):328–40

Fairchild MD, Wyble DR. 2007. Mean observer metamerism and the selection of display primaries. *Proc. Soc. Imaging Sci. Technol., Albuquerque, NM, November*, pp. 151–56

Favalora GE, Napoli J, Hall DM, Dorval RK, Giovinco M, et al. 2002. 100-million-voxel volumetric display. *Proc. SPIE* 4712:300

Fielding R. 1985. *Techniques of Special Effects Cinematography*. Oxford, UK: Focal Press. 4th ed.

Fincham EF, Walton J. 1957. The reciprocal actions of accommodation and convergence. *J. Physiol.* 137:488–508

Formankiewicz MA, Mollon JD. 2009. The psychophysics of detecting binocular discrepancies of luminance. *Vis. Res.* 49(15):1929–38

Frisby JP, Buckley D, Horsman JM. 1995. Integration of stereo, texture, and outline cues during pinhole viewing of real ridge-shaped objects and stereograms of ridges. *Perception* 24:181–98

Fry G. 1939. Further experiments on the accommodative convergence relationship. *Am. J. Optom.* 16:325–34

Fukushima T, Torii M, Ukai K, Wolffsohn JS, Gilmartin B. 2009. The relationship between CA/C ratio and individual differences in dynamic accommodative responses while viewing stereoscopic images. *J. Vis.* 9(13):21

Gershun A. 1939. *The Light Field*, transl. P Moon, G Timoshenko. *J. Math. Phys.* 18:51–151

Glasser A, Campbell MC. 1998. Presbyopia and the optical changes in the human crystalline lens with age. *Vis. Res.* 38(2):209–29

Hakala JH, Oittinen P, Häkkinen JP. 2015. Depth artifacts caused by spatial interlacing in stereoscopic 3D displays. *ACM Trans. Appl. Percept.* 12(1):3

Häkkinen J, Pölönen M, Takatalo J, Nyman G. 2006. Simulator sickness in virtual display gaming: a comparison of stereoscopic and non-stereoscopic situations. In *Proc. 8th Conf. Hum.-Comp. Interact. Mob. Devices Serv.*, pp. 227–30. New York: ACM

Hands P, Smulders TV, Read JCA. 2015. Stereoscopic 3-D content appears relatively veridical when viewed from an oblique angle. *J. Vis.* 15(5):6

Heath GG. 1956. Components of accommodation. *Am. J. Optom. Arch. Am. Acad. Optom.* 33(11):569–79

Held RT, Banks MS. 2008. Misperceptions in stereo displays: a vision science perspective. *Proc. 5th Symp. Appl. Percept. Graph. Vis.*, pp. 23–32. New York: ACM

Held RT, Cooper EA, Banks MS. 2012. Blur and disparity are complementary cues to depth. *Curr. Biol.* 22:426–31

Held RT, Cooper EA, O'Brien JF, Banks MS. 2010. Using blur to affect perceived distance and size. *ACM Trans. Graph.* 29(2):19

Held RT, Hui TT. 2011. A guide to stereoscopic 3D displays in medicine. *Acad. Radiol.* 18:1035–48

Hirsch M, Wetzstein G, Raskar R. 2014. A compressive light field projection system. *ACM Trans. Graph.* 33:58

Hoffman DM, Girshick AR, Akeley K, Banks MS. 2008. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *J. Vis.* 8(3):33

Hoffman DM, Johnson PV, Kim J, Vargas AD, Banks MS. 2014. 240 Hz OLED technology properties that can enable improved image quality. *J. Soc. Inf. Disp.* 22(7):346–56

Hoffman DM, Karasev VI, Banks MS. 2011. Temporal presentation protocols in stereoscopic displays: flicker visibility, perceived motion, and perceived depth. *J. Soc. Inf. Disp.* 19(3):271–97

Hofstetter HW. 1945. The zone of clear single binocular vision. *Am. J. Optom. Physiol. Opt.* 22:361–84

Howard IP, Allison RS, Zacher JE. 1997. The dynamics of vertical vergence. *Exp. Brain Res.* 116:153–59

Howarth PA. 2011. Potential hazards of viewing 3-D stereoscopic television, cinema and computer games: a review. *Ophthalmic Physiol. Opt.* 31:111–22

Hu X, Hua H. 2013. An optical see-through multi-focal-plane stereoscopic display prototype enabling nearly correct focus cues. *Proc. SPIE* 8648:86481A

Hu X, Hua H. 2014. Design and assessment of a depth-fused multi-focal-plane display prototype. *J. Disp. Technol.* 10(4):308–16

Huang F-C, Chen K, Wetzstein G. 2015. The light field stereoscope: immersive computer graphics via factored near-eye light field displays with focus cues. *ACM Trans. Graph.* 34:60

Huang F-C, Wetzstein G, Barsky BA, Raskar R. 2014. Eyeglasses-free display: towards correcting visual aberrations with computational light field displays. *ACM Trans. Graph.* 33:59

Inoue T, Ohzu H. 1997. Accommodative responses to stereoscopic three-dimensional display. *Appl. Opt.* 36(19):4509–515

Ives FE. 1903. *Parallax stereogram and process of making same*. US Patent No. 725,567

Johnson PV, Kim J, Banks MS. 2015b. Stereoscopic 3D display technique using spatiotemporal interlacing has improved spatial and temporal properties. *Opt. Expr.* 23(7):9252–75

Johnson PV, Kim J, Hoffman DM, Vargas AD, Banks MS. 2015a. Motion artifacts on 240-Hz OLED stereoscopic 3D displays. *J. Soc. Inf. Disp.* 22(8):393–403

Johnson PV, Parnell JA, Kim J, Saunter CD, Love GD, Banks MS. 2016. Dynamic lens and monovision 3D displays to improve viewer comfort. *Opt. Expr.* 24(11):11808–27

Jones A, McDowall I, Yamada H, Bolas M, Debevec P. 2007. Rendering for an interactive 360° light field display. *ACM Trans. Graph.* 26:40

Jones A, Nagano K, Liu J, Busch J, Yu X, et al. 2014. Interpolating vertical parallax for an autostereoscopic three-dimensional projector array. *J. Electron. Imaging* 23:011005

Jordan JR, Geisler WS, Bovik AC. 1990. Color as a source of information in the stereo correspondence process. *Vis. Res.* 30:1955–70

Jorke H, Simon A, Fritz M. 2009. Advanced stereo projection using interference filters. *J. Soc. Inf. Disp.* 17(5):407–10

Julesz B, White B. 1969. Short term memory and the Pulfrich phenomenon. *Nature* 222:639–41
Kane D, Guan P, Banks MS. 2014. The limits of human stereopsis in space and time. *J. Neurosci.* 34(4):1397–408
Kane D, Held RT, Banks MS. 2012. Visual discomfort with stereo 3D displays when the head is not upright. *Proc. SPIE* 8288:828814
Kelley EF. 2011. Resolving resolution. *Inf. Displ.* 27(9):18–21
Kelly DH. 1972. Flicker. In *Handbook of Sensory Physiology*, pp. 273–302, ed. D Jameson, LM Hurvich. Berlin, Ger.: Springer Berlin Heidelberg
Kelly DH. 1979. Motion and vision. II. Stabilized spatio-temporal threshold surface. *J. Opt. Soc. Am.* 69:1340–49
Kim D, Jung YJ, Han Y, Choi J, Kim E, et al. 2014. fMRI analysis of excessive binocular disparity on the human brain. *Int. J. Imaging Syst. Technol.* 24(1):94–102
Kim J, Johnson PV, Banks MS. 2014a. Stereoscopic 3D display with color interlacing improves perceived depth. *Opt. Expr.* 22(26):31924–34
Kim J, Kane D, Banks MS. 2014b. The rate of change of vergence-accommodation conflict affects visual discomfort. *Vis. Res.* 105:159–65
Kim JS, Banks MS. 2012. Effective spatial resolution of temporally and spatially interlaced stereo 3D televisions. *SID Symp. Digest Tech. Pap.* 43(1):879–82
Kim S-K, Yoon K-H, Yoon SK, Ju H. 2015. Parallax barrier engineering for image quality improvement in an autostereoscopic 3D display. *Opt. Expr.* 23:13230–44
Klompenhouwer MA. 2006. *Flat panel display signal processing*. PhD dissertation, Eindhoven University, Neth.
Konrad R, Cooper EA, Wetzstein G. 2016. Novel optical configurations for virtual reality: evaluating user preference and performance with focus-tunable and mono vision near-eye displays. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 1211–20. New York: ACM
Kooi FL, Toet A. 2004. Visual comfort of binocular and 3D displays. *Displays* 25:99–108
Krishnan VV, Phillips S, Stark L. 1973. Frequency analysis of accommodation, accommodative vergence and disparity vergence. *Vis. Res.* 13:1545–54
Krishnan VV, Shirachi D, Stark L. 1977. Dynamic measures of vergence accommodation. *Am. J. Optom. Physiol. Opt.* 54:470–73
Krol JD, van de Grind WA. 1983. Depth from dichoptic edges depends on vergence tuning. *Perception* 12:425–38
Kubovy M. 1986. *The Psychology of Perspective and Renaissance Art*. New York: Cambridge Univ. Press
Kuroki Y. 2012. Improvement of 3D visual image quality by using high frame rate. *J. Soc. Inf. Disp.* 20:566–74
Laforet V. 2007. A really big show. *New York Times*, May 31
Lambooij M, Fortuin MF, IJsselsteijn WA, Heynderickx I. 2012. Reading performance as screening tool for visual complaints from stereoscopic content. *Displays* 33(2):84–90
Lambooij M, Fortuin M, IJsselsteijn WA, Evans BJW, Heynderickx I. 2011. Susceptibility to visual discomfort of 3-D displays by visual performance measures. *IEEE Trans. Circuits Syst. Video Technol.* 21(12):1913–23
Lambooij M, IJsselsteijn W, Fortuin M, Heynderickx I. 2009. Visual discomfort and visual fatigue of stereoscopic displays: a review. *J. Imaging Sci. Technol.* 53:1–14
Lang M, Hornung A, Wang O, Poulakos S, Smolic A, Gross M. 2010. Nonlinear disparity mapping for stereoscopic 3D. *ACM Trans. Graph.* 29(4):75
Lanman D, Hirsch M, Kim Y, Raskar R. 2010. Content-adaptive parallax barriers: optimizing dual-layer 3D displays using low-rank light field factorization. *ACM Trans. Graph.* 29(6):163
Lanman D, Luebke D. 2013. Near-eye light field displays. *ACM Trans. Graph.* 32(6):220
Lee B, Park J-H. 2010. Overview of 3D/2D switchable liquid crystal display technologies. *Proc. SPIE* 7618:761806
Lippmann G. 1908. La photographie intégrale. *Acad. Sci.* 146:446–51
Liu S, Cheng D, Hua H. 2008. An optical see-through head mounted display with addressable focal planes. In *7th IEEE International Symposium on Mixed and Augmented Reality*, pp. 33–42. Piscataway, NJ: Inst. Electr. Electron. Eng.
Liu S, Hua H. 2010. A systematic method for designing depth-fused multi-focal plane three-dimensional displays. *Opt. Expr.* 18(11):11562–73

Love GD, Hoffman DM, Hands PJW, Gao J, Kirby AK, Banks MS. 2009. High-speed switchable lens enables the development of a volumetric stereoscopic display. *Opt. Expr.* 17:15716–25

Lu C, Fender DH. 1972. The interaction of color and luminance in stereoscopic vision. *Investig. Ophthalmol. Vis. Sci.* 11:482–90

MacKenzie KJ, Dickson RA, Watt SJ. 2012. Vergence and accommodation to multiple-image-plane stereoscopic displays: "real world" responses with practical image-plane separations? *J. Electron. Imaging* 21(1):011002

MacKenzie KJ, Hoffman DM, Watt SJ. 2010. Accommodation to multiple-focal-plane displays: implications for improving stereoscopic displays and for accommodation control. *J. Vis.* 10(8):22

Maiello G, Manuela C, Solari F, Bex PJ. 2014. Simulated disparity and peripheral blur interact during binocular fusion. *J. Vis.* 14(8):13

Maimone A, Wetzstein G, Hirsh M, Lanman D, Raskar R, Fuchs H. 2013. Focus 3D: compressive accommodation display. *ACM Trans. Graph.* 32(5):153

Marshall J, Burbeck C, Ariely D, Rolland J, Martin K. 1996. Occlusion edge blur: a cue to relative visual depth. *J. Opt. Soc. Am. A* 13:681–88

Martens TG, Ogle KN. 1959. Observations on accommodative convergence; especially its nonlinear relationships. *Am. J. Ophthalmol.* 47:455–62

Mather G. 2006. *Foundations of Perception*. New York: Taylor & Francis

Mather G, Smith DRR. 2000. Depth cue integration: stereopsis and image blur. *Vis. Res.* 40:3501–6

Mather G, Smith DRR. 2002. Blur discrimination and its relation to blur-mediated depth perception. *Perception* 31:1211–19

McIntire JP, Havig PR, Geiselman EE. 2014a. Stereoscopic 3D displays and human performance: a comprehensive review. *Displays* 35:18–26

McIntire JP, Wright ST, Harrington LK, Havig PR, Watamaniuk SN, Heft EL. 2014b. Optometric measurement predict performance but not comfort on a virtual object placement task with a stereoscopic three-dimensional display. *Opt. Eng.* 53(6):061711

Mitchell DE, O'Hagan S. 1972. Accuracy of stereoscopic localization of small line segments that differ in size or orientation for the two eyes. *Vis. Res.* 12:437–54

Morgan MJ. 1979. Perception of continuity in stroboscopic motion: a temporal frequency analysis. *Vis. Res.* 19:491–500

Mun S, Park M-C, Park S, Whang M. 2012. SSVEP and ERP measurement of cognitive fatigue caused by stereoscopic 3D. *Neurosci. Lett.* 525(2):89–94

Narain R, Albert RA, Bulbul A, Ward GJ, Banks MS, O'Brien JF. 2015. Optimal presentation of imagery with focus cues on multi-plane displays. *ACM Trans. Graph.* 34(4):59

Nefs HT. 2012. Depth of field affects perceived depth-width ratios in photographs of natural scenes. *Seeing Perceiving* 25:577–95

Nojiri Y, Yamanoue H, Hanazato A, Emoto M, Okano F. 2004. Visual comfort/discomfort and visual fatigue caused by stereoscopic HDTV viewing. *Proc. SPIE* 5291:303

Nojiri Y, Yamanoue H, Hanazato A, Okana F. 2003. Measurement of parallax distribution, and its application to the analysis of visual comfort for stereoscopic HDTV. *Proc. SPIE* 5006:195

Norcia AM, Tyler CW. 1984. Temporal frequency limits for stereoscopic apparent motion processes. *Vis. Res.* 24:395–401

Palmer SE. 1999. *Vision Science: Photons to Phenomenology*. Cambridge, MA: MIT press

Palmer SE, Brooks JL. 2008. Edge-region grouping in figure-ground organization and depth perception. *J. Exp. Psychol.: Hum. Percept. Perform.* 34(6):1353–71

Palmisano S. 1996. Perceiving self-motion in depth: the role of stereoscopic motion and changing-size cues. *Percept. Psychophys.* 58(8):1168–76

Palmisano S. 2002. Consistent stereoscopic information increases the perceived speed of vection in depth. *Perception* 31:463–80

Park S, Won MJ, Mun S, Lee EC, Whang M. 2014. Does visual fatigue from 3D displays affect autonomic regulation and heart rhythm? *Int. J. Psychophysiol.* 92(1):42–48

Pastoor S. 1993. Human factors of 3D displays in advanced image communications. *Displays*. 14:150–57

Peli E. 1998. The visual effects of head-mounted display (HMD) are not distinguishable from those of desk-top computer display. *Vis. Res.* 38:2053–66

Pentland AP. 1987. A new sense for depth of field. *IEEE Trans. Pattern Anal. Mach. Intell.* 9:523–31

Percival AS. 1928. *The Prescribing of Spectacles*. Bristol, UK: J. Wright & Sons

Pollock B, Burton M, Kelly JW, Gilbert S, Winer E. 2012. The right view from the wrong location: depth perception in stereoscopic multi-user virtual environments. *IEEE Trans. Vis. Comput. Graph.* 18:581–88

Pulfrich C. 1922. Die Stereoskopie im Dienste der isochromen und heterochromen Photometrie. *Die Naturwissenschaften* 10:751–61

Ravikumar S, Akeley K, Banks MS. 2011. Creating effective focus cues in multi-plane 3D displays. *Opt. Expr.* 19(21):20940–52

Read JCA, Bohr I. 2014. User experience while viewing stereoscopic 3D television. *Ergonomics* 57:1140–53

Read JCA, Cumming BG. 2005. The stroboscopic Pulfrich effect is not evidence for the joint encoding of motion and depth. *J. Vis.* 5(5):3

Read JCA, Godfrey A, Bohr I, Simonotto J, Galna B, Smulders TV. 2016. Viewing 3D TV over two months produces no discernible effects on balance, coordination or eyesight. *Ergonomics.* In press

Ross J, Hogben JH. 1975. The Pulfrich effect and short-term memory in stereopsis. *Vis. Res.* 15:1289–90

Rovamo J, Raninen A. 1988. Critical flicker frequency as a function of stimulus area and luminance at various eccentricities in human cone vision: a revision of Granit-Harper and Ferry-Porter laws. *Vis. Res.* 28(7):785–90

Schechner YY, Kiryati N. 2000. Depth from defocus versus stereo: How different really are they? *Int. J. Comput. Vis.* 29(2):141–62

Schor CM. 1992. A dynamic model of cross-coupling between accommodation and convergence: simulations of step and frequency responses. *Optom. Vis. Sci.* 69(4):258–69

Semmlow J, Wetzel P. 1979. Dynamic contributions of the components of binocular vergence. *J. Opt. Soc. Am.* 69:639–45

Seuntiëns PJ, Meesters LM, IJsselsteijn WA. 2005. Perceptual attributes of crosstalk in 3D images. *Displays* 26(4–5):177–83

Sheard C. 1930. Zones of ocular comfort. *Am. J. Optom.* 7(1):9–25

Sheedy JE, Hayes J, Engle J. 2003. Is all asthenopia the same? *Optom. Vis. Sci.* 80(11):732–39

Shibata T, Kim J, Hoffman DM, Banks MS. 2011. The zone of comfort: predicting visual discomfort with stereo displays. *J. Vis.* 11(8):11

Shibata T, Muneyuki F, Oshima K, Yoshitake J, Kawai T. 2013. Comfortable stereo viewing on mobile devices. *Proc. SPIE* 8648:86481D

Simon A, Jorke H. 2011. Interference filter system for high-brightness and natural-color stereoscopic imaging. *SID Symp. Digest Tech. Pap.* 42:317–19

Son J-Y, Saveljev VV, Choi Y-J, Bahn J-E, Kim S-K, Choi H-H. 2003. Parameters for designing autostereoscopic imaging systems based on lenticular, parallax barrier, and integral photography plates. *Opt. Eng.* 42(11):3326–33

Soneira RM. 2012. *3D TV display technology shoot-out*. DisplayMate. **http://www.displaymate.com/3D_TV_ShootOut_1.htm**

Sugawara M, Masaoka K, Emoto M, Matsuo Y, Nojiri Y. 2008. Research on human factors in ultrahigh-definition television (UHDTV) to determine its specifications. *SMPTE Mot. Imaging J.* 117(3):23–29

Sullivan A. 2004. DepthCube solid-state 3D volumetric display. *Proc. SPIE* 5291:279

Torii M, Okada Y, Ukai K, Wolffsohn JS, Gilmartin B. 2008. Dynamic measurement of accommodative responses while viewing stereoscopic image. *J. Mod. Opt.* 55(4–5):557–67

Trentacoste M, Mantiuk R, Heidrich W. 2011. Blur-aware image downsampling. *Comput. Graph. Forum* 30:573–82

Tsirlin I, Wilcox LM, Allison RS. 2011. The effect of crosstalk on the perceived depth from disparity and monocular occlusions. *IEEE Trans. Broadcast.* 57(2):445–53

Turner TL, Hellbaum RF. 1986. LC shutter glasses provide 3-D display for simulated flight. *Inf. Disp.* 9(2):22–24

Tyler CW. 1974. Depth perception in disparity gratings. *Nature* 251:140–42

Urvoy M, Barkowsky M, Le Callet P. 2013. How visual fatigue and discomfort affect 3D-TV quality of experience: a comprehensive review of technological, psychophysical, and psychological factors. *Ann. Telecommun.* 68:641–55

van Beurden MHPH, IJsselsteijn WA, Juola JF. 2012. Effectiveness of stereoscopic displays in medicine: a review. *3D Res.* 3:3

van Ee R, Anderson BL. 2001. Motion direction, speed and orientation in binocular matching. *Nature* 410:690–94

Vishwanath D, Blaser E. 2010. Retinal blur and the perception of egocentric distance. *J. Vis.* 10(10):26

Vishwanath D, Girshick AR, Banks MS. 2005. Why pictures look right when viewed from the wrong place. *Nature Neurosci.* 8(10):1401–10

Wandell BA, Silverstein L. 2003. Digital color reproduction. In *The Science of Color*, ed. S Shevell. Oxford, UK: Elsevier

Watson AB. 2010. Display motion blur: comparison of measurement methods. *J. Soc. Inf. Disp.* 18(2):179–90

Watson AB. 2013. High frame rates and human vision: a view through the window of visibility. *SMPTE Mot. Imaging J.* 122:18–32

Watson AB, Ahumada AJ, Farrell JE. 1986. Window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays. *J. Opt. Soc. Am. A* 3:300–7

Watt SJ, Akeley K, Ernst MO, Banks MS. 2005. Focus cues affect perceived depth. *J. Vis.* 5(10):7

Wetzstein G, Lanman D, Heidrich W, Raskar R. 2011. Layered 3D: tomographic image synthesis for attenuation-based light field and high dynamic range displays. *ACM Trans. Graph.* 30:95

Wetzstein G, Lanman D, Hirsch M, Raskar R. 2012. Tensor displays: compressive fight field synthesis using multilayer displays with directional backlighting. *ACM Trans. Graph.* 31:80

Wheatstone C. 1838. Contributions to the physiology of vision. Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philos. Trans. R. Soc. Lond.* 128:371–94

Wilcox LM, Allison RS, Helliker J, Dunk B, Anthony RC. 2015. Evidence that viewers prefer higher frame-rate film. *ACM Trans. Appl. Percept.* 12:15

Wilcox LM, Stewart JAD. 2003. Determinants of perceived image quality: ghosting vs. brightness. *Proc. SPIE* 5006:263

Woods AJ. 2012. Crosstalk in stereoscopic displays: a review. *J. Electron. Imaging* 21(4):040902

Woods AJ, Docherty T, Koch R. 1993. Image distortions in stereoscopic displays. *Proc. SPIE 1915, Stereosc. Disp. Appl. IV, San Jose, CA*, p. 36

Woods AJ, Harris CR. 2010. Comparing levels of crosstalk with red/cyan, blue/yellow, and green/magenta anaglyph 3D glasses. *Proc. SPIE* 7254:75240Q

Woods AJ, Yuen KL, Karvinen KS. 2007. Characterizing crosstalk in anaglyphic stereoscopic images on LCD monitors and plasma displays. *J. Soc. Inf. Disp.* 15(11):889–98

Wöpking M. 1995. Viewing comfort with stereoscopic pictures: an experimental study on the subjective effects of disparity magnitude and depth of focus. *J. Soc. Inf. Disp.* 101–3

Yang S, Schlieski T, Salmons B, Cooper SC, Doherty RA, et al. 2012. Stereoscopic viewing and reported perceived immersion and symptoms. *Optom. Vis. Sci.* 89:1068–80

Yang S, Sheedy JE. 2011. Effects of vergence and accommodative responses on viewer's comfort in viewing 3D stimuli. *Proc. SPIE* 7863:78630Q

Yano S, Emoto M, Mitsuhashi T. 2004. Two factors in visual fatigue caused by stereoscopic HDTV images. *Displays* 25:141–50

Yano S, Ide S, Mitsuhashi T, Thwaites H. 2002. A study of visual fatigue and visual comfort for 3D HDTV/HDTV images. *Displays* 23:191–201

Yun JD, Kwak Y, Yang S. 2013. Evaluation of perceptual resolution and crosstalk in stereoscopic displays. *J. Disp. Technol.* 9:106–11

Zannoli M, Love GD, Narain R, Banks MS. 2016. Blur and the perception of depth at occlusions. *J. Vis.* 16(6):17

Zhang T, O'Hare L, Hibbard PB, Nefs HT, Heynderickx I. 2014. Depth of field affects perceived depth in stereographs. *ACM Trans. Appl. Percept.* 11(4):18