Stereoscopy and the Human Visual System

Martin S. Banks, Jenny R. Read, Robert S. Allison, & Simon J. Watt

Stereoscopic displays have become very important for many applications, including operation of remote devices, medical imaging, surgery, scientific visualization, computer-assisted design, and more. But the most significant and exciting development is the incorporation of stereo technology into entertainment: specifically, cinema, television, and video games. It is important in these applications for stereo 3D imagery to create a faithful impression of the 3D structure of the scene being portrayed. It is also important that the viewer is comfortable and does not leave the experience with eye fatigue or a headache. And that the presentation of the stereo images does not create temporal artifacts like flicker or motion judder.

Here we review current research on stereo human vision and how it informs us about how best to create and present stereo 3D imagery. The paper is divided into four parts: 1) Getting the geometry right; 2) depth cue interactions in stereo 3D media; 3) focusing and fixating on stereo images; and 4) temporal presentation protocols: Flicker, motion artifacts, and depth distortion.

Getting the Geometry Right

What are we trying to do when we present stereo displays? Are we trying to recreate the scene as a physically-present viewer would have seen it? Or simply give a good depth percept? How should we capture and display the images to achieve each of these? Vision science does not yet have the answers regarding what makes a good depth percept, but in this section we aim to cover the geometrical constraints and lay out what is currently known about how the brain responds to violations of those constraints.

The puppet theater

Figure 1 depicts a 3D display as reproducing a visual scene as a miniature model—a puppet theater if you will—in front of the viewer. Of course, conventional stereoscopic displays cannot recreate the optical wavefronts of a real visual scene. For example, the images are all presented on the same physical screen, and therefore they cannot reproduce the varying accommodative demand of real objects at different distances. And they cannot provide the appropriate motion parallax as the viewers moves his/her head left and right. However, they can in principle reproduce the exact binocular disparities of a real scene. In many instances, this will be an impossible or inappropriate goal. However, we argue that it is important to understand the underlying geometrical constraints of the “puppet theater” in order to understand what we are doing when we violate the constraints. Thus, it will be a helpful exercise to consider what we would have to do to reproduce the disparities that would be created by a set of physical objects seen by the viewer.

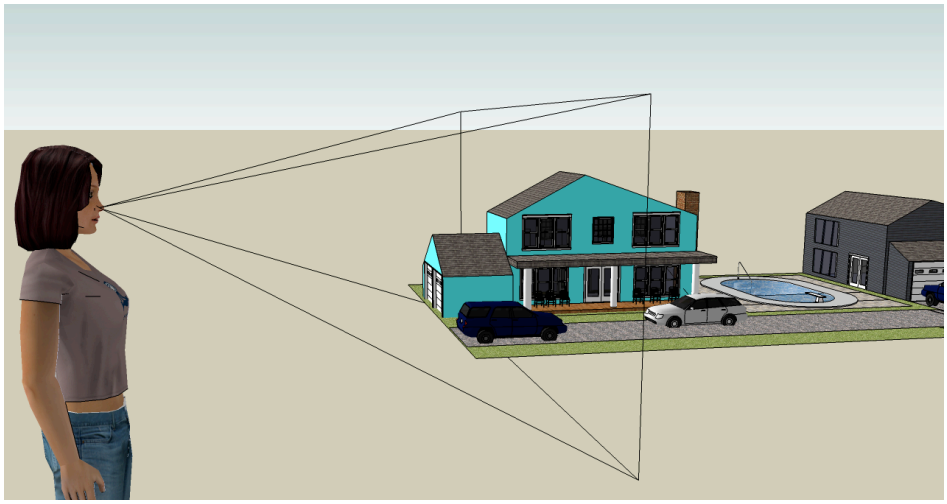

Figure 1. A visual scene as a miniature model in front of the viewer.

Epipolar geometry and vertical disparity

The images we need to create are of course dependent on how the they will be displayed. In the real world, a point in space and the projection centers of the two eyes define a plane; this is the so-called *epipolar plane*. To recreate this situation on a stereo 3D display, we have to recreate such an epipolar plane. Let us assume the images will be displayed on a screen frontoparallel to the viewer, so that horizontal lines on the screen are parallel to the line joining the two eyes (**Figure 1**, **Figure 2**). This is approximately the case for home viewing of stereo 3D TV. The first constraint to note in this situation is that, to simulate real objects in the puppet theater, there must be no vertical parallax on the screen: otherwise, the points on the display screen seen by the left and right eyes will not lie on an epipolar plane. (We will use the convention that *parallax* refers to separation on the screen, and *disparity* to separation on the retina. **Figure 2** illustrates why. Irrespective of where the eyes are looking (providing the viewer does not tilt their head), the rays joining each eye to a single object in space intersect the screen at points that are displaced horizontally on the screen: epipolar-plane geometry is preserved. Thus, in order to simulate objects physically present in front of the viewer, the left and right images must be presented on the display screen with no vertical separation.

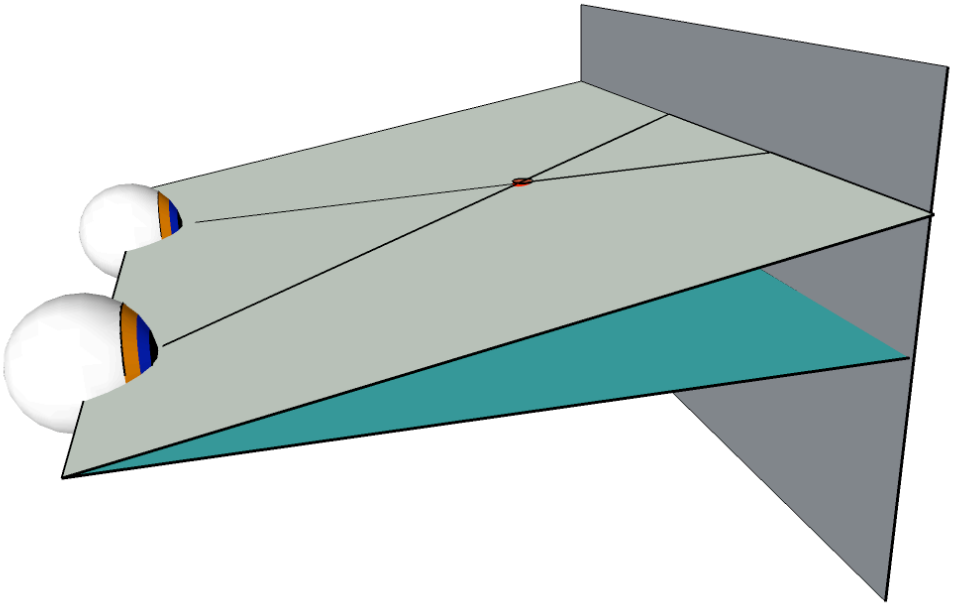

Figure 2. An object in space and the centers of projection of the eyes define an *epipolar plane*. If the screen displaying the stereo 3D content contains a vector parallel to the interocular axis, then the intersection of this plane with the screen will also be parallel to the interocular axis. In the usual case where the interocular axis is horizontal, this means that to reproduce the disparity of the real object, its two images must have zero vertical parallax. Their horizontal parallax will depend on how far the simulated object is in front of or behind the screen.

What happens when we get the geometry wrong?

If the stereo images on the display do contain vertical parallax, they are not consistent with a physically present object. Vertical parallax can be introduced by obvious problems such as misalignments of the cameras during filming or misalignments of the images during presentation, but they can also be introduced by more subtle issues such as filming with converged cameras ("toe in"; Allison, 2007). These two sources of vertical parallax cause a change in the vertical disparities at the viewer's retinas and are very likely to affect the 3D percept.

Vertical disparities arising from misalignments. The eyes move in part to minimize retinal disparities. For example, vergence eye movements work to ensure that the lines of sight of the two eyes intersect at a desired point in space. Horizontal vergence (convergence or divergence) is triggered by horizontal disparities. If the eyes are vertically misaligned, the lines of sight will not intersect in space. There will be a constant vertical offset between the two eyes' images (*Figure 3*, left). The human visual system contains self-correcting mechanisms designed to detect such a vertical offset and correct for it by moving the eyes back into alignment (Allison, Howard, & Fang, 2000; Howard, Allison, & Zacher, 1997; Howard, Fang, Allison, & Zacher, 2000). The eye movement that accomplishes this is a vertical vergence.

In stereo displays, small vertical misalignments of the images activate vertical vergence. This could occur, for example, if the cameras were misaligned by a rotation about the axis joining the centers of the two cameras, or in a cinema if the projectors were offset vertically. The viewers will automatically diverge their eyes vertically, so as to remove the offset between the retinal images. This happens automatically, so inexperienced viewers are usually not consciously aware of it. It is likely to cause fatigue and eyestrain if it persists.

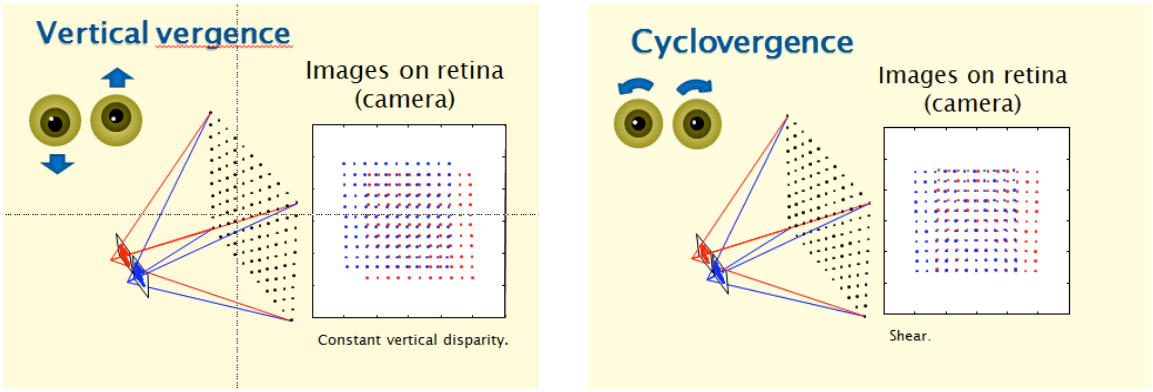

Figure 3. Different eye postures cause characteristic patterns of vertical disparity on the retina, largely independent of the scene viewed. Here, the eyes view an array of points on a grid in space, directly in front of the viewer. The eyes are not converged, so the points have a large horizontal disparity on the retina. In the left panel, the eyes have a vertical vergence misalignment. This introduces a constant vertical disparity across the retina. In the right panel, the eyes are slightly cyclodiverged (rotated in opposite directions about the lines of sight). This introduces a shear-like pattern of vertical disparity across the retina.

Passive stereo displays where the left and right images are presented on alternate pixel rows could introduce a constant vertical disparity corresponding to one pixel, if the left and right images were captured from vertically aligned cameras and then presented with an offset (**Figure**

4A). If instead the images are captured at twice the vertical resolution of each eye's image, with the left and right images pulled off from the odd and even pixel rows respectively, there is no overall vertical disparity, just slightly different sampling (**Figure 4**B). In any case, the vertical disparity corresponding to one pixel viewed at 3 picture heights is only about a minute of arc, which is probably too small to cause eyestrain.

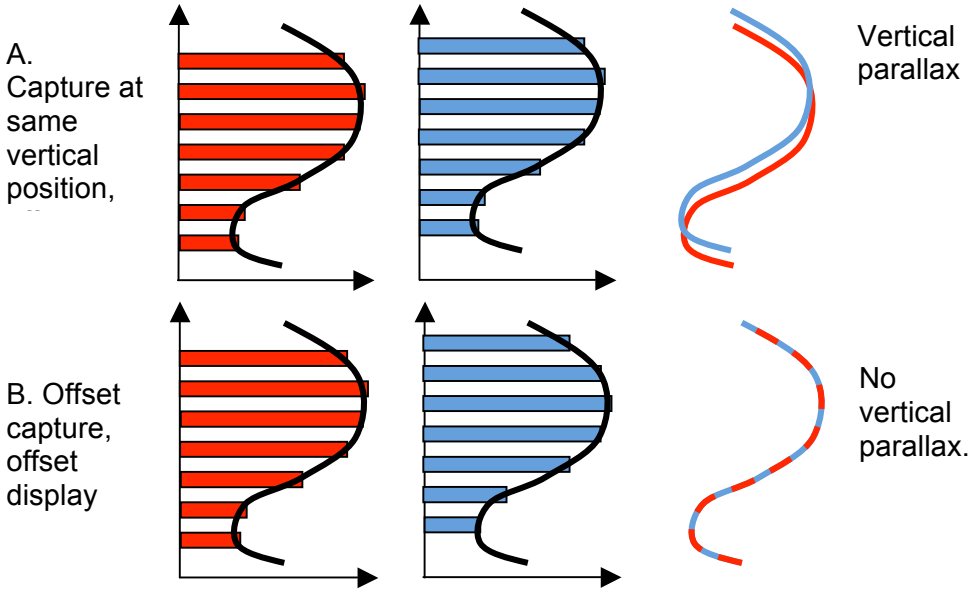

Figure 4. Passive stereo where left and right images are displayed on different pixel rows can introduce vertical parallax (A), but need not do so if created appropriately (B).

A similar situation occurs if the images are misaligned by being rotated about an axis perpendicular to the screen. Once again, the brain automatically seeks to null out rotations of up to a few degrees by rotating the eyes about the lines of sight, an eye movement known as *cyclovergence* (Rogers & Bradshaw, 1999); **Figure 3**, right. This again produces discomfort, fatigue, and eyestrain.

Vertical disparities arising from viewing geometry. The human visual system (and human stereographers) works hard to avoid such misalignments. But even if both eyes (or both cameras) are perfectly aligned, vertical disparities between the retinal (or filmed) images can still occur. ***Figure 5*** shows two cameras converged on a square structure in front of them. Because each camera is viewing the square obliquely, its image on the film is a trapezoid, due to keystoning. The corners of the square are thus in different vertical positions on the two films, and this violates epipolar-plane geometry (Figure 2). A viewer looking at the stereo display of these trapezoids receives a pattern of vertical disparities that is inconsistent with the original scene.

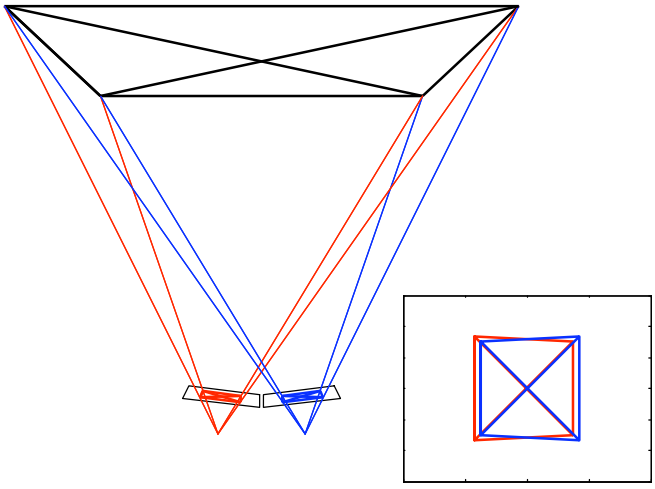

Figure 5. Vertical parallax introduced by camera convergence.

One can deduce the relative alignment of the cameras from the pattern of vertical disparities (Longuet-Higgins, 1981). The exact position of objects in space can then be estimated by back-projecting from the retinal images. The visual system uses the pattern of vertical disparities across the retina to interpret and scale the information available from horizontal disparities (Garding, Porrill, Mayhew, & Frisby, 1995; Read, Phillipson, & Glennerster, 2009; Rogers & Bradshaw, 1993). For this reason, vertical disparities in stereo displays do not necessarily just degrade the 3D experience, but can produce systematic distortions in depth perception.

One well known example is the induced effect (Ogle, 1938). In this illusion, a vertical magnification of one eye's image relative to the other causes a perception that the whole screen is slightly slanted, i.e. rotated about a vertical axis. This is thought to be because a similar vertical magnification occurs naturally when we view a surface obliquely.

Estimating convergence from vertical disparity. For the purpose of stereo 3D displays, a more pertinent example concerns vertical disparities associated with convergence. For the brain to interpret 3D information correctly, it must estimate the current convergence angle with which it is viewing the world. This is because, as *Figure 6* shows, a given retinal disparity can specify vastly different depth estimates depending on whether the eyes are converging on a point close to the viewer, or far away.

The brain has several sources of information about convergence. Some of these are entirely independent of the visual content: for example, sensory information from the eye muscles. However, the pattern of vertical disparities also provides a purely retinal source of information. Consider the example in Figure 5. The equal and opposite keystoning in the two images instantly tells us that this images must have been acquired by converged cameras. The larger the keystoning, the more converged the cameras.

There is an extensive vision science literature examining humans' ability to use these cues. This shows that humans do use both retinal and extra-retinal information about eye

position (Backus, Banks, van Ee, & Crowell, 1999; Banks & Backus, 1998; Bradshaw, Glennerster, & Rogers, 1996). As one would expect from a well-engineered system, more weight is placed on whichever cue is most reliable. Generally, the visual system places more weight on the retinal information, relying on the physical convergence angle only when the retinal images are less informative (Backus & Banks, 1999; Backus, et al., 1999). For example, because the vertical disparity introduced by convergence is larger at the edge of the visual field, less weight is given to the retinal information when it is available only in the center of the visual field (Bradshaw, et al., 1996).

These vertical disparities can have substantial effects on the experience of depth. For example, in one experiment, the same horizontal disparity (10 arc min) resulted in a perceived depth difference of 5cm when the vertical disparity pattern indicated viewing at infinity, but only 3cm when the vertical disparity pattern indicated viewing at 28cm – although in both cases, the physical viewing distance was 57cm (Rogers & Bradshaw, 1993).

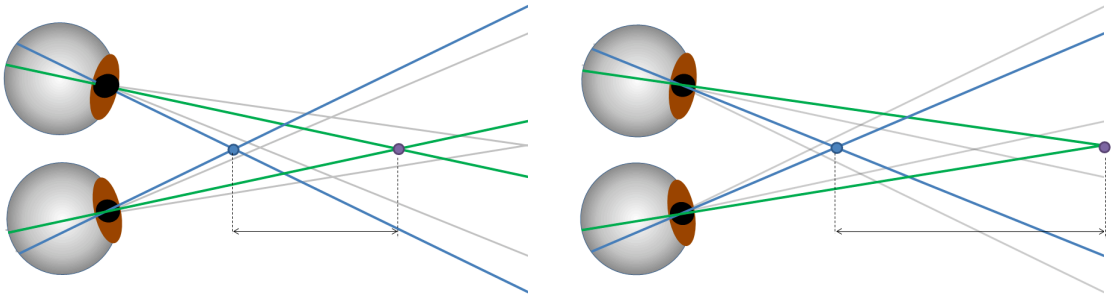

Figure 6. Mapping from disparity to depth depends on the convergence angle. In both panels, the eyes are fixating the purple sphere. In A, the sphere is close so the eyes are more strongly converged. The retinal disparity between the two sphere is the same in both panels, but the physical distance this maps onto is much larger in B, where the convergence angle is smaller.

The effect of filming with converged cameras. This geometry is relevant to the vexing issue of whether stereo content should be shot with camera axes parallel or converged ("toe in"). Some stereographers have argued that cameras should converge on the subject of interest in filming because the eyes converge in natural viewing. While there are good reasons for filming "toe-in", this particular justification is not correct. It depends on the fallacy that cameras during filming are equivalent to eyes during viewing. This would be the case only if the images recorded during filming were presented directly to the audience's retinas, without distortion. Instead, the images recorded during filming are presented on a screen that is usually roughly frontoparallel to the interocular axis (*Figure 2*). The images displayed on the screen are thus viewed obliquely by each eye, introducing keystoning at the retinas. As described in ***Figure 5***, the retinal images therefore contain vertical disparities even if there was no vertical parallax on the screen. If the images displayed on the screen have vertical parallax because they were captured with converged cameras, this will add to the vertical disparity introduced by the viewer's own convergence. The resulting vertical disparity will indicate that the viewer's eyes are more converged than they really are. As we have seen, this could potentially reduce the amount of perceived depth for a given horizontal disparity.

To correctly simulate physical objects, one should film with the camera axes parallel, as shown in **Figure 7**. To display the resulting images, one should shift them horizontally so that objects that are meant to have the same simulated distance as the screen distance have zero horizontal parallax on the screen. Provided that the viewer keeps the interocular axis horizontal and parallel to the screen, this ensures that all objects have correct horizontal and vertical disparity on the retina, independent of the viewer's convergence angle.

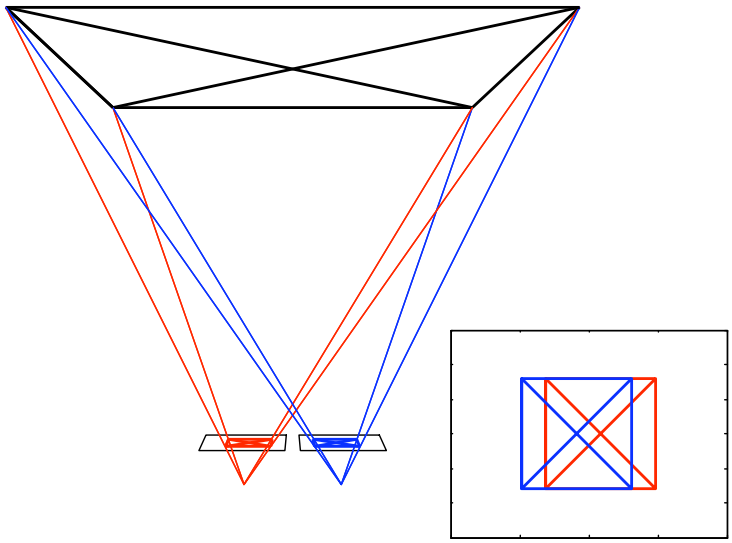

Figure 7. Filming with parallel camera axes.

Some back-of-the-envelope calculations. To get a feel for how serious these effects might be, we can do some back-of-the-envelope calculations. For convergence on the midline (i.e., looking straight ahead, not to left and right), vertical disparity is independent of scene structure and simply scales with convergence angle. To close approximation, the retinal vertical disparity at different points in the visual field is given by the following equation (Read, Phillipson, & Glennerster, 2009):

[retinal vertical disparity] = [convergence angle] × 0.5*sin(2*elevation) × tan(azimuth),

where azimuth and elevation refer to location in the visual field. This equation is for the vertical disparity in natural viewing. That is, even if an object is displayed with zero screen parallax, it will still have a vertical disparity of 7 arc min when viewed with 1° convergence at 20° elevation and 20° azimuth. The same equation can of course be used to compute the on-screen vertical disparity resulting from filming toe-in. For example, one can ask what degree of toe-in is necessary to cause a 1-pixel vertical disparity. For 36mm film with a 50mm focal length, the corners of the image are at azimuth=20°, elevation=13°. If the 36mm is represented by 2048 pixels, a vertical disparity of 1 pixel is 1.2 arc min. This can be caused by a toe-in of just 14 arcmin.

Is this enough to alter perception? Let's suppose that the images on the screen have a pattern of on-screen vertical parallax resulting from having been filmed toed-in:

[on-screen vertical parallax] = [some scale factor K] × 0.5*sin(2*elevation) × tan(azimuth).

This will combine with the natural vertical disparity, indicating the wrong convergence angle. The scale-factor K, which has angular units, is the additional, artefactual component of the convergence estimate that would be added if the visual system worked solely on the retinal information.

Let us suppose the viewer is in an IMAX cinema, screen size 22m × 16m, viewing it at one screen-height: 16m. Their true convergence angle is therefore 14 arcmin. At the corner of the screen, elevation = 27° and azimuth = 35°. Physical objects at the corners of the screen produce a retinal vertical disparity of 2.0 arcmin just by virtue of the geometry. Suppose the toe-in vertical disparity is such that it is just 1 cm even at the corners of the screen (clearly it is smaller everywhere else). This means that the toe-in contributes an additional 2.1 arcmin of vertical disparity at elevation = 27°, azimuth = 35°. That is, the barely noticeable on-screen parallax more than doubles the vertical disparity at the retina, and hence the retinal cue to convergence is 29 arcmin instead of the physical value of 14 arcmin.

Roughly speaking, the convergence overestimate in degrees = 180/π * [viewing distance] [on-screen vertical separation at (x,y)] / x / y. In the above calculation, the viewing distance was 16m, the vertical separation was 1cm at x=11m and y=8m, implying a convergence angle that is too large by around 0.1^o.

What implications might this convergence error have for perceived shape? Suppose the images accurately simulate a transparent sphere, radius 1m, at the center of the screen. The sphere has an angular radius of 3.6^o, and its front and back surfaces have a horizontal disparity of -0.93 arcmin and 0.82 arcmin, respectively. If these disparities were interpreted with the actual viewing distance of 16m, convergence 14 arcmin, the viewer should correctly perceive a spherical object, radius 1m, 16m away. But if the images are interpreted assuming a convergence of 29 arcmin, viewing distance 8m, then the on-screen parallax implies a spheroid with an aspect ratio of 2: i.e. a radius of 0.5m in the screen plane and just 0.25m perpendicular to the plane of the screen. That is, for the same horizontal parallax and the same viewing position, a supposedly spherical object could be perceived as flattened by a factor of 2, simply because of toe-in vertical parallax, even where this is just 1cm at the very corners of the screen.

In practice, the distortion may not be so obvious. For example, there may be other powerful perspective and shading cues indicating that the object is spherical. Nevertheless, these calculations suggest that small vertical parallax can potentially have a significant effect on perception.

As yet, little work has been done to investigate depth distortions caused by toe-in filming. From the vision science literature to date, we would predict different effects for stereo 3D cinema versus television. In a cinema, the display typically occupies much of the visual field. Thus, we would expect convergence estimates to be dominated by the retinal information, rather than the physical value. In this situation, it is possible that the same horizontal disparities could produce measurably different depth percepts if acquired with converged camera axes versus parallel. In home viewing of 3D television, the visual periphery will generally consist of physical objects in the room. These will necessarily produce vertical disparities consistent with the viewer's physical convergence angle, while vertical disparities within the relatively small television screen are likely to have less effect.This means that horizontal parallax on the television screen is likely to be converted into depth estimates using the viewer's physical convergence angle. Thus, we would expect the angle between the camera axes to have less effect on the depth perceived in this situation.

Interaxial distance. The separation of the cameras during filming is another important topic. To exactly recreate the puppet theater, one should film with the cameras one interocular distance apart. However, stereographers regularly play with interaxial distance (i.e., the separation between the optical axes of the cameras). For example, they might start with a large interaxial distance to produce measurable parallax in a shot of distant mountains, then reduce the interaxial distance as the scene changes to a close-up of a dragon on the mountain. A remarkable recent experiment has demonstrated that most observers are insensitive to changes in interaxial distance within a scene. Although we could detect the resulting changes in disparity if they occurred in isolation, when they occur within a given scene we do not perceive them, because we assume the objects stay the same size (Glennerster, Tcheang, Gilson, Fitzgibbon, & Parker, 2006). In the words of the authors, "Humans ignore motion and stereo cues [to absolute size] in favor of a fictional stable world".

Why we don't need to get it right

Ultimately, the central mystery for vision science may be why stereo 3D television and cinema works as well as it does. By providing an additional, highly potent depth cue, stereo 3D content risks alerting the visual system to errors it might have forgiven in 2D content. As an example, an actor's head on a cinema screen may be 10 foot high, but we do not perceive it as gigantic. One could argue that this is because a 10-foot head viewed from a distance of 30 feet subtends the same angle on the retina as a 1-foot head viewed from 3 feet. Stereo displays,

however, potentially provide depth information confirming that the actor is indeed gigantic. Additionally, stereo displays often depict disparities that are quite unnatural: i.e., disparities that are physically impossible for any real scene to produce given the viewer's eye position, or disparities that conflict with highly reliable real-world statistics (mountains are hundreds of feet high, people are around 6 feet high, etc.). This is reminiscent of the "uncanny valley" in robotics, where improving the realism of a simulated human can produce revulsion (Geller, 2008).

Presumably, such conflicts are the reason why a minority of people find stereo 3D content disturbing or nauseating. However, most of us find stereo 3D content highly compelling despite these violations of the natural order. An analogy can be drawn with the way we perceive most photographs as veridical depictions of the world. We do not usually perceive objects in photographs as distorted or straight lines as curved, despite the fact that – unless we are viewing the photograph from the exact spot the camera was located to take it – the image on our retina is substantially different from that produced by the real scene (Vishwanath, Girshick, & Banks, 2005). It is not yet known to what extent this is a learned ability, raising the possibility that as stereo displays become more commonplace, our visual systems will become even better at interpreting them without adverse effects.

Depth Cue Interactions in Stereoscopic 3D Media

Adding binocular disparity enriches media with a vivid sense of depth, solidity and space. However, the traditional pictorial depth cues—shading, shadows, blur, aerial perspective (haze, smoke), linear perspective, texture gradient, occlusion, etc.—used to provide a sense of depth, space, and texture are still present. In stereo 3D (S3D) displays, as in 2D displays, these cues are important and active as are the cues not normally provided by either 2D or S3D displays, such as accommodation and motion parallax due to head motion. These are not subsidiary or secondary cues that are replaced by binocular disparity when it is available, but continue to contribute to the qualitative sense of three-dimensionality and the quantitative depth experienced with two eyes as well as one. However, in S3D media (as in the real world) these multiple sources often provide incomplete, imprecise, ambiguous, and even contradictory depth information. The visual system has the challenge of reconstructing a coherent 3D percept from these myriad and changing sensory signals.

Variety and Ambiguity of Stereoscopic Percepts

Stereopsis has two inherent ambiguities that are important for cue interactions. The first results from the fact that the image in one eye must be matched with that in the other. This correspondence problem can be non-trivial especially with repetitive textures. It has been a major challenge for computer stereo vision. In contrast, the human visual system seems to solve this problem effectively and effortlessly (Julesz, 1960). This capacity and the fact that most S3D media is rich and varied makes this the lesser of the ambiguities for our purposes. The other ambiguity is that retinal disparity does not directly specify depth. As described earlier in this paper, the amount of depth corresponding to a given disparity depends on distance and to a lesser extent on direction. In the absence of good information for distance, a given disparity can correspond to a large range of possible depths. Horizontal disparity does not provide this distance information and binocular information from vergence or vertical disparity is limited to close range and has limited accuracy.

Stereopsis can support the perception of a 3D world in many respects including discriminating a difference in depth, ordering objects in depth, judging slant or curvature, obtaining shape and relief, judging speed or direction of motion in depth, recovering surface properties, or obtaining accurate measures of depth between objects. Depending on the nature of the task, the ambiguities of stereopsis become more or less important. For example, to determine whether one object is placed in front of another does not require calibration for viewing distance, but estimating the size of the gap between them does.

In the visual appreciation of S3D film and other content, these perceptions are complex and multifaceted. Depth ordering and segregation help reduce clutter and separate subject from background, recovering shape and relief provides volume and depth, recovering binocular highlights gives a sense of gloss and luster, and so on. In S3D media, cue integration and combination needs to be considered on all these levels as they occur simultaneously and often seamlessly.

Ambiguity, Reliability, and Accuracy

The problem of vision is to 'invert' the imaging process and recover the 3D world. But information is lost in the many-to-one transformation inherent in perspective projection. A given monocular image is compatible with multiple real scenes (Figure 8); one of the possible scenes is that we are simply viewing a 2D image on a plane, which is of course the case in painting, film, and television. Of course, not all possible interpretations are equally likely. The structure and regularities of the world greatly constrain the problem. A long tradition in perception holds that depth perception relies on recreating the most likely 3D world consistent with the retinal image(s). Helmholtz called this process 'unconscious inference (Helmholtz, 1909). Modern variants on this idea codify this probabilistic interpretation of sensory signals in ways that as we will see seem natural if not obvious to engineers.

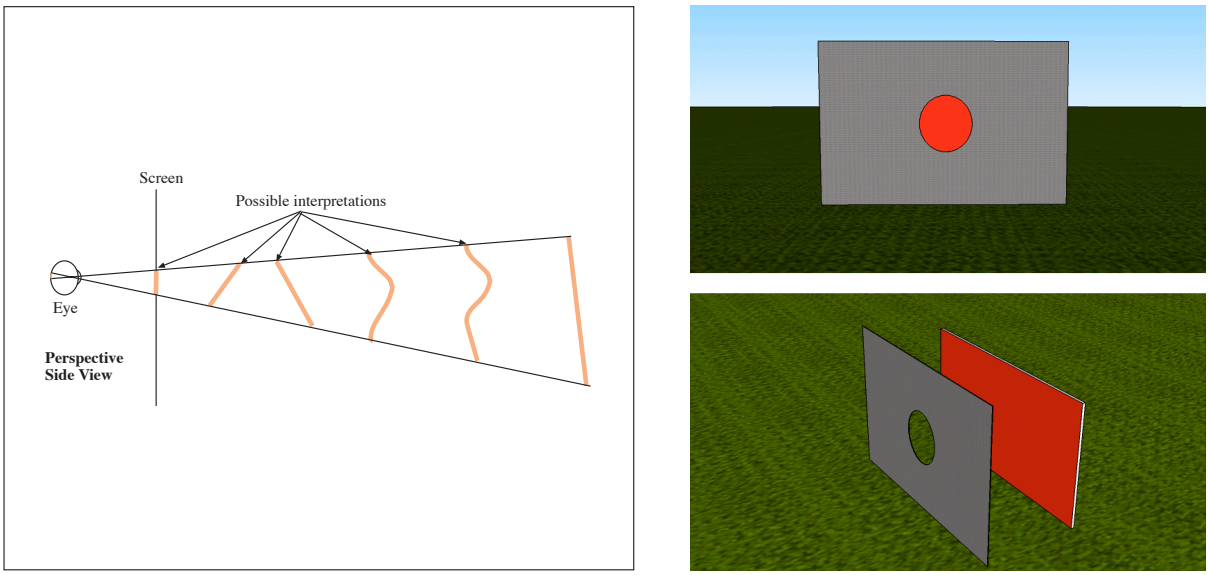

Figure 8. Ambiguity in perspective projection. Left panel shows how a perspective projection is compatible with many real scenes, including a drawing of a 2D surface. Right panel shows that even occlusion can be ambiguous if it is uncertain which surface is the occluder.

Just as stereopsis supports many types of 3D judgments and suffers from ambiguities, monocular cues vary in the degree that they support such judgments. For example, occlusion is one of the least ambiguous depth cues. If one object blocks the view of another, it must be between the latter object and the viewer. Of course, one object may be cut away so it looks like it blocks another or both 'objects' could be paint on a canvas. But for the most part occlusion is a very unambiguous source of relative depth information. Interestingly occlusion tells you nothing about the amount of depth between the two objects. One can intuitively imagine how occlusion and stereopsis might interact to determine depth order, but it is less clear how that would work for depth magnitude. That is, the two cues are not commensurate: they are apples and oranges that cannot be directly compared or combined (but see Burge et al., 2010).

Even when two cues are commensurate, they are not always comparable in terms of precision, reliability or range. For example, the stereoscopic baseline in humans is much larger than the pupil aperture, so stereopsis is a more precise relative depth signal than blur (Mather &

Smith, 2000; Held et al., 2010). Finally, the cues can differ in accuracy and provide biased estimates of depth or other 3D properties. For instance, shape estimates from shading are usually consistent with the assumption of a light source located above the viewer. When this assumption is not correct, shape from shading can provide biased estimates.

Cue Integration, Cue Combination, and Cue Conflict

Cue combination refers to the combination of sensory information to derive a percept of an object, feature, or scene. Cue integration refers more narrowly to combining multiple sources of commensurate information; i.e., about the depth, shape, velocity, or some other aspect of an object.

Cue conflict occurs when two or more cues provide different and incompatible information. This is often thought of in terms of cue integration, but can apply to other cue combination scenarios. Cue conflict can take place within binocular cues (e.g., vergence, stereopsis) and/or between stereopsis and other cues.

It is important to recognize that S3D media almost *always produce a cue conflict*. Many of these conflicts come with the technology such as the conflict with accommodation which always indicates S3D objects are at the screen plane not where we portray them. Other conflicts are due to the nature of the medium. For instance, scaling of images that arises from choice of lens or display size affects depth from disparity (stereopsis) and perspective differently. Sitting off-center in the theater also has a different effect on depth from disparity and perspective. Finally some natural conflicts simply arise from incompatible solutions to the ambiguities of vision. For instance, unusual lighting direction can make shape from shading incompatible with shape from disparity.

Conceptual and Computational Models of Cue Combination

There are many conceptual models of how cues can be perceptually combined including (Howard & Rogers, 2002):

1) *Cue dominance or vetoing* where one cue determines the percept. A familiar example is ventriloquism where the sound is ‘captured’ by the visual input. In S3D, occlusion cues can veto depth from disparity at window violations.
2) *Summation and averaging* which are additive interactions. These can be generalized to rather complex nonlinear interactions referred to as cooperative interactions (Bülthoff & Mallot, 1988).
3) *Disambiguation* where one cue disambiguates another (or they mutually disambiguate each other). For instance, the sign of blur is ambiguous and a given amount of blur can result from focus in front of or beyond an object; stereopsis and other depth signals could disambiguate. Information from other depth cues can also disambiguate which interpretation of a perspective image should be favored or confirm and stabilize a bistable perception.
4) *Calibration and adaptation* occurs when one cue provides information necessary to interpret another. For instance, motion or perspective can provide the distance signal necessary to obtain depth magnitude from stereopsis.
5) *Dissociation*. To be integrated, the cues should be *bound* together to apply to a common object feature or location. In contrast, dissociation of cues refers to interactions in which the cues are applied differentially, either interpreting them as arising from different objects or applying them to different aspects of the same object.

Cue integration is the *fusing* together of redundant (typically commensurate) information usually involving vetoing or averaging processes. The computational and behavioral literature (Clark & Yuille, 1990) has distinguished between weak and strong fusion models. In strong fusion models, sensory inputs are combined without constraining how the information is combined. There is considerable anatomical and psychophysical evidence for modularity in the visual system. To a significant extent, various depth cues such as shading, stereopsis, and

perspective may be processed independently to arrive at depth estimates for points in the scene. Models that combine the outputs of such depth modules are referred to as weak-fusion models. Landy et al. (1995) recognized the difficulty of integrating incommensurate cues and proposed the model of 'modified weak fusion' in which outputs of depth modules are combined linearly but limited nonlinear interaction is allowed to 'promote' cues so that they have a common measurement scales and can be combined. This promotion may include calibration, scaling and other effects driven by secondary cues. Although simple, this idea of a linear combination of quasi-independent depth modules has been very successful in practice.

The solution that often arises will be familiar to engineers, particularly those trained in communications theory. The brain must arrive at the optimal or most probable percept α consistent with the set of depth estimates **x** (that is maximize the conditional probability of $P(\alpha \mid \mathbf{x})$). It will not surprise most readers that classical techniques such as maximum likelihood estimators (MLE; Fisher, 1925) have been applied successfully. While theoretically limited, such linear estimators and more general Bayesian estimators have proven surprisingly successful in describing quantitatively how cues interact in the laboratory. The basic MLE solution predicts that observers will weigh the depth cues according to their reliability which is the inverse of their variance ($1/\sigma_{cue}^2$; Landy et al., 1995). For example, with depth estimates, *D*, from two cues we obtain

$$D_{optimal} = w_1 \cdot D_{cue1} + w_2 \cdot D_{cue2}$$

$$w_1 = \frac{\frac{1}{\sigma_{cue1}^2}}{\frac{1}{\sigma_{cue1}^2} + \frac{1}{\sigma_{cue2}^2}} \quad \text{and} \quad w_2 = \frac{\frac{1}{\sigma_{cue2}^2}}{\frac{1}{\sigma_{cue1}^2} + \frac{1}{\sigma_{cue2}^2}}$$

$$\frac{1}{\sigma_{optimal}^2} = \frac{1}{\sigma_{cue1}^2} + \frac{1}{\sigma_{cue2}^2}$$

If the cues have equal reliabilities, their weights are both ½ (averaging) and the reliability of the combined estimate increases by a factor of two.

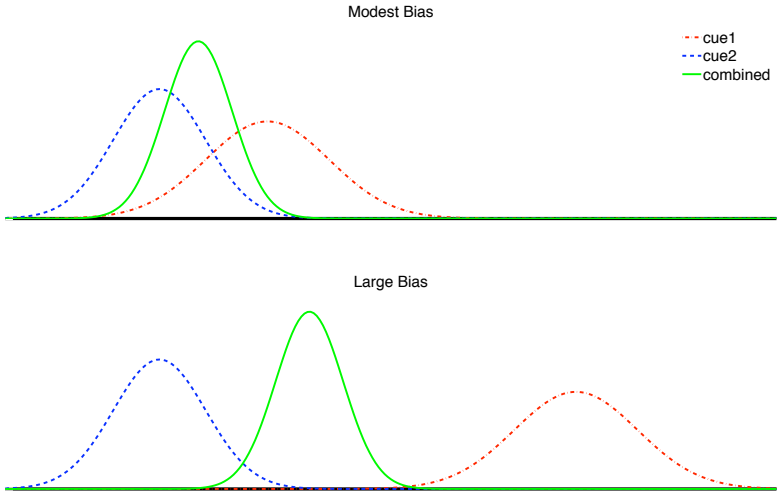

Figure 9. Cue combination. Curves show the likelihood of a given depth value (horizontal axis) provided by two cues and the MLE combination (all normalized for unit area). When the estimates of the two cues are similar (top), the weighted combination gives a more precise estimate. When bias is large (bottom), the cue combination may not be consistent with either cue.

More generally, cue integration can be formulated to take into account the likelihood of various perceptions and changes in the reliability of cues with distance, slant, and other factors (Knill, 2007). Recent research has turned to issues of dependencies among the cues and robustness when they disagree (see below).

Cue trading. If we can successfully express depth perception as a weighted depth cue combination as described above, an obvious question arises as to what extent we can trade one cue for another. One could imagine reducing interaxial distance (i.e., camera separation) and hence disparity for visual comfort while turning up the perspective, motion parallax or shading to compensate (Siegel & Nagata, 2000). To a certain extent this is possible and even mandatory if cues are combined at a perceptual level with the viewer having access to only the final perception (Hillis et al., 2002). On the other hand, there are limits to the degree this can be accomplished and automated:

- MLE and other weighted averages of depth cues are only appropriate if the modules provide (noisy) estimates of the same value. If one cue is suspected to be strongly biased or inaccurate, the visual system should discount it. By analogy, if you made ten measures of a parameter and nine measured 50+/-2 units, you would consider a tenth measurement of 500 to be likely due to error and you would therefore not average it with the others. It has been proposed that the visual system is similarly robust when cues are discrepant, for instance vetoing unreliable cues (Knill, 2007; Landy et al., 1995; Meese & Holmes, 2004). On the other hand, linear cue integration sometimes seems to occur even when cues are very discrepant (Muller et al., 2009).
- The weights adopted can vary by viewer, even if they all have good stereopsis. For instance, in judgments of the slant of surfaces, some observers preferentially weight perspective and others disparity (Alllison et al., 1998; Girshick & Banks, 2009).
- The weights assigned can vary with the type of task, previous experience, type of scene, and location in the image.
- Van Ee claims that discrepant depth cues can result in alternation of discrepant perceptions over time rather than stable cue integration (van Ee et al., 2003) though one of us has failed to replicate this finding (Girshick & Banks, 2009).

Thus cue trading is a complex scene-dependent and often idiosyncratic process.

Cue conflict examples.

Depth sign: Cue conflict in depth sign (in front versus behind) or depth order can often be considered especially strong conflicts. The standard example of this in S3D film is the window violation. Occlusion cues indicate the frame edge is in front of the stereoscopic imagery that is portrayed in front of it. Cue dominance may be perceived with the occlusion cue pinning the surface to the edge of the screen. In other cases, strange and uncomfortable cue dissociations can be perceived. Similar issues arise with depth-sign errors in automated 2D-to-3D conversion.

Depth magnitude: As described above, if cue conflicts are modest there tends to be trading or weighted averaging of cues to depth magnitude. Thus, other cues such as perspective and shading act to modulate the depth from disparity. Many of these conflicts are a consequence of differential effects of rig and projection parameters (such as focal length, interaxial distance, depth of field, screen distance, and screen size) on different depth cues. In many cases, cue integration can impact other aspects besides depth such as apparent size.

Slant: The orientation of surfaces in depth, or slant, is important for shape and object recognition. The relationship between slant specified by perspective-based cues (e.g., texture gradients) and disparity-gradient cues varies with focal length and magnification on the one hand and rig parameters such as interaxial distance on the other. Studies have shown that observers weight perspective and disparity to arrive an estimate of surface slant (Allison et al., 1998; Hillis et al., 2004). When surface slant from perspective and disparity differ greatly, observers tend to rely on one cue vetoing the other; but the cue preference is idiosyncratic and does not necessarily favor the most reliable (Girshick & Banks, 2009; Allison & Howard, 2000).

Mis-alignment of the stereo rig can also produce slant distortions (see earlier section). Rotational misalignment of the images about the z-axis produces horizontal disparity patterns consistent with the scene being slanted in depth about a horizontal axis. Similarly, size mis-

calibration (e.g., due to difference in focal length in the two cameras) produces disparities consist with slant about a vertical axis. Another distortion is due to keystone distortions arising from toe-in convergence of the cameras. This predicts perceived curvature of stereoscopic space (Woods et al., 1993; Held & Banks, 2008). While undesirable, these distortions are most noticeable when monocular cues are weak and strong perspective can attenuate or eliminate them.

Qualitative depth and appearance: Interaction of stereopsis with shading and lighting in an S3D context is important. Much work needs to be done here, but it seems that lighting for depth can enhance the sense of volume and space in S3D content. Similarly beyond geometrical properties stereopsis like slant, stereopsis can influence perception of material properties like transparency (Tsirlin et al., 2008). Many stereographers feel that specular highlights should be avoided at all costs. In everyday experience though, binocular differences in intensity produce percepts of luster that supports the perception of surface gloss (Sakano & Ando, 2010). An effect of lighting on perceived depth is provided in Figure 10. In S3D content, however, you can obtain intensity disparities that are not associated with surface glossiness and thus can conflict with monocular information on shininess. For instance, if the beam-splitter in a mirror rig is polarization sensitive (i.e., preferentially reflects one polarization state while transmitting the other), the two images can have large differences in intensity for reflecting surfaces like water and glass. These artifacts are due to beam-splitter characteristics rather than the interocular difference in vantage point, so would be very difficult for the brain to interpret ecologically (e.g., the entire surface of a pond might be bright in one eye but not the other). By their very nature, specularities are highly directional hence constrained phenomenon. Binocular specular highlights are very informative, but fairly small changes can likely make them geometrically implausible. It is possible that we might be more sensitive to incorrect binocular specularities than to other cue conflicts.

Figure 10. Use of lighting to enhance the perception of depth. Top and bottom stereo pairs are arranged for cross-eyed fusion and have the same camera parameters. The top pair has high-contrast lighting the bottom has flat lighting. Viewers generally report that the top pair has more depth than the bottom pair.

Tolerance to cue conflict. A key concern is the tolerance of the typical observer to these cue conflicts. How much can they tolerate? When problematic, how much does it bother us?

Unfortunately, particularly in the context of rich cinematic content, these are still very much open questions. Cue conflict has been linked to simulator sickness effects and degraded perception. We understand in certain situations how cue conflict can cause issues (see the section on vergence and accommodation conflict, for example). Most of this data has come from either non-specific image quality and comfort surveys or laboratory experiments. Generalizing these results to a viewer watching rich and varied content for a full-length motion picture is important but not straightforward.

Motion pictures, S3D or not, are not normally viewed from the seat equivalent to the center of the perspective projection. Banks et al. (2009) has shown that we do not experience the distortion predicted from perspective geometry as we view the image off axis. This is expected from our ability to watch television. As we move our head, the percept is consistent with a flat picture and we have presumably learned a type of constancy where we interpret the image as essentially a projection normal to the plane. In S3D content, the screen plane is shattered and we see vivid depth. Banks et al found that when seated off-axis while viewing a simple S3D scene, observers saw the scene according to the stereo geometry. One can demonstrate this by translating the head side to side while viewing an S3D display. The scene rotates with you and distorts as the 3D world morphs to be appropriate with the current viewpoint. Banks et al. found essentially none of the constancy effect they found for 2D images with their simple hinged surface stimulus. It remains to be determined whether stronger perspective information could produce partial perspective constancy in rich media like S3D films or whether the difference in the viewer's amount of experience with 2D and S3D media plays a role.

Focusing and Fixating on Stereoscopic Images: What We Know, and Need to Know

Technological advances have improved stereo media since previous 3D fads. Nonetheless, problems of discomfort and fatigue (and poor stereoscopic depth perception) remain prevalent. For example, in a recent large-scale survey (n > 7000) by the Russian movie website *Kinopoisk.ru*, 36% of respondents reported experiencing headaches or eye tiredness while watching stereo 3D movies (Berezin, 2010). As stereo 3D viewing enters the mainstream, and becomes a daily activity for the general population, there is a need to better understand how the human visual system responds to stereoscopic media if safe and effective content is to be developed.

Vergence-accommodation Conflicts

There are several reasons why viewing stereo media can have unpleasant effects on viewers (Lambooij et al., 2009; Mendiburu, 2009). Arguably the most important is the unnatural stimulus to the eye's focusing response. When we look at objects that are nearer or farther away, our eyes make two distinct oculomotor responses. The muscles in our eyes change the shape of the lens to try to focus the image on the retinas: a process called *accommodation*. At the same time, we rotate our two eyes equal-and-opposite amounts to try to bring the object of interest to the center of each retina: referred to as *vergence.* In natural viewing, we accommodate and converge to the same distance. Stereo cinema and TV systems, however, present images on a single, fixed image plane (the screen) and so viewers must often make vergence eye movements to one distance—to an object nearer than the screen for example—while accommodating at a different distance—the screen surface (Figure 11). Thus, there is mismatch between the stimulus to accommodation and the stimulus to vergence; this is the *vergence-accommodation conflict*.

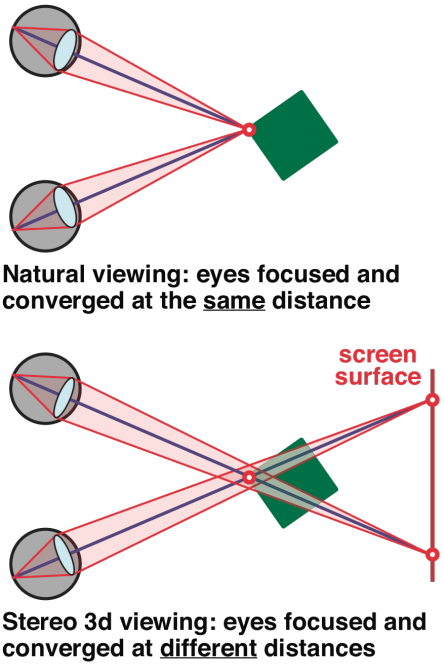

Figure 11. The vergence-accommodation conflict in stereoscopic displays. In natural viewing (upper panel), vergence and accommodation are to the same distance (upper panel). In stereo displays these two oculomotor responses must be decoupled (lower panel) for the viewer to have clear, single binocular vision.

Accommodation and vergence do not operate independently, but are synergistically coupled. Under natural-viewing conditions, each response makes the other quicker and more precise. The "decoupling" of accommodation and vergence required by stereo media is difficult and effortful for many people and has been shown to cause discomfort and fatigue, and to degrade the percept of depth (Akeley et al., 2004; Emoto et al., 2005; Hoffman et al., 2008; Ukai, 2007; Ukai & Howarth, 2008; Wann & Mon-Williams, 1997; Watt et al., 2005; Yano et al., 2004).

There are other causes of aversive symptoms in stereo media, including: (*i*) misalignments/mis-scaling of the two eyes' images that arise from differences in the optics in pairs of stereo cameras, or inaccuracies in camera rigs, (*ii*) non-natural binocular disparities due to, for example, camera toe-in, (*iii*) cross-talk, or 'ghosting', where imperfect separation of the two eyes' images results in the left eye's image being partially visible to the right eye and *vice versa*. Tractable solutions exist for these problems, however. Modern display technologies (active LC shutter glasses, polarizing or chromatic filters on the projector/TV/glasses, and line-by-line pattern polarization on some TVs) have all but eliminated ghosting. Perhaps more significantly, the switch to digital film (and displays) means that distortions and misalignments of the left- and right-eye's images can be fixed in post-production, using stereo image-processing software. In contrast, the vergence-accommodation conflict is fundamental to all existing stereo cinema and TV systems. Some researchers have developed multi-focal-plane displays (Akeley et al., 2004; Liu et al., 2008; Love et al., 2009; MacKenzie et al., 2010; McDowall & Bolas, 1994) that can successfully eliminate the conflict (MacKenzie et al., 2010, 2011). However, these displays do not permit multiple viewers, or even multiple single viewpoints, and so do not offer a practical solution for cinema and TV.

If vergence-accommodation conflicts in cinema and TV cannot be eliminated, we must instead understand and quantify the exact conditions that cause aversive side effects. This will allow guidelines to be developed for the amount of variation in stereo depth that is acceptable, the timescale over which variations can occur, and whether some sections of the population are more affected than others.

It has long been suspected that vergence-accommodation conflicts cause fatigue and discomfort, but it has only recently been confirmed empirically. Most studies of fatigue/discomfort compare effects of viewing stereo images with viewing normal 2D images. This approach is problematic for two reasons. First, 2D viewing typically differs from stereo viewing in several ways that can, themselves, cause aversive symptoms (including ghosting, motion judder from the temporal properties of the stereo system, and incorrectly aligned stereo images). Determining the effects of the vergence-accommodation conflict unambiguously therefore requires that the conflict be manipulated while keeping all other stimulus properties constant. Second, increased discomfort from viewing stereo 3D media, compared to 2D viewing, could result not from the vergence-accommodation conflict *per se*, but simply from the requirement to make vergence eye movements to fixate objects nearer and farther away. To rule out this possibility, the eye movements must also be equivalent in the two conditions. Thus, conventional stereo viewing should also be compared to equivalent real-world viewing conditions in which accommodation and vergence demands are varied together.

Hoffman et al. (2008) used a multiple-focal-plane display to create real-world variations in accommodation and vergence, while holding all other stimulus properties constant. They compared viewers' reports of fatigue and discomfort in two viewing conditions: (*i*) conventional stereo display conditions, in which the stereoscopic depth of points in the images varied, but the accommodation distance (screen distance) was fixed, and (*ii*) real-world conditions, in which the accommodative distance varied with the variations in stereoscopic depth. In the conventional-display condition, viewers reported significantly higher levels of symptoms related to visual fatigue, indicating that vergence-accommodation conflicts *per se* can cause these aversive symptoms. They also showed that vergence-accommodation conflicts degrade depth perception, causing a reduction in the ability to discriminate fine detail in stereoscopic depth (stereoacuity), and increased time to fuse stereo images (see also Akeley et al., 2004; Watt et al., 2005).

Decoupling vergence and accommodation responses. The eyes must focus and converge reasonably accurately or the resulting perceptual experience will be very poor. The accommodation error must be within the eye's depth of focus—approximately ±0.25 D (diopters; Campbell, 1957; Charman & Whitefoot, 1977)—for the image to appear clear and sharp. And the vergence error must be within Panum's fusion area (0.25-0.5 deg, or 0.07-0.14 D) or stereoscopic fusion does not occur, resulting in double vision (diplopia). The coupling of the accommodation and vergence systems means that these two responses cannot be varied entirely independently, and so with large conflicts in the stimuli to accommodation and vergence, stereo images are likely to appear blurred, diplopic, or both. It is therefore critical to understand the range within which accommodation and vergence responses can be decoupled without causing aversive side effects. Most of what we known about this comes from ophthalmological studies, designed to establish limits for prism and lens prescriptions for spectacles (Scheiman & Wick, 1994). This work has given rise to two important concepts: the zone of clear single binocular vision (ZCSBV), and Percival's zone of comfort. The ZCSBV describes the extent to which accommodation and vergence responses can be decoupled while maintaining a clear, single binocular percept. It describes the maximum attainable decoupling of accommodation and vergence responses, but fatigue/discomfort can be induced with much less decoupling. Based on experiments with prescribing spectacles, Percival (1892) suggested that the middle third of the ZCSBV represented the range of vergence-accommodation postures that could be achieved without causing discomfort. This is referred to as Percival's zone of comfort (Figure 12).

The stereoscopic zone of comfort. Percival's zone is useful conceptually, but it may be of only limited value in describing the zone of comfort (ZoC) for stereo displays. Vergence-accommodation conflicts resulting from lens or prism corrections in spectacles are likely to be easier for the system to adapt to (by adapting the vergence accommodation-coupling) because (*i*) they introduce a fixed offset between the stimuli to vergence and accommodation, whereas in

stereo viewing the conflict constantly changes, and (*ii*) spectacles are worn continuously, and so people are exposed to a constant conflict for very long durations, while stereo viewing occurs for relatively short durations. Thus, it is important to measure the ZoC for stereo viewing in a relevant context.

To our knowledge, only one study has attempted to map out the ZoC for stereo displays, while appropriately isolating the vergence-accommodation conflict. Shibata et al. (2011) used an adaptive optics multiple-focal-plane display (Love et al., 2009), and recorded subjective ratings of discomfort (with questionnaires) as a function of (*i*) viewing distance, and (*ii*) the sign of the conflict (stereo objects nearer vs. farther than the display surface). They found effects of both factors. A given vergence-accommodation conflict resulted in overall slightly higher fatigue/discomfort ratings at far viewing distances than at near. The sign of conflict also had a small but significant effect that interacted with viewing distance. At near distances, fatigue/discomfort ratings to a given conflict magnitude were greater for objects nearer than the screen, and at far distances they were greater for objects farther than the screen. Interestingly, this asymmetry was related to individual's phorias. Phoria is the vergence position adopted by the eyes when there is no stimulus to vergence but accommodation is stimulated. Thus, a person's phoria can be thought of as the extent to which his/her accommodation and vergence responses are naturally decoupled at different accommodation distances (Scheiman & Wick, 1994). Although there are significant individual differences in phoria, the typical pattern is to converge farther than the accommodation distance at near distances, and nearer than the accommodation distance at far distances (Sheedy & Saladin, 1977; Figure 12). Thus, one might expect, as Shibata et al. found, that it is most demanding to converge nearer than the screen distance at near viewing distances and farther than the screen at far viewing distances.

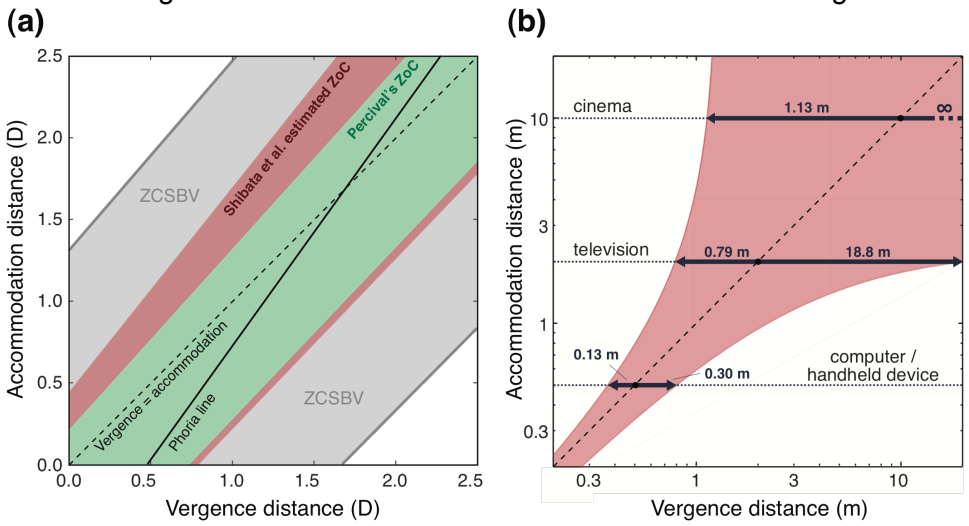

Figure 12. Zone of comfort. The left panel plots accommodation distance as a function of vergence distance, both in diopters. It shows an estimate of the zone of clear single binocular vision (ZCSBV) in gray. It also shows Percival's zone of comfort in green and the estimate of the zone of comfort for stereo 3D viewing from Shibata et al. (2011) in red. The phoria line for a typical viewer is also shown. The right panel shows the zone of comfort from Shibata et al. when plotted in units of distance rather than diopters. The horizontal lines represent the typical viewing distances for various common devices.

Figure 12 plots the ZCSBV, and Percival's zone, as estimated from the literature by Shibata et al. It also plots Shibata et al.'s estimate of the stereoscopic ZoC, based on their questionnaire data. This estimate is approximate because it is based on noisy questionnaire data, and relatively few measurements (fatigue ratings to just one conflict magnitude for each distance and sign of conflict), but it nonetheless represents the current best guess of the shape of the ZoC.

Screen distance and the ZoC. The left panel of Figure 12 plots the various zones in units of diopters—the reciprocal of distance in meters. Using diopters is appropriate because the

amount of blur in the retinal image is proportional to defocus in diopters, not physical distance. Changes in vergence angle have a very similar relationship with physical distance. Thus, a given change in the dioptric distance to a stimulus requires approximately the same change in accommodation and/or vergence independent of the overall distance to the stimulus. Perhaps unsurprisingly then, the width of the comfort zone is quite similar (though not identical) in diopters for screens positioned at different distances. This has important implications for the width of the ZoC in physical distance for different viewing situations. The right panel of Figure 12 plots Shibata et al.'s ZoC estimate as a function of physical distance. In meters, the width of the comfort zone is very small at near viewing distances. It is of course much larger at far viewing distances, but it can be seen that it is still perfectly possible to exceed the ZoC at TV and even cinema viewing distances (by presenting objects too near to the viewer). Thus, the often-made assumption that vergence-accommodation conflicts do not matter at far viewing distances is not true.

The zone of comfort in cinematography. The television and movie industry is well aware that large vergence-accommodation conflicts are problematic, but there is no commonly agreed rule to deal with them. Widespread practice appears to be to control the maximum amount of horizontal disparity as a proportion of screen width, so that screen parallax (the horizontal separation between left- and right-eye's image points on the screen) is within 2-3% of the screen width for objects nearer than the screen, and 1-2% of screen width for objects farther than the screen (Mendiburu, 2009; ATSC report, 2011). This rule-of-thumb has practical value to filmmakers because the range of on-screen parallax resulting from a given scene/camera configuration can be examined readily (i.e., on the film set) by overlaying each eye's image on a standard monitor. This rule is fundamentally incorrect, however, because it does not take into account the size of the screen that the content will be displayed on, or the viewing distance. The disparities at the viewer's eyes (and therefore the vergence-accommodation conflict) depend on the differences in angular direction of image points at the two eyes, and so they will vary considerably if the same on-screen parallax—specified in pixels, or as a proportion of screen size—is viewed on a small vs. large screen, or at a near vs. far viewing distance. In practice, the consequences of using this incorrect rule may not be catastrophic because we tend to view large screens at farther distances than we view small screens (that is, the screen size, measured in visual angle, does not vary dramatically; Ardito, 1994). But importantly, the asymmetry in tolerance to stereo depths nearer and farther than the screen (Shibata et al., 2011) varies with viewing distance. Clearly this has implications for how content should be optimized for different viewing situations, including scaling movies down to television format; even if the screen has constant angular size, different on-screen parallax limits may be needed for near (computer or TV) and far (cinema) viewing.

What We Need to Know to Specify General Guidelines

Existing studies demonstrate that the underlying concept of a *zone of comfort* for accommodation and vergence responses is valid and useful. They fall well short, however, of the specific knowledge required for comprehensive guidelines on producing stereo 3D content.

Factors predicting an individual's susceptibility to aversive symptoms remain largely unknown. There are large individual differences in a range of ophthalmological variables that could conceivably affect a person's susceptibility to discomfort from vergence-accommodation conflicts. For instance, people's ability to decouple accommodation and vergence responses differs significantly, as do their phorias and their ability to accommodate to different distances (Scheiman & Wick, 1994). Large-scale population studies are required to establish the relationships between these ophthalmological variables and aversive symptoms during stereo viewing. If there are indeed large differences in each individual's ZoC, the placement of content in stereo depth may need to be conservative to remain acceptable to the majority of people.

The viewer's age is likely to be a particularly important factor. There is a belief in the stereo industry that older viewers are more affected by vergence-accommodation conflicts than younger viewers. For instance, "oculo-motor exercising [decoupling accommodation and

vergence] can be painful and can increase in difficulty with age. Kids would just not care, when elderly persons may be unable to practice it" (Mendiburu, 2009). In fact, the opposite is probably true. The ability to vary one's accommodation state decreases significantly with age so, under natural viewing, older adults experience vergence-accommodation conflicts most of the time (because they cannot vary their accommodation response with vergence when looking nearer and farther). Indeed, their oculomotor responses more closely resemble those required for viewing stereo media: changing vergence while accommodating to a fixed distance. Consistent with this, Yang et al. (2011) recently found that people aged 24–34 years reported more discomfort than people aged 45 and over when viewing the same stereo 3D content.

It also remains to be determined whether there are any short- or long-term effects of prolonged, repeated exposure to the unnatural stimulus presented by stereoscopic displays. In adults, accommodation-vergence coupling is quite adaptable (Schor & Kotulak, 1986), and so there is a possibility that accommodation function may take some time to return to normal following prolonged viewing of stereo 3D media. Moreover, as the stereo 3D industry continues to develop, our use of stereo media will change from an occasional activity to an everyday one. The introduction of stereo computer games, in particular, will expose viewers to vergence-accommodation conflicts regularly for potentially long periods of time. We may need to be particularly cautious about long-term effects of vergence-accommodation conflicts on younger children because their visual systems are still developing (Rushton & Riddell, 1999). We know of no specific causes for concern at this time, but the research required to identify relevant issues has not yet been done. It is reasonable to assume that vergence-accommodation coupling exists because it is beneficial, and so one should be cautious when systematically disrupting its natural operation. Of course, the ZoC could be measured in children in the same way it has been measured in adults. Clearly, however, it would not be acceptable to carry out the long-term experimental studies that would be required to understand any potential long-term effects (although, ironically, young people may expose themselves to such a regime voluntarily). Thus, clinical, research, and industry communities should remain alert to the development of unwanted symptoms in users of stereo media.

Temporal Presentation Protocols: Flicker, Motion Artifacts, and Depth Distortions

Temporal Protocols in Stereo Displays

It is clearly desirable to be able to present flicker-free image content without noticeable motion artifacts or distortions of perceived depth. Here we investigate how the means of presenting stereo images over time affects the visibility of flicker, motion, and depth.

Stereo 3D (S3D) displays generally use one conventional 2D display to present different images to the left and right eyes. Because S3D displays are so similar to conventional non-stereo displays, many of the standards, protocols, technical analyses, and artistic effects that have been developed for non-stereo displays also apply to S3D. There are, however, important differences between non-stereo and stereo displays that can produce artifacts unique to stereo presentation.

There are a variety of ways to present different images to the two eyes. The field-sequential approach presents images to the left and right eyes in temporal alternation (e.g., RealD, Dolby). Among field-sequential approaches, there are several ways to present the alternating images in time including multiple-flash methods. In addition to field-sequential approaches, one can present images to the two eyes simultaneously by using multiple projectors (IMAX), wavelength-multiplexing techniques (Infitec and anaglyph), or spatial multiplexing (micropol) on one 2D display. Figure 13 schematizes some protocols. Column 6 shows the RealD and Dolby approach. The IMAX approach is similar to column 1.

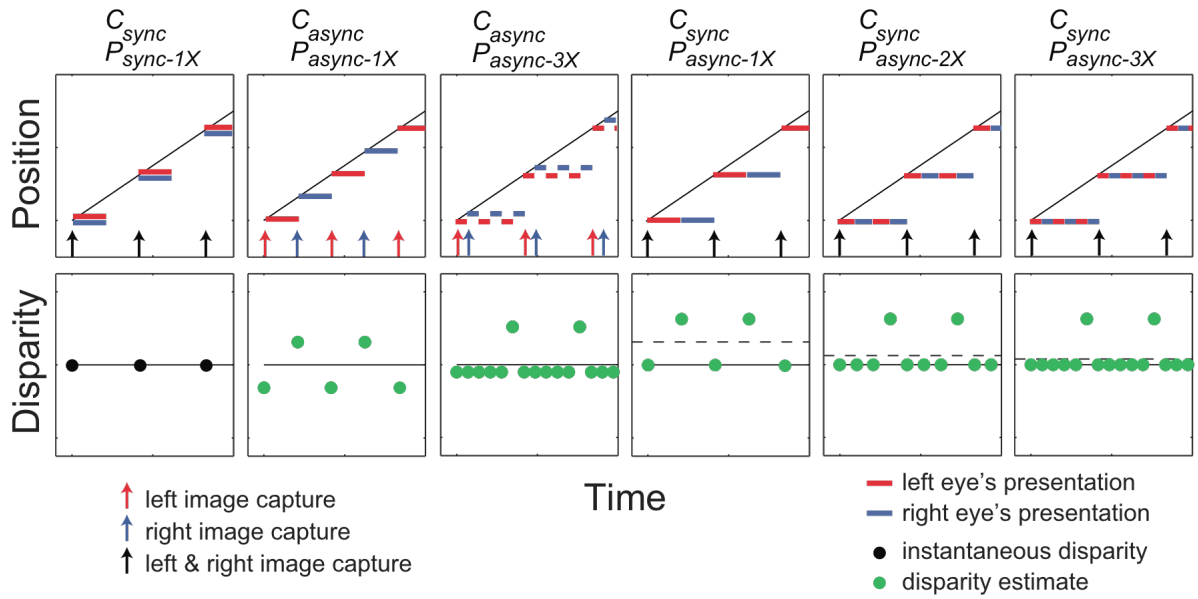

Figure 13. Temporal protocols used in S3D displays. The columns represent different protocols. Upper row: Each panel plots the position of stimulus moving at constant speed in the plane of the screen as a function of time. The red and blue line segments represent the presentations of the images to the left and right eyes, respectively. The arrows indicate the times at which the stimulus was captured (or computed). Black arrows indicate left and right images captured simultaneously. The red and blue arrows indicate left and right images captured in alternating fashion. The black diagonal lines represent the correct positions for the left and right images as a function of time. Lower row: Each panel plots disparity as a function of time. The black horizontal lines represent the correct disparities. The black dots represent the disparities when the two eyes' images are presented simultaneously. The green dots represent the disparities that would be calculated if the left-eye image is matched to the successive right-eye image and the right-eye image is matched to the successive left-eye image. The dashed horizontal lines represent the time-average disparities that would be obtained by such matching. Wherever a horizontal line is not visible, the average disparity is the same as the correct disparity, so the two lines superimpose.

Spatio-temporal frequencies. To examine how various temporal presentation methods affect the viewer's perceptual experience with stereo displays, it is very useful to examine the temporal and spatial frequencies created by these methods. We begin by considering stroboscopic presentation of a moving object presented to one eye (Watson et al., 1986). To create the appearance of a high-contrast vertical line moving smoothly at speed s, one presents a sequence of very brief snapshots of the line at time intervals of Δt with each view displaced by $\Delta x = s\Delta t$. The temporal presentation rate τ_p is the reciprocal of the time between presentations: $\tau_p = 1/\Delta t$.

The left panel of Figure 14 depicts a real stimulus moving at speed s and the stroboscopic version of that stimulus. Spatial position is plotted as a function of time. By using the Fourier transform, we determine the temporal and spatial frequencies in these two stimuli. These are depicted in the right panel of Figure 14. Spatial frequency is plotted as a function of temporal frequency. The Fourier transform of the real moving stimulus is the black line; it has a slope of $-1/s$. The transform of the stroboscopic stimulus is represented by the black and green lines. The black line is the same line as for the real stimulus. The green lines are *aliases*: artifacts created by the stroboscopic presentation. Their slopes are $-1/s$ and they are separated horizontally by τ_p. Thus, the spatiotemporal frequencies of the stroboscopic stimulus contain a signal component (the black line) plus a series of aliases (the green ones). As the speed of the stimulus (s) increases, the slope of the signal and aliases decreases. As the presentation rate (τ_p) increases, the separation between the aliases increases. When the aliases are visible to a viewer, the percept will contain flicker and/or motion artifacts. When the aliases are not visible, the percept will be non-flickering and smooth.

To assess the visibility of the aliases, we need to consider what human viewers can and cannot see. The system's sensitivity to different temporal and spatial frequencies is described by the spatio-temporal contrast sensitivity function (CSF). Figure 15 plots the CSF for a typical viewer under room-light conditions. This function has been called the *window of visibility* because it characterizes the spatio-temporal stimuli that can be seen as opposed to the ones that cannot be seen. The CSF is represented in Figure 14 by the white ellipse. The dimensions of the ellipse's principal axes are the highest visible temporal frequency (*cff*) and the highest visible spatial frequency (*va*). Aliases falling within the ellipse will generally be visible and those falling outside will not. We can now see how a stroboscopic stimulus could appear identical to a smoothly moving real stimulus. If the two are moving at the same speed and the stroboscopic presentation τ_p is fast enough, the aliases would fall outside the window of visibility and therefore the stroboscopic and real stimuli could not be discriminated. Similarly, the stroboscopic and real stimuli could be discriminated when the aliases fall within the window of visibility: the stroboscopic stimulus would exhibit flicker and/or motion artifacts.

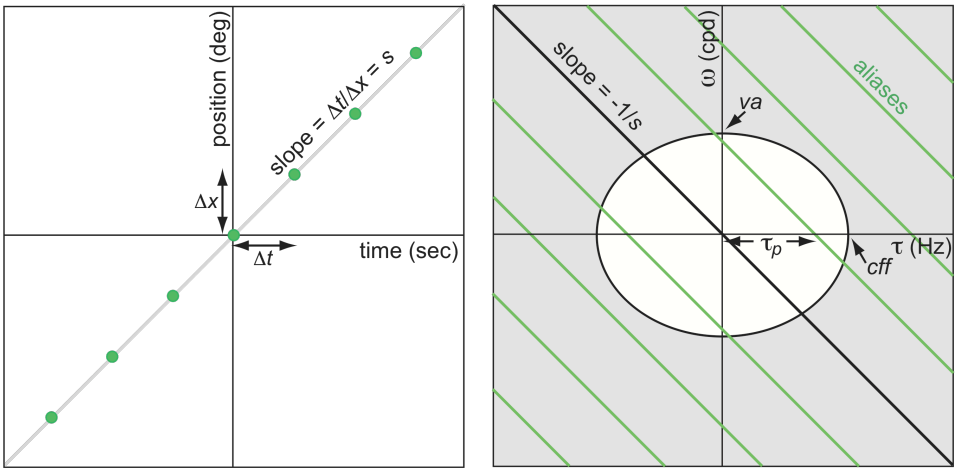

Figure 14. Properties of a smoothly moving stimulus and a stroboscopic stimulus. The gray diagonal line in the left panel represents the motion of a smoothly moving vertical line on axes of time and horizontal position. The green dots represent the stroboscopic presentation of that stimulus; brief flashes occur at multiples of Δ*t*. The right panel shows the Fourier transform (technically the amplitude spectrum) for the smoothly moving and stroboscopic stimuli plotted on axes of temporal frequency (cycles/sec or Hz) and spatial frequency (cycles/deg or cpd). The black diagonal line represents the temporal and spatial frequencies of the smoothly moving stimulus. The green lines are the additional frequencies from the stroboscopic stimulus; they are temporal aliases separated by $\tau_p = 1/\Delta t$. The ellipse contains combinations of temporal and spatial frequency that are visible to the visual system. The highest visible temporal frequency is indicated by *cff* and the highest visible spatial frequency by *va*. The shaded region contains combinations of temporal and spatial frequency that are not visible.

We discussed stroboscopic presentation in Figure 14 because such presentation determines the spatio-temporal frequencies of the aliases and those frequencies remain for other protocols that are actually used in S3D displays. Other presentation methods (e.g., sample and hold, multi-flash) do not change the pattern of aliases; they only change their amplitudes.

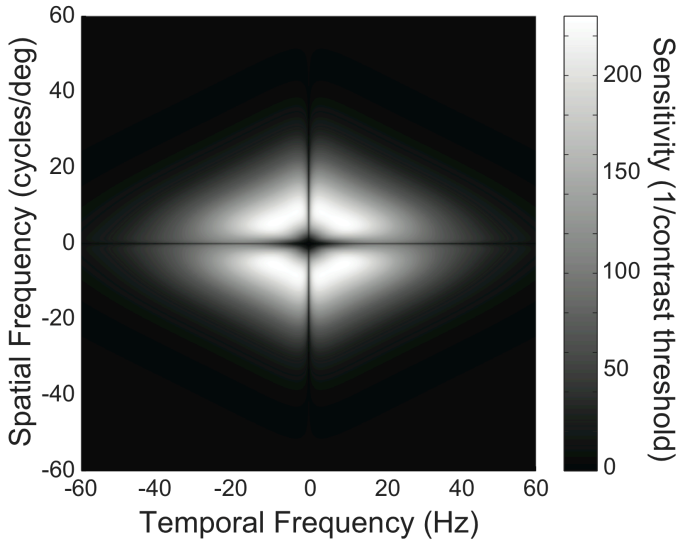

Figure 15. The human spatio-temporal contrast sensitivity function. The sensitivity to a moving sinusoidal grating is plotted as a function of temporal frequency and spatial frequency. Sensitivity is the reciprocal of the contrast required to detect the stimulus and is represented by gray scale; brighter values corresponding to higher sensitivity. Adapted from Kelly (1979).

Flicker Visibility

Let us now consider when flicker will be visible in an S3D display. We define visible flicker as perceived fluctuations in the brightness of the stimulus. We assume that flicker is perceived when aliases such as those in the right panel of Figure 14 encroach the window of visibility near a spatial frequency of zero (i.e., along the temporal-frequency axis).

In field-sequential stereo displays (e.g., active shutter glasses or passive glasses with active switching in front of the projector), the monocular images consist of presentation intervals alternating with dark intervals. In some cases, each presented image of the moving stimulus is a new one; we refer to this as a *single-flash protocol*; it is schematized in the upper left panel of Figure 16. There is also a *double-flash protocol* in which the images are presented twice before updating and a *triple-flash protocol* in which they are presented three times before updating. We will use f to represent the number of such flashes in a protocol. Those protocols are also schematized in the left column of Figure 16 and in Figure 13. Of course, multi-flashing is similar to the double and triple shuttering that is done with film-based movie projectors. The double- and triple-flash protocols are used to reduce the visibility of flicker (RealD and Dolby use triple-flash, and IMAX uses field-simultaneous double-flash). We will refer to the rate at which new images are presented as the *capture rate* τ_c (or $1/t_c$, where t_c is the time between image updating). We will refer to the rate at which images, updated or not, are delivered to an eye as the *presentation rate* τ_p (or $1/t_p$). Note that $\tau_c = \tau_p/f$ (or $t_c = ft_p$).

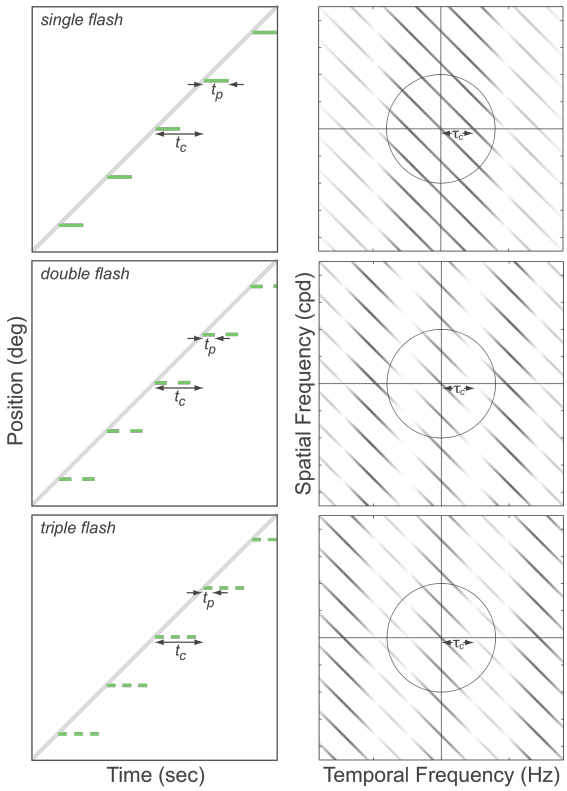

Figure 16. Properties of stimuli presented with multiple-flash protocols. The left panels schematize the single-, double-flash, and triple-flash protocols. In each case, the same images are presented during the interval t_c until updated images are presented in the next interval. In multi-flash protocols, the duration of each image presentation t_p is t_c/f, where f is the number of flashes. The right panels show the corresponding Fourier transforms of the multi-flash stimuli plotted as a function of temporal and spatial frequency. The transform of a smoothly moving real stimulus would again be a diagonal line with slope –1/s. Amplitude is represented by gray scale, dark values corresponding to higher amplitudes. The presentation rate τ_p (or $1/t_p$) is indicated by arrows. The aliases are separated by τ_c ($1/t_p$), which is also indicated by arrows. The circles represent the window of visibility.

The Fourier transform for the single-flash, field-sequential protocol is shown in the upper right panel of Figure 16. With the insertion of dark frames, the amplitude of the aliases at a temporal frequency of τ_c is rather high. As a result, flicker should be quite visible, whether the stimulus is moving or not, unless a high presentation rate is used. The transforms for the double- and triple-flash protocols are also shown in the right column of Figure 16. The frequencies of the aliases are the same in the single- and double-flash protocols, but their amplitudes go to zero at τ_c in double flash and at $2\tau_c$ in single flash. In the triple-flash protocol, the aliases are again the same, but their amplitudes go to zero at τ_c and again at $2\tau_c$, remaining small in between. The first alias with non-zero amplitude along the temporal-frequency axis occurs at a temporal frequency of τ_p ($1/t_p$), which is the presentation rate. Thus, we predict that presentation rate, not capture rate, will determine flicker visibility. This prediction was confirmed in a perceptual experiment (Hoffman et al., 2011). Because presentation rate should be the primary determinant, one should be able to reduce flicker visibility for a fixed capture rate by using multi-flash protocols. Specifically, flicker should be less visible in the triple-flash than in the double-flash protocol and less visible in the double-flash than in the single-flash protocol. This prediction has been shown to be correct (Hoffman et al., 2011).

Stereo processing in the visual system is sluggish (Tyler, 1974) and therefore the visual system is less sensitive to rapidly changing disparities than to time-varying luminance signals. From this observation, we predict little if any difference in flicker visibility between stereo and

non-stereo presentations provided that the temporal protocols are the same. This prediction is basically correct (Hoffman et al., 2011).

Motion Artifacts

We now turn to the visibility of motion artifacts. These artifacts include judder (jerky or unsmooth motion appearance), edge banding (more than one edge seen at the edge of a moving stimulus), and motion blur (perceived blur at a moving edge). The analysis of motion artifacts is somewhat more complicated than the one for flicker because with a given capture rate, multi-flash protocols do not change the spatio-temporal frequency of the aliases. Instead they differentially attenuate the amplitudes of the aliases at certain temporal frequencies. Thus, the visibility of motion artifacts is determined by the spatio-temporal frequencies and amplitudes of the aliases.

Viewers typically track a moving stimulus with smooth-pursuit eye movements that keep the stimulus on the fovea, and this affects what motion artifacts look like. With smooth pursuit, the image of a smoothly moving stimulus becomes fixed on the retina; that is, for a real object moving smoothly at speed s relative to the observer, and an eye tracking at the same speed, the retinal speed of the stimulus is 0. With a digitally displayed stimulus moving at the same speed, the only temporally varying signal on the retina is created by the difference between smoothly moving and discretely moving images. Each image presentation of duration t_p displaces across the retina by $\Delta x = -st_p$. Thus, significant displacement can occur with high stimulus speeds and low frame rates thereby blurring the stimulus on the retina ("motion blur"; Klompenhouwer & Velthoven, 2004; Feng, 2006).

From the analysis of temporal and spatial frequencies, we can make a number of predictions about the visibility of motion artifacts. (1) The visibility of motion artifacts should increase with increasing stimulus speed and decrease with increasing capture rate. More specifically, combinations of speed and capture rate that yield a constant ratio (s/τ_c) should have approximately equivalent motion artifacts. Hoffman et al. (2011) tested this prediction and found that it is essentially correct. (2) Although speed and capture rate should be the primary determinants of motion artifacts, multi-flash protocols for a fixed capture rate should produce more visible motion artifacts. This too has been tested empirically and found to be correct (Hoffman et al., 2011). (3) Edge banding should be determined by the number of flashes in multi-flash protocols: two bands being perceived with double-flash, three with triple-flash, etc. This prediction is borne out empirically (Feng, 2006; Klompenhouwer, 2006; Hoffman et al., 2011).

Distortions of Perceived Depth

A temporal delay to one eye's input can cause a moving object to appear displaced in depth (Morgan, 1979; Burr & Ross, 1979). Many of the protocols in Figure 13 introduce such a delay to one eye. If the delay alters the visual system's estimate of the disparity over time, this would in turn produce distortion in the depth percept. Consider, for example, the C_{sim}/P_{alt-1X} protocol (fourth column in Figure 13). The solid horizontal line in the lower panel represents the correct disparity over time; i.e., the disparity that would occur with the presentation of a moving real object in the plane of the screen. To compute disparity, the visual system must match images in one eye with images in the other. But the images in this protocol are presented to the two eyes at different times, so non-simultaneous images must be matched. If each image in one eye is matched with the succeeding image in the other eye, the estimated disparities would be the green dots in the lower panel. For every two successive matches (three images), one disparity estimate is equal to the correct value and one is greater. As a result, the time-average disparity is biased relative to the correct value, and this should cause a change in perceived depth: a perceptual distortion. Notice that the difference between the time-average disparity and the correct disparity depends on the protocol: largest with single flash (Figure 13, column 4) and smallest with triple flash (column 6). For this reason, the largest distortions should occur with single-flash protocols and the smallest with triple-flash protocols. The magnitude of the

distortions should also depend on speed because the difference between the time-average disparity and the correct disparity is proportional to speed. We will refer to the distortions predicted from the average disparity over time as the *time-average model*.

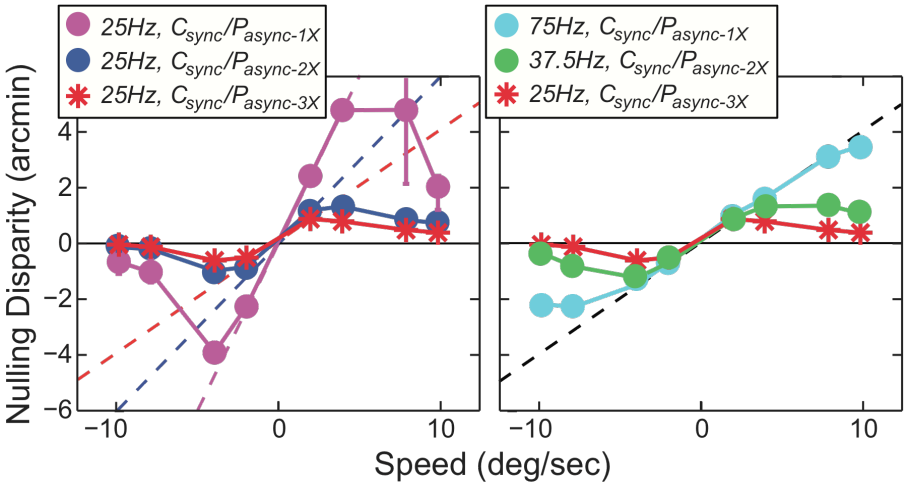

Figure 17. Distortions of perceived depth with simultaneous capture and alternating presentation. The disparity distortion is plotted as a function of the speed of a stimulus moving in the plane of the display screen. The left panel shows the data from protocols with a 25-Hz capture rate. The purple circles represent the data with the single-flash protocol (C_{sim}/P_{alt-1X}). The blue circles represent the data with the double-flash protocol (C_{sim}/P_{alt-2X}). The red asterisks represent the data from the triple-flash protocol (C_{sim}/P_{alt-3X}). The predictions for the time-average disparity model (lower row of Figure 13) are the dashed lines with the colors corresponding to the appropriate temporal protocol. The right panel shows the data from the same protocols, but with different capture rates. In each case, the presentation rate was 75 Hz, so the right-eye's image was delayed relative to the left-eye's image by 1/150 sec. The predictions for the time-average model are the dashed line. The cyan circles, green circles, and red asterisks are the data from the single-, double-, and triple-flash protocols, respectively.

The most frequently used protocols employ simultaneous capture and alternating presentation where one eye's image is delayed. Figure 17 plots the predictions and data from such protocols with different capture rates and numbers of flashes. The predictions from the time-average model are the dashed lines. Experimental data in which the magnitude of the depth distortion was measured are represented by the colored symbols. As expected, the size of the distortion increases as stimulus speed increases. With 75-Hz capture, the distortion increases up to the fastest speed tested. With 25-Hz capture, the distortion levels off at ~3 deg/sec and then decreases at yet higher speeds. We conclude that perceived depth distortions do occur, as predicted by the time-average model, when capture and presentation synchrony are not matched. The model's predictions are accurate at slow speeds, but smaller distortions than predicted are observed at fast speeds. The prediction failure at fast speeds is the consequence of a temporal disparity-gradient limit (Hoffman et al., 2011).

Thus, distortions of perceived depth occur with moving objects in some stereo presentation protocols because they delay the input to one eye relative to the other eye. As a consequence, objects moving in one direction can be perceived as closer than they are meant to be and objects moving in the opposite direction can be perceived as farther than they are meant to be. Such distortions can be readily observed in stereo TV and cinema. For example, in S3D broadcasts of World Cup soccer in 2010, a ball kicked along the ground would appear to recede in depth when moving in one direction (paradoxically seeming to go beneath the playing field!) and would appear to come closer in depth when moving in the opposite direction. This speed-dependent effect can be quite disturbing, so it is clearly useful to understand its cause and how one might minimize or eliminate it.

Summary of Temporal Protocols

The work described here adds to the theoretical and empirical foundation for determining what display parameters are likely to yield noticeable flicker, motion artifacts, and depth distortions. From this foundation, one can make effective decisions about how to minimize or even eliminate these undesirable effects.

References

Advanced Television Systems Committee (ATSC) report on 3D digital television (2011). *ATSC Planning Team 1 Interim Report*. Retrieved from "http://www.atsc.org/PT1/PT-1-Interim-Report.pdf".

Akeley, K. Watt, S. J., Girshick, A. R., & Banks, M. S. (2004). A stereo display prototype with multiple focal distances. *ACM Transactions on Graphics*, 23, 804–813.

Allison, R. S. (2007). Analysis of the influence of vertical disparities arising in toed-in stereoscopic cameras. *Journal of Imaging Science & Technology*, 51, 317-327.

Allison, R. S. & Howard, I. P. (2000). Temporal dependencies in resolving monocular and binocular cue conflict in slant perception, *Vision Research*, 40, 1869-1885.

Allison, R. S., Howard, I. P., & Fang, X. (2000). Depth selectivity of vertical fusional mechanisms. *Vision Res, 40*, 2985-2998.

Allison, R. S., Howard, I. P., Rogers, B. J. et al. (1998). Temporal aspects of slant and inclination perception, *Perception*, 27, 1287-1304.

Ardito, M. (1994). Studies of the influence of display size and picture brightness on the preferred viewing distance for HDTV programs. *SMPTE Journal,* 103, 517-522.

Backus, B. T. & Banks, M. S. (1999). Estimator reliability and distance scaling in stereoscopic slant perception. *Perception, 28*, 217-242.

Backus, B. T., Banks, M. S., van Ee, R., & Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Res, 39*, 1143-1170.

Banks, M. S. & Backus, B. T. (1998). Extra-retinal and perspective cues cause the small range of the induced effect. *Vision Res, 38*, 187-194.

Banks, M. S., Held, R. T., & Girshick, A. R. (2009). Perception of 3-D layout in stereo displays. *Information Display*, 25, 12-16.

Berezin, O. (2010). Digital cinema in Russia: Status of 2D/3D DC rollout. Is 3D still a driver for the development of the cinema market? *3Dmedia2010*, statistic from kinopoisk.ru.

Bradshaw, M. F., Glennerster, A., & Rogers, B. J. (1996). The effect of display size on disparity scaling from differential perspective and vergence cues. *Vision Res,* 36, 1255-1264.

Burge, J., Fowlkes, C. C., & Banks, M. S. (2010). Natural-scene statistics predict how the figure–ground cue of convexity affects human depth perception. *J Neurosci*. 30, 7269-7280.

Burr, D. C. & Ross, J. (1979) How does binocular delay give information about depth? *Vision Research,* 19, 523-532.

Bülthoff, H. H. & Mallot, H. A. (1988). Integration of Depth Modules: Stereo and Shading," *J Opt Soc Am A*, 5, 1749-1758, 1988.

Campbell, F. W. (1957). The depth of field of the human eye. *Optica Acta*, 4*,* 157–164.

Charman, W. N., & Whitefoot, H. (1977). Pupil diameter and the depth-of field of the human eye as measured by laser speckle. *Journal of Modern Optics,* 24*,* 1362–3044.

Clark, J. J. & Yuille, A. L. (1990). *Data fusion for sensory information processing systems.* Boston: Cluwer.

Emoto, M., Niida, T., & Okano, F. (2005). Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television. *Journal of Display Technology*, 1*,* 328–340.

Feng, X. F. (2006). LCD motion-blur analysis, perception, and reduction using synchronized backlight flashing. *Proceedings of the SPIE,* No. 6057.

Fisher, R. A. (1925). Theory of statistical estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 22, 700-725.

Garding, J., Porrill, J., Mayhew, J. E., & Frisby, J. P. (1995). Stereopsis, vertical disparity and relief transformations. *Vision Res,* 35, 703-722.

Geller, T. (2008). Overcoming the uncanny valley. *IEEE Comput Graph Appl,* 28, 11-17.

Girshick, A. R. & Banks, M. S. (2009). Probabilistic combination of slant information: weighted averaging and robustness as optimal percepts. *Journal of Vision*, 9, 1-20.

Glennerster, A., Tcheang, L., Gilson, S. J., Fitzgibbon, A. W., & Parker, A. J. (2006). Humans ignore motion and stereo cues in favor of a fictional stable world. *Current Biology,* 16, 428-432.

Helmholtz, H. (1909). *Physiological Optics. English Translation 1962 by J. P. C. Southall from the 3rd German Edition of Handbuch Der Physiologischen Optik. Vos, Hamburg.*, New York, Dover.

Hillis, J. M., Ernst, M. O., Banks, M. S., Landy, M. S. (2002). Combining sensory information: mandatory fusion within, but not between, senses. *Science*, 298, 1627-1630.

Hillis, J.M., Watt, S., Landy, M.S., & Banks, M.S. (2004). Slant from texture and disparity cues: Optimal cue combination*. Journal of Vision*, 4, 967-992.

Hoffman, D. L., Girshick, A. R., Akeley, K., & Banks, M. S. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision,* 8*,* 1–30.

Hoffman, D. M., Karasev, V. I., & Banks, M. S. (2011). Temporal presentation protocols: Flicker visibility, perceived motion, and perceived depth. *Journal of SID*, 19, 255-281.

Howard, I. P., Fang, X., Allison, R. S., & Zacher, J. E. (2000). Effects of stimulus size and eccentricity on horizontal and vertical vergence. *Exp Brain Res,* 130, 124-132.

Howard, I. P. & Rogers, B. J. (2002). *Depth perception*, Vol. 2. Toronto: Porteous..

Julesz, B. (1960). Binocular depth perception of computer generated patterns. *Bell System Technical Journal*, 39, 1125-1162.

Kelly, D. H. (1979). Motion and vision. II. Stabilized spatio-temporal threshold surface. *J. Opt. Soc. Am.* 69, 1340-1349.

Klompenhouwer, M. A. (2006). *Flat panel display signal processing: analysis and algorithms for improved static and dynamic resolution* (PhD Thesis). Eindhoven University Press.

Klompenhouwer, M. A. & Velthoven, L. J. (2004). Motion blur reduction for liquid crystal displays: motion-compensated inverse filtering. *Proceedings of the SPIE,* 5308, 690.

Knill, D. C. (2007). Learning Bayesian priors for depth perception. *Journal of Vision*, 7, 13.

Knill, D. C. (2007). Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *Journal of Vision*, 7, 1-24.

Lambooij, M., Ijsselsteijn, W., Fortuin, M., & Heynderickx, I. (2009). Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology 53*(3), 1–14.

Landy, M. S., Maloney, L. T., Johnston, E. B., Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, 35, 389-412.

Liu, S., Cheng, D., & Hua, H. (2008). An optical see-through head mounted display with addressable focal planes. In *International Symposium on Mixed and Augmented Reality*, IEEE, 33–41.

Longuet-Higgins, H. C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature,* 293, 133-135.

Love, G. D., Hoffman, D. M., Hands, P. J. W., Gao, J., Kirby, A. K., & Banks, M. S. (2009). High-speed switchable lens enables the development of a volumetric stereoscopic display. *Optics Express,* 17*,* 15715–15725*.*

Howard, I. P., Allison, R. S., & Zacher, J. E. (1997). The dynamics of vertical vergence. *Exp Brain Res,* 116, 153-159.

MacKenzie, K. J., Dickson, R., & Watt, S. J. (2011). Vergence and accommodation to multiple-image-plane stereoscopic displays: 'Real world' responses with practical image-plane separations? In *Stereoscopic Displays and Applications XXII, Proceedings of the SPIE,* 7863, 7863-1 – 7863-11.

MacKenzie, K. J., Hoffman, D. M., & Watt, S. J. (2010). Accommodation to multiple-focal-planes displays: implications for improving stereoscopic displays, and for accommodation control. *Journal of Vision, 10(8)*:22.

Mather, G. & Smith, D. R. (2000). Depth cue integration: Stereopsis and image blur. *Vision Research*, 40, 3501-3506.

McDowall, I., & Bolas, M. (1994). FakeSpace Labs accommodation display research. Unpublished report.

Meese, T. S. & Holmes, D. J. (2004). Performance data indicate summation for pictorial depth-cues in slanted surfaces. *Spat Vis*, 17, 127-151.

Mendiburu, B. (2009) *3D Movie Making: Stereoscopic digital cinema from script to screen*. Focal Press.

Morgan, M. J. (1979). Perception of continuity in stroboscopic motion: A temporal frequency analysis. *Vision Research,* 19, 523-532.

Muller, C. M., Brenner, E., & Smeets, J. B. (2009). Testing a counter-intuitive prediction of optimal cue combination. *Vision Res*, 49, 134-139.

Ogle, K. N. (1938). Induced size effect I: A new phenomenon in binocular vision associated with the relative size of the images in the two eyes. *Archives of Ophthalmology,* 20, 604.

Percival, A. S. (1892). The relation of convergence to accommodation and its practical bearing. *Ophthalmological Review,* 11*,* 313-328.

Read, J. C. A., Phillipson, G. P., & Glennerster, A. (2009). Latitude and longitude vertical disparities. *Journal of Vision,* 9(13), 11-37.

Rogers, B. J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspective and binocular stereopsis. *Nature,* 361, 253-255.

Rogers, B. J. & Bradshaw, M. F. (1999). Disparity minimisation, cyclovergence, and the validity of nonius lines as a technique for measuring torsional alignment. *Perception,* 28, 127-141.

Rushton, S. K., & Riddell, P. M. (1999). Developing visual systems and exposure to virtual reality and stereo displays: some concerns and speculations about the demands on accommodation and vergence. *Applied Ergonomics,* 30*,* 69–78.

Sakano, Y. & Ando, H. (2010). Effects of head motion and stereo viewing on perceived glossiness. *Journal of Vision*, 10, 1-14.

Scheiman, M., & Wick, B. (1994). *Clinical management of binocular vision: Heterophoric, accommodative, and eye movement disorders*. Philadelphia: J.B. Lippincott.

Schor, C. M., & Kotulak, J. C. (1986). Dynamic interactions between accommodation and convergence are velocity sensitive. *Vision Research,* 26*,* 927–942.

Sheedy, J. E. & Saladin, J. J. (1977). Phoria, vergence, and fixation disparity in oculomotor problems. *American Journal of Optometry & Physiological Optics*, 54, 474-478.

Shibata, T., Kim, J., Hoffman, D. M., & Banks, M. S. (2011). The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision,* in press*.*

Siegel, M. & Nagata, S. (2000). Just Enough reality: Comfortable 3-d viewing via microstereopsis. *IEEE T Circ Syst Vid*, 10, 387-396.

Tsirlin, I., Allison, R. S., & Wilcox, L. M. (2008). Stereoscopic Transparency: Constraints on the perception of multiple surfaces. *Journal of Vision*, 8, 1-10.

Tyler, C.W. (1974). Depth perception in disparity gratings." *Nature* 251, 140 –142.

Ukai, K. (2007). Visual fatigue caused by viewing stereoscopic images and mechanism of accommodation. In *Proceedings of the First International Symposium on University Communication,* 176–179.

Ukai, K., & Howarth, P. A. (2008). Visual fatigue caused by viewing stereoscopic motion images: Background, theories and observations. *Displays,* 29*,* 106–116.

van Ee, R., Adams, W. J., & Mamassian, P. (2003). Bayesian modeling of cue interaction: Bistability in stereoscopic slant perception. *J Opt Soc Am A Opt Image Sci Vis*, 20, 1398-1406.

Vishwanath, D., Girshick, A. R., & Banks, M. S. (2005). Why pictures look right when viewed from the wrong place. *Nat Neurosci,* 8, 1401-1410.

Wann, J. P., & Mon-Williams, M. (1997). Health issues with virtual reality displays: What we do know and what we don't. *ACM SIGGRAPH Computer Graphics,* 31, 53–57.

Watson, A. B., Ahumada, A. J. & Farrell, J. E. (1986). Window of visibility: psychophysical theory of fidelity in time-sampled visual motion displays. *J. Opt. Soc. Am.* 3, 300-307.

Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision*, 5, 834-862.

Yang, S.-N., Schlieski, T., Selmins, B, Cooper, S., Doherty, R. A., Corriveau, P. J., & Sheedy, J. E. (2011). Individual differences and seating position affect immersion and symptom in stereoscopic 3D viewing. *Technical Report: Vision Performance Institute, Pacific University.*

Yano, S., Emoto, M., & Mitsuhashi, T. (2004). Two factors in visual fatigue caused by stereoscopic HDTV images. *Displays,* 25, 141-150.