Journal of Imaging Science and Technology® 51(4): 317–327, 2007.

Analysis of the Influence of Vertical Disparities Arising in Toed-in Stereoscopic Cameras

Robert S. Allison
Department of Computer Science and Centre for Vision Research, York University, 4700 Keele St., Toronto, Ontario M3J 1P3, Canada
E-mail: allison@cs.yorku.ca

Abstract. *A basic task in the construction and use of a stereoscopic camera and display system is the alignment of the left and right images appropriately—a task generally referred to as camera convergence. Convergence of the real or virtual stereoscopic cameras can shift the range of portrayed depth to improve visual comfort, can adjust the disparity of targets to bring them nearer to the screen and reduce accommodation-vergence conflict, or can bring objects of interest into the binocular field of view. Although camera convergence is acknowledged as a useful function, there has been considerable debate over the transformation required. It is well known that rotational camera convergence or "toe-in" distorts the images in the two cameras producing patterns of horizontal and vertical disparities that can cause problems with fusion of the stereoscopic imagery. Behaviorally, similar retinal vertical disparity patterns are known to correlate with viewing distance and strongly affect perception of stereoscopic shape and depth. There has been little analysis of the implications of recent findings on vertical disparity processing for the design of stereoscopic camera and display systems. I ask how such distortions caused by camera convergence affect the ability to fuse and perceive stereoscopic images.* © *2007 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.(2007)51:4(317)]

INTRODUCTION

In many stereoscopic viewing situations it is necessary to adjust the screen disparity of the displayed images for viewer comfort, to optimize depth perception or to otherwise enhance the stereoscopic experience. Convergence of the real or virtual cameras is an effective means of adjusting portrayed disparities. A long-standing question in the stereoscopic imaging and display literature is what is the best method to converge the cameras? Humans use rotational movements to binocularly align the visual axes of their eyes on targets of interest. Similarly, one of the easiest ways to converge the cameras is to pan them in opposite directions to "toe-in" the cameras. However, convergence through camera toe-in has side effects that can lead to undesirable distortions of stereoscopic depth.[1,2] In this paper we reanalyze these geometric distortions of stereoscopic space in the context of recent findings on the role of vertical disparities in stereoscopic space perception. We focus on a number of issues related to converged cameras and the mode of convergence: The effect of rectification; relation between the geometry of the imaging device and the display device; fused and augmented displays; orthostereoscopy; the relation between parallax distortions in the display and the resulting retinal disparity; and the effect of these toe-in induced retinal disparities on depth perception and binocular fusion.

Our interests lie in augmented-reality applications and stereoscopic heads for tele-operation applications. In these systems a focus is on the match and registration between the stereoscopic imagery and the "real world" so we will concentrate on orthostereoscopic or near orthostereoscopic configurations. These configurations have well known limitations for applications such as visualization and cinema, and other configurations may result in displays that are more pleasing and easier to fuse. However, it is important to note that our basic analysis generalizes to other configurations, and we will discuss other viewing arrangements when appropriate.[3,4] In a projector-based display system with separate right and left projectors, or in binocular head mounted display (HMD) with independent left and right displays, the displays/projectors can also be converged mechanically or optically. In this paper we will also assume a single flat, fronto-parallel display (i.e., a monitor or projector display) so that the convergence of the projectors is not an issue. Since the left and right images are projected or displayed into the same plane we will refer to these configurations as a "parallel display." In most cases similar considerations will apply for a HMD with parallel left and right displays.

OPTIONS FOR CAMERA CONVERGENCE

We use the term convergence here to refer to a variety of means of realigning one stereoscopic half-image with respect to the other, including toe-in (or rotational) convergence and translational image shift.

Convergence can shift the range of portrayed depth to improve visual comfort and composition. Looking at objects presented stereoscopically further or nearer than the screen causes a disruption of the normal synergy between vergence and accommodation in most displays. Normally accommodation and vergence covary but, in a stereoscopic display, the eyes should remain focused at the screen regardless of disparity. The accommodation-vergence conflict can cause visual stress and disrupt binocular vision.[5] Convergence of the cameras can be used to adjust the disparity of targets of interest to bring them nearer to the screen and reduce this conflict.

Received Dec. 5, 2006; accepted for publication Mar. 7, 2007.
1062-3701/2007/51(4)/317/11/$20.00.

Table I. Typical convergence for stereoscopic sensors and displays. "Natural" modes of convergence are shown in bold.

DISPLAY/SENSOR GEOMETRY	REAL OR VIRTUAL CAMERA CONVERGENCE	
	Translation	Rotation
Flat	**Horizontal Image Translation**	**Toed-in camera, toed-in projector combination**
	Differential translation of computer graphics images	Toed-in stereoscopic camera or robot head
	Image sensor shift	
	Variable baseline camera	
Spherical	Human viewing of planar stereoscopic displays?	**Haploscope**
		Human physiological vergence

Convergence can also be used to shift the range of portrayed depth. For example, it is often preferable to portray stereoscopic imagery in the space behind rather than in front of the display. With convergence a user can shift stereoscopic imagery to appear "inside" the display and reduce interposition errors between the stereoscopic imagery and the edges of the displays.

Cameras used in stereoscopic imagers have limited field of view and convergence can be used to bring objects of interest into the binocular field of view.

Finally, convergence or more appropriately translation of the stereoscopic cameras can also be used to adjust for differences in a user's interpupillary distance. The latter transformation is not typically called convergence since the stereoscopic baseline is not maintained.

In choosing a method of convergence there are several issues one needs to consider. What type of 2D image transformation is most natural for the imaging geometry? Can a 3D movement of the imaging device accomplish this transformation? In a system consisting of separate acquisition and display systems is convergence best achieved by changing the imaging configuration and/or by transforming the images (or projector configuration) prior to display? If an unnatural convergence technique must be used, what is the impact on stereoscopic depth perception?

Although camera convergence is acknowledged as a useful function, there has been considerable debate over the correct transformation required. Since the eyes (and the cameras in imaging applications) are separated laterally, convergence needs to be an opposite horizontal shift of left and right eyes images on the sensor surface or, equivalently, on the display. The most appropriate type of transformation to accomplish this 2D shift—rotation or translation—depends on the geometry of the imaging and display devices. We agree with the view that the transformation should reflect the geometry of the display and imaging devices in order to minimize distortion (see Table I). One could argue that a "pure" vergence movement should affect the disparity of all objects equally, resulting in a change in mean disparity over the entire image without any change in relative disparity between points.

For example, consider a spherical imaging device such as the human eye where expressing disparity in terms of visual angle is a natural coding scheme. A rotational movement about the optical centre of the eye would scan an image over the retina without distorting the angular relationships within the image. Thus the natural convergence movement with such an imaging device is a differential rotation of the two eyes, as occurs in physiological convergence (although freedom to choose various spherical coordinate systems complicates the definition of disparity[6]).

A flat sensor is the limiting form of spherical sensor with an infinite radius of curvature, and thus the rotation of the sensor becomes a translation parallel to the sensor plane. For displays that rely on projection onto a single flat, frontoparallel display surface (many stereoscopic displays with the notable exception of some head-mounted displays and haploscopic systems) depth differences should be represented as linear horizontal disparities in the image plane. The natural convergence movement is a differential horizontal shift of the images in the plane of the display. Acquisition systems with parallel cameras are well-matched to such display geometry since a translation on the display corresponds to a translation in the sensor plane. This model of parallel cameras is typically used for the virtual cameras in stereoscopic computer graphics[7] and the real cameras in many stereoscopic camera setups.

Thus horizontal image translation of the images on the display is the preferred minimal distortion method to shift convergence in a stereoscopic rig with parallel cameras when presented on a parallel display. This analysis corresponds to current conventional wisdom. If the stereo baseline is to be maintained then this vergence movement is a horizontal translation of the images obtained from the parallel cameras rather than a translation of the cameras themselves. For example, in computer-generated displays, the left and right half images can be shifted in opposite directions on the display surface to shift portrayed depth with respect to the screen. With real camera images, a problem with shifting the displayed images to accomplish convergence is that in doing so, part of each half-image is shifted off of the display resulting in a smaller stereoscopic image.

An alternative is to shift the imaging device (e.g., CCD array) behind the camera lens, with opposite sign of shift in the two cameras forming the stereo rig. This avoids some of the problems associated with rotational convergence discussed below. Implementing a large, variable range of convergence with mechanical movements or selection of subarrays from a large CCD can be complicated. Furthermore, many lenses have significant radial distortion and translating the center of the imaging device away from the optical axis increases the amount of radial distortion. Worse, for matched lenses the distortions introduced in each sensor image will be opposite if the sensors are shifted in opposite directions. This leads to increased disparity distortion. Toed-in cameras can center the image on the optical axis and reduce this particular problem.

If we converge nearer than infinity using horizontal im-

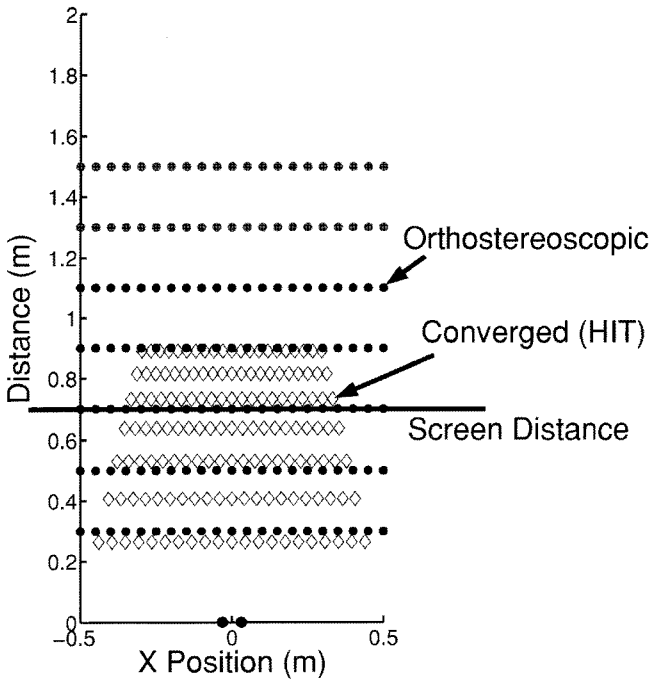

Figure 1. A plan view of an array of points located in the *X-Z* plane at eye level. The solid dots show the true position of the points and also their reconstruction based on images from a parallel camera orthostereoscopic rig presented at a 0.7 m viewing distance. The open diamond shaped markers show the reconstructed position of the points in the array when the cameras are converged using horizontal image translation (HIT). As predicted the points that are truly at 1.1 m move in to appear near the screen distance of 0.7 m. Also depth and size should appear scaled appropriately for the nearer distance. But notice that depth ordering and planarity are maintained. Circles at a distance of zero denote the positions of the eyes.

age shift, then far objects should be brought toward the plane of the screen. With convergence via horizontal image shift, a frontal plane at the camera convergence distance should appear flat and at the screen distance. However, depth for a given retinal disparity increases approximately with the square of distance. Thus if the cameras are converged at a distance other than the screen distance to bring a farther (or nearer) target toward the screen, then the depth in the scene should be distorted nonlinearly but depth ordering and planarity are maintained (Figure 1). This apparent depth distortion is predicted for both the parallel and toed-in configurations. In the toed-in case it would be added to the curvature effects discussed below. Similar arguments can be made for size distortions in the image (or equivalently the apparent spacing of the dots in Fig. 1). See Woods[1] and Diner and Fender[2] for an extended discussion of these distortions.

It is important to note that these effects are predicted from the geometry and do not always correspond to human perception. Percepts of stereoscopic space tend to deviate from the geometric predictions based on the Keplerian projections and Euclidean geometry[6]). Vergence on its own is not a strong cue to distance and other depth cues in the display besides horizontal disparity can affect the interpretation of stereoscopic displays. For example, it has been known for over 100 years that observers can use vertical disparities in the stereoscopic images to obtain more veritical estimates of stereoscopic form.[8] In recent years, a role for vertical disparities in human stereoscopic depth perception has been confirmed.[9,10]

Translation of the images on the display or of the sensors behind the lenses maintains the stereoscopic camera baseline and hence the relative disparities in the acquired or simulated image. Shifting of the images can be used to shift this disparity range to be centered on the display to ease viewing comfort. However, in many applications this disparity range is excessive and other techniques may be more suitable. Laterally shifting the cameras toward or away from each other increases or decreases the range of disparities corresponding to a given scene. Control of the stereo rig baseline serves a complementary function to convergence by adjusting the "gain" of stereopsis instead of simply the mean disparity. This function is often very useful for mapping a depth range to a useful or comfortable disparity range in applications such as computer graphics,[4,11] photogrammetry, etc.

In augmented reality or other enhanced vision systems that fuse stereoscopic imagery with direct views of the world (or with displays from other stereoscopic image sources), orthostereoscopic configurations (or at least consistent views) are important. In these systems, proper convergence of the camera systems and calibration of image geometry is required so that objects in the display have appropriate disparity relative to their real world counterparts. A parallel camera orthostereoscopic configuration presents true disparities to the user if presented on a parallel display. Thus, geometrically at least, we should expect to see true depth. In practice this seldom occurs because of the influence of other depth cues (accommodation-vergence conflict, changes in effective interpupillary distance with eye movements, flatness cues corresponding to viewing a flat display, etc.).

In summary, an orthostereoscopic parallel-camera/parallel-display configuration can present accurate disparities to the user.[1,7] On parallel displays, convergence by horizontal shift of the images obtained from parallel cameras introduces no distortion of horizontal or vertical screen disparity (parallax). Essentially, convergence by this method brings the two half images into register with out changing relative disparity. This can reduce vergence-accommodation conflict and improve the ability to fuse the imagery. Geometrically, one would predict effects on perceived depth—the apparent depth of imagery with respect to the screen and the depth scaling in the image are affected by the simulated vergence.[1,13] However, this amounts to a relief transformation implying that depth ordering and coplanarity should be maintained.[2,10]

CAMERA TOE-IN

While horizontal image translation is attractive theoretically, there are often practical considerations that limit use of the method and make rotational convergence attractive. For example, with a limited camera field of view and a nonzero stereo baseline there exists a region of space near to the

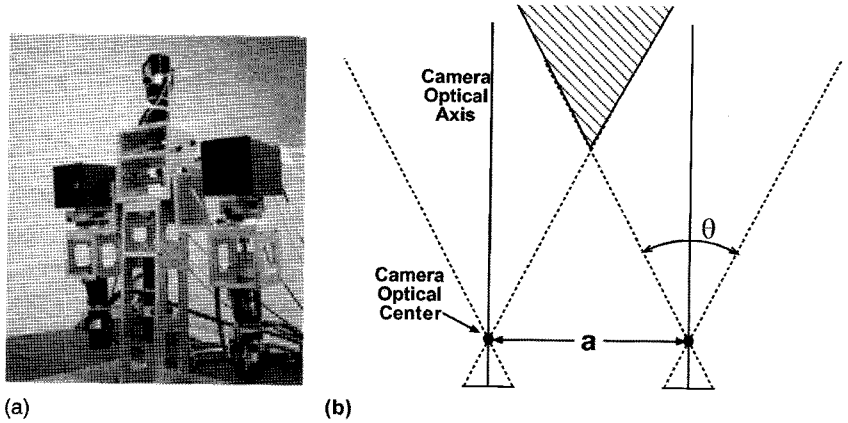

Figure 2. (a) The Toronto IRIS Stereoscopic Head 2 (TRISH II), an example of a robot head built for a wide range of working distances. With such a system, a wide range of camera convergence is required to bring objects of interest into view of the cameras. With off-the shelf cameras this can be most conveniently achieved with camera toe-in. (b) A hypothetical stereo rig with camera field of view θ. Objects in near working space are out of the binocular field of view which is indicated by the cross hatch pattern.

cameras that cannot be seen by one or both cameras. In some applications such as landscape photography this region of space may be irrelevant; in other applications such as augmented reality or stereoscopic robot heads this may correspond to a crucial part of the normal working range (see Figure 2). Rotational convergence of the cameras can increase the near working space of the system and center the target in the camera images.[14] Other motivations for rotational convergence include the desire to center the target on the camera optics (e.g., to minimize camera distortion) and the relative simplicity and large range of motion possible with rotational mechanisms. Given that rotational convergence of stereo cameras is often implemented in practice, we ask what effects the distortions produced by these movements have on the perception of stereoscopic displays?

It is well known that the toed-in configuration distorts the images in the two cameras producing patterns of horizontal and vertical screen disparities (parallax). Geometrically, deviations from the parallel-camera configuration may result in spatial distortion unless compensating transformations are introduced mechanically, optically or electronically in the displayed images,[2,12] for example unless a pair of projectors (or HMD with separate left and right displays) with matched convergence or a parallel display with special distortion correction techniques are used.[15,16] For the rest of this paper we will assume a single projector or display system (parallel display) and a dual sensor system with parallel or toed-in cameras.

The effects of the horizontal disparities have been well described in the literature and we review them before turning to the vertical disparities in the next section. The depth distortions due to the horizontal disparities introduced can be estimated geometrically.[1] The geometry of the situation is illustrated in Figure 3. The imaging space world coordinate system is centered between the cameras, a is the intercamera distance and the angle of convergence is β (using the conventional stereoscopic camera measure of convergence rather than the physiological one).

Let us assume the cameras converge symmetrically at point C located at distance F. A local coordinate system is attached to each camera and rotated $\pm\beta$ about the y axis with respect to the imaging space world coordinate system. The coordinates of a point $P=[XYZ]^T$ in the left and right cameras is

$$\begin{bmatrix} X_l \\ Y_l \\ Z_l \end{bmatrix} = \begin{bmatrix} \left(X+\frac{a}{2}\right)\cos(\beta) - Z\sin(\beta) \\ Y \\ Z\cos(\beta) + \left(X+\frac{a}{2}\right)\sin(\beta) \end{bmatrix},$$
$$\begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix} = \begin{bmatrix} \left(X-\frac{a}{2}\right)\cos(\beta) + Z\sin(\beta) \\ Y \\ Z\cos(\beta) - \left(X-\frac{a}{2}\right)\sin(\beta) \end{bmatrix}. \quad (1)$$

After perspective projection onto the converged CCD array (coordinate frame u-v centered on the optic axis and letting $f=1.0$) we get the following image coordinates for the left, $[u_l, v_l]^T$, and right, $[u_r, v_r]^T$, arrays:

$$\begin{bmatrix} u_l \\ v_l \end{bmatrix} = \begin{bmatrix} X_l/Z_l \\ Y_l/Z_l \end{bmatrix} = \begin{bmatrix} \dfrac{\left(X+\frac{a}{2}\right)\cos(\beta) - Z\sin(\beta)}{Z\cos(\beta) + \left(X+\frac{a}{2}\right)\sin(\beta)} \\ \dfrac{Y}{Z\cos(\beta) + \left(X+\frac{a}{2}\right)\sin(\beta)} \end{bmatrix}, \quad (2)$$

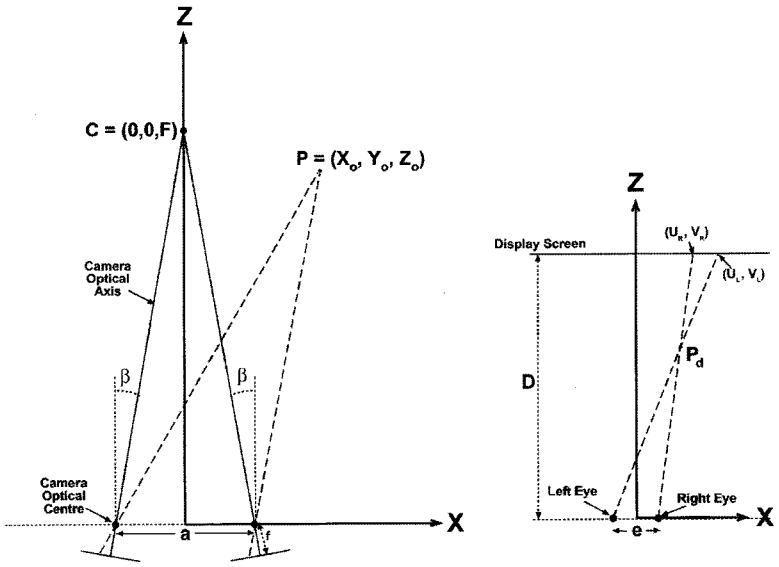

Figure 3. Imaging and display geometry for symmetric toe-in convergence at point C and viewing at distance D (plan view).

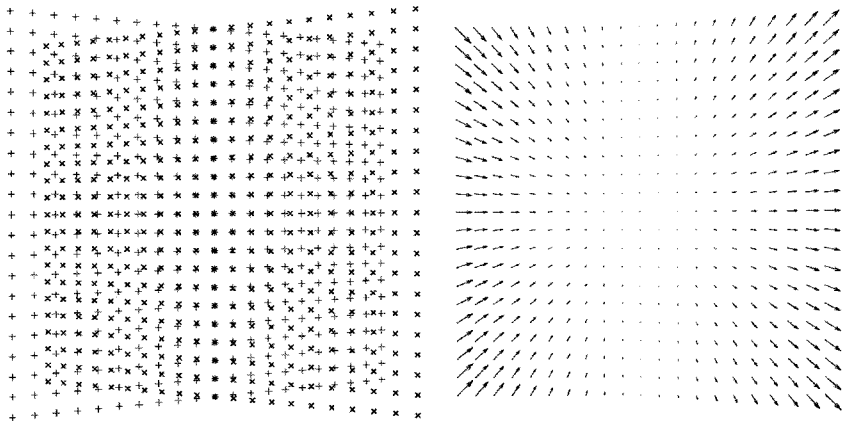

Figure 4. Keystone distortion due to toe-in. (a) Left (+) and right (×) images for a regularly spaced grid of points with the stereo camera converged (toed-in) on the grid. (b) Corresponding disparity vectors comparing left eye with right eye views demonstrate both horizontal and vertical components of the keystone distortion.

$$\begin{bmatrix} u_r \\ v_r \end{bmatrix} = \begin{bmatrix} X_r/Z_r \\ Y_r/Z_r \end{bmatrix} = \begin{bmatrix} \dfrac{\left(X-\dfrac{a}{2}\right)\cos(\beta)+Z\sin(\beta)}{Z\cos(\beta)-\left(X-\dfrac{a}{2}\right)\sin(\beta)} \\ \dfrac{Y}{Z\cos(\beta)-\left(X-\dfrac{a}{2}\right)\sin(\beta)} \end{bmatrix}.$$

The CCD image is then reprojected onto the display screen. We assume a single display/projector model with central projection and a magnification of M with respect to the CCD sensor image resulting in the following screen coordinates for the point in the left, $[U_l, V_l]^T$, and right, $[U_r, V_r]^T$, eye images:

$$\begin{bmatrix} U_l \\ V_l \end{bmatrix} = M\begin{bmatrix} u_l \\ v_l \end{bmatrix}, \quad \begin{bmatrix} U_r \\ V_r \end{bmatrix} = M\begin{bmatrix} u_r \\ v_r \end{bmatrix}. \tag{3}$$

Toeing-in the stereoscopic rig to converge on a surface centers the images of the target in the two cameras but also introduces a keystone distortion due to the differential perspective (Figure 4). In contrast convergence by shifting the CCD sensor behind the camera lens (or shifting the half images on the display) changes the mean horizontal disparity but does not entail keystone distortion. For a given focal length and camera separation, the extent of the keystone distortion is a function of the convergence distance and not the distance of the target.

To see how the keystoning affects depth perception, assume the images are projected onto a screen at distance D and viewed by a viewer with interocular distance of e. If the magnification from the CCD sensor array to screen image is

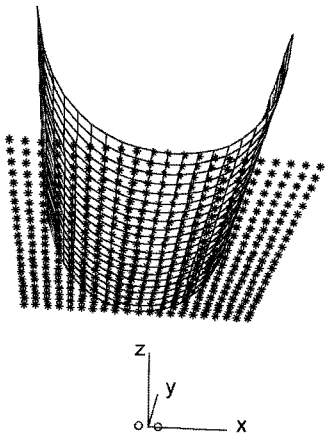

Figure 5. Geometrically predicted perception (curved grid) of displayed images taken from a toed-in stereoscopic camera rig converged on a fronto-parallel grid made with 10 cm spacing (asterisks) based on horizontal disparities (associated size distortion not shown). Camera convergence distance (F) and display viewing distance (D) are 0.70 cm ($e=a=62.5$ mm; $f=6.5$ mm; see Fig. 3 and text for definitions). The icon at the bottom of the figure indicates the position of the world coordinate frame and the eyeballs.

M and both images are centered on the display then geometrically predicted coordinates of the point in display space is (after Ref. 1)

$$P_d = \begin{bmatrix} X_d \\ Y_d \\ Z_d \end{bmatrix} = \begin{bmatrix} \dfrac{e(U_l + U_r)}{2[e - (U_r - U_l)]} \\ \dfrac{e(V_l + V_r)}{2[e - (U_r - U_l)]} \\ \dfrac{eD}{e - (U_r - U_l)} \end{bmatrix} \quad (4)$$

where $(U_r - U_l)$ is the horizontal screen parallax of the point.

If we ignore vertical disparities for the moment, converging the camera causes changes in the *geometrically* predicted depth. For instance, if the cameras toe-in to converge on a frontoparallel surface (parallel to the stereobaseline), then from geometric considerations the center of the object should appear at the screen distance but the surface should appear curved (Figure 5). This curvature should be especially apparent in the presence of undistorted stereoscopic reference imagery as would occur in augmented reality applications.[16] In contrast, if convergence is accomplished via horizontal image translation then a frontal plane at the camera convergence distance should appear flat and at the screen distance although depth and size will be scaled as discussed in the previous section.

USE OF VERTICAL DISPARITY IN STEREOPSIS

The pattern of vertical disparities in a stereoscopic image depends on the geometry of the stereoscopic rig. With our spherical retinas disparity is best defined in terms of visual angle. An object that is located eccentric to the median plane of the head is closer to one eye than the other (Figure 6). Hence, it subtends a larger angle at the nearer eye than at the further. The vertical size ratio (VSR) between the images of an object in the two eyes varies as a function of the object's eccentricity with respect to the head. Figure 6 also shows the variation of the vertical size ratio of the right eye image to the left eye image for a range of eccentricities and distances.

It is evident that, for centrally located targets, the gradient of vertical size ratios varies with distance of the surface from the head. This is relatively independent of the vergence state of the eyes and the local depth structure.[17] Howard[18] turned this relationship around and suggested that people could judge the distance of surfaces from the gradient of the VSR. Gillam and Lawergren[19] proposed a computational model for the recovery of surface distance and eccentricity based upon processing of VSR and VSR gradients. An alternative computational framework[10,20] uses vertical disparities to calculate the convergence posture and gaze eccentricity of the eyes rather than the distance and eccentricity of a target surface. For our purposes, these models make the same predictions about the effects of camera toe-in. However, the latter model uses projections onto flat projection surfaces (hypothetical flat retinae) which is easier for visualization and matches well with our previous discussion of camera toe-in.

With flat imaging planes, disparities are usually measured in terms of linear displacement in the image plane. If the cameras in a stereoscopic rig are toed in (or if eyes with flat retinae are converged), then the left and right camera images have opposite keystone distortion. It is interesting to note that in contrast to the angular disparity case the gradients of vertical disparities are a function of camera convergence but are affected little by the distance of the surface. These vertical disparity gradients on flat cameras/retinae provide an indication of the convergence angle of the cameras and hence the distance of the fixation point.

For a pair of objects or for depth within an object, the relationship between relative depth and relative disparity is a function of distance from the observer. To an extent, the visual system is able to maintain an accurate perception of depth of an object at various distances despite disparity varying inversely with the square of the distance between the object and the observer. This "depth constancy" demonstrates an ability to account for the effects of viewing distance on stereoscopic depth. The relationship between the retinal image size of an object and its linear size in the world is also a function of distance. To the degree that vertical disparity gradients are used as an indicator of the distance of a fixated surface for three-dimensional reconstruction, toe-in produced vertical disparity gradients would be expected to indirectly affect depth and size perception. Psychophysical experiments have demonstrated that vertical disparity gradients strongly affect perception of stereoscopic shape, size and depth[9,10,21] and implicate vertical disparity processing in human size and depth constancy.

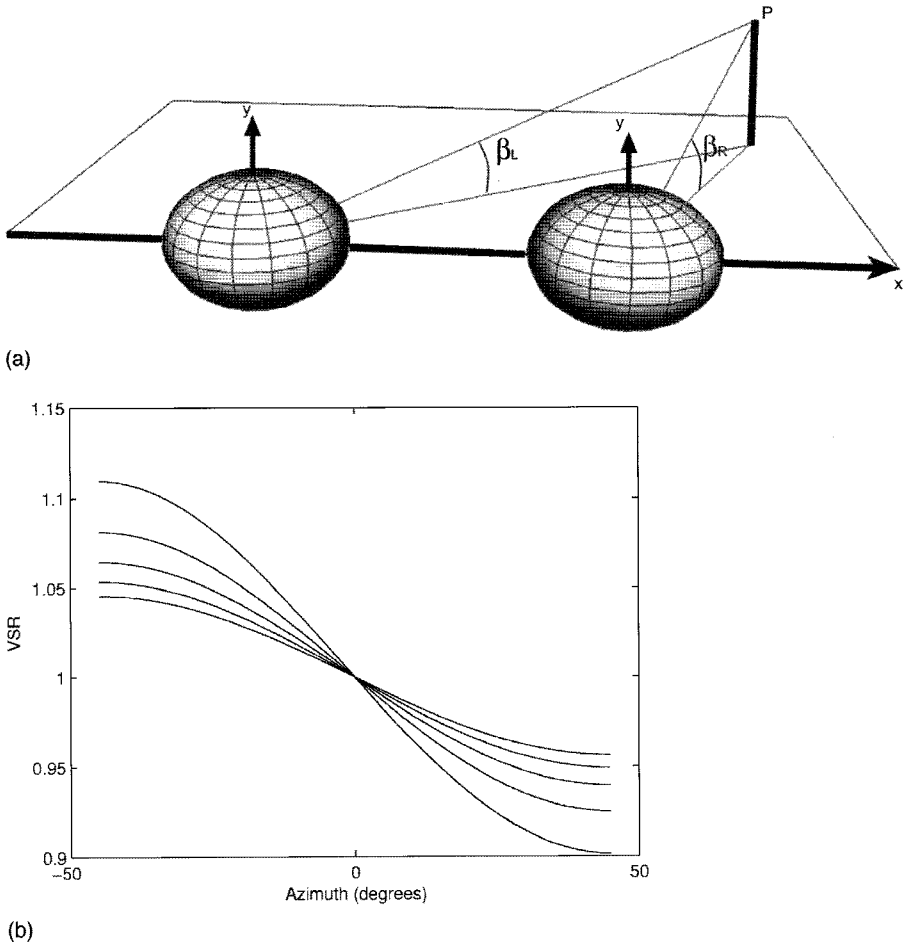

Figure 6. (a) A vertical line located eccentric to the midline of the head is nearer to one eye than the other. Thus it subtends a larger angle in the nearer eye than the further (adapted from Howard and Rogers[6]). (b) The gradient of vertical size ratio of the image of a surface element in the left eye to that in the right eye varies as a function of distance of the surface (shown as a series of lines: distances of 70, 60, 50, 40, and 30 cm in order of steepness).

VERTICAL DISPARITY IN TOED-IN STEREOSCOPIC CAMERAS

First, consider a stereoscopic camera and parallel display system that intends to portray realistic depth and that has camera separation equal to the eye separation. If the camera is converged using the toe-in method at a fronto-parallel surface at the distance of the screen, then the center of the target will have zero horizontal screen disparity. However, the camera toe-in will introduce keystone distortion into the two images with the pattern of horizontal disparities predicting curvature as discussed above. What about the pattern of vertical disparities? The pattern of vertical disparities produced by a toed-in camera configuration resembles the gradient of vertical size disparities on the retinae that can arise due to differential perspective of the two eyes. As discussed in the previous section, this differential perspective forms a natural and rich source of informative parameters contributing to human stereoscopic depth perception.

Given that camera toe-in generates such gradients of vertical disparity in stereoscopic imagery, is it beneficial to use camera toe-in to provide distance information in a stereoscopic display? In other words, should the toed-in configuration be used to converge the cameras and preserve the sense of absolute distance and size, shape and depth constancy? Perez-Bayas[22] argued that toed-in camera configurations are more natural since they present these vertical disparities. The principal problem with this claim is that it considers the screen parallax of stereoscopic images rather than their retinal disparities. These keystone distortions are in addition to the natural retinal vertical disparities present when viewing a scene at the distance of the screen.

In order to estimate the effect on depth perception we need to consider the *retinal* disparities generated by the stereoscopic image. The keystone distortion occurs in addition to the retinal vertical disparity pattern inherent in the image because it is portrayed on the flat screen. Consider a fronto-parallel surface located at the distance of the screen away from the camera and that we intend to display the surface at the screen. Projections onto spherical retinas are hard to visualize so let us consider flat retinae converged (toed-in) at the screen distance. Alternatively one could imagine another pair of converged cameras viewing the display, one centered at the center of each eye. The images on these converged flat retinae would of course have differential keystone distortion

when viewing a frontal surface such as the screen. When displaying images from the toed-in stereoscopic camera, which already have keystone distortion, the result is an exaggerated gradient of vertical disparity in the retinal images appropriate for a much nearer surface. For a spherical retina the important measure is the gradient of vertical size ratios in the image. The vertical size ratios in the displayed images imposed by the keystone distortion are in addition to the natural VSR for a frontal surface at the distance of the screen. Clearly, the additional keystone distortion indicates a nearer surface in this case as well [Figure 7(a)].

From either the flat camera or spherical retina model we predict spatial distortion if disparities are scaled according to the vertical disparities, which indicate a closer target. Such a misjudgement of perceived distance would be predicted to have effects on perceived depth and size [open circles in Fig. 7(b)]. There is little evidence that observers actually mislocalize surfaces at a nearer distance when a vertical disparity gradient is imposed. However, there is strong evidence for effects of VSR gradients on depth constancy processes.

If a viewer fixates a point on a fronto-parallel screen, then at all screen distances nearer than infinity the images of other points on the screen have horizontal disparity (retinal but not screen disparity). This is because the theoretical locus of points in three-dimensional space with zero retinal disparity, which is known as the horopter (the Vieth-Muller circle), curves inward toward the viewer and away from the frontal plane. The curvature of the horopter increases at nearer distances (Figure 8).[23] Thus a frontal plane presents a pattern of horizontal disparities that varies with distance. If depth constancy is to be maintained for fronto-parallel planes then the distance of the surface needs to be taken into account. Rogers and Bradshaw[21] showed that vertical disparity patterns can have a strong influence on frontal plane judgements, particularly for large field of view displays. Specifically, "flat"—or zero horizontal screen disparity—planes are perceived as curved if vertical disparity gradients indicate a distance other than the screen distance.

In our case, the toe-in induced vertical disparity introduces a cue that the surface is nearer than specified by the horizontal screen disparity. Thus a zero horizontal screen disparity pattern for a frontal surface at the true distance would be interpreted as at nearer distance. The disparities would be less than expected from a frontal plane at the nearer distance. As a result, surfaces in a scene should appear curved more concavely than they are in the real scene. Notice that the distortion is in the opposite direction than the distortion created by horizontal disparities due to the keystoning.

Thus the effect of vertical disparity introduced by the keystone distortion is complicated. The vertical disparity introduces a cue that the surface is nearer than specified by the horizontal screen disparity. Thus, from vertical disparities, we would expect a bias in depth perception and concave distortion of stereoscopic space. This may counter the convex distortions introduced by the horizontal disparities discussed above. So the surface may appear flatter than expected from the distorted horizontal disparities. But the percept is not more "natural" than the parallel configura-

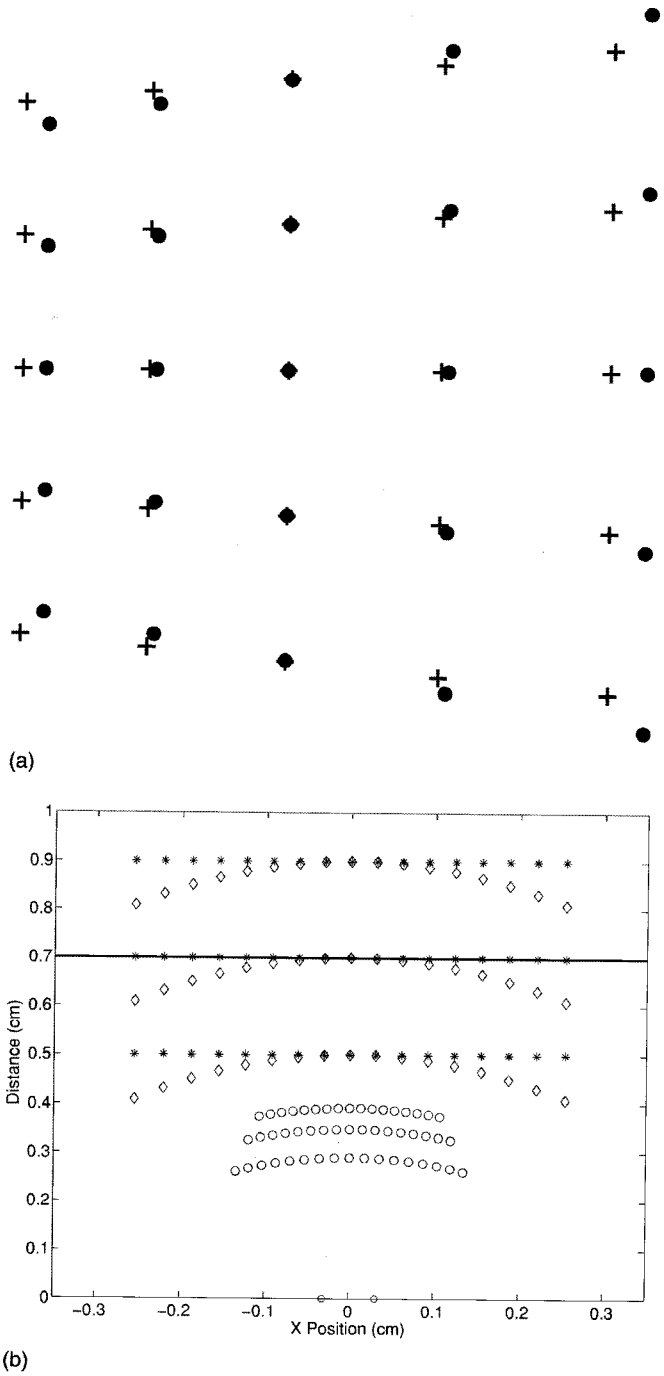

Figure 7. (a) Simulation of the keystone distortion and gradient of VSR present in a stereo half image for a toed-in configuration. The plus symbols show the keystone distortion in the displayed image of a grid for a camera converged at 70 cm and the circle symbols indicated the exaggerated VSR distortion present in the retinal half image for an observer viewing the display at 70 cm (flat retina). (b) Predicted distorted appearance (circles) in a set of frontal plane surfaces (asterisks) if depth from disparity is scaled according to the distance indicated by an exaggerated VSR. Typically the surface is not mislocalized in depth but curvature is induced. The predicted curvature based on the on the equations provided by Duke and Wilcox[28] is also shown (diamonds). The simulated positions of the eyes are denoted by circles at zero distance and the screen by a line at 70 cm.

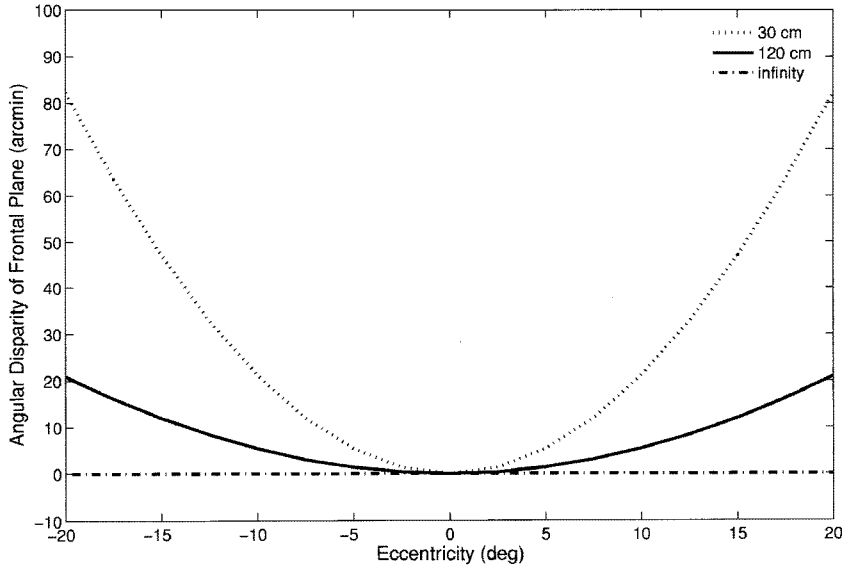

Figure 8. Disparity of a point on a fronto-parallel surface as a function of distance. Horizontal disparity for a given eccentricity increases with nearness due to the increasing curvature of the Veith-Muller circle (see text).

tion. Rather two distortions due to camera toe-in act to cancel each other out.

Do toed-in configurations provide useful distance information for objects at other distances or for nonorthostereoscopic configurations?

Since the toe-in induced vertical disparity gradients are superimposed upon the natural vertical disparity at the retinae they do not provide natural distance cues for targets near the display under orthostereoscopic configurations. Nonorthostereoscopic configurations are more common than orthostereoscopic and we should consider the effects of toe-in on these configurations. Magnification and minification of the images will scale the disparities in the images as well so that the vertical gradient of vertical size ratio will be relatively unchanged under uniform magnification. Hence we expect a similar curvature distortion under magnification or minification.

Hyperstereoscopic and hypostereoscopic configurations exaggerate and attenuate, respectively, the horizontal and vertical disparities due to camera toe-in and the magnitude of the stereoscopic distortions will be scaled. However, for both configurations the sign of the distortion is the same and vertical disparities from camera toe-in predict concave curvature of stereoscopic space with increased distortion with an increased stereobaseline.

For surfaces outside the plane of the screen, vertical keystone distortion from toe-in still introduces spatial distortion. A surface located at a distance beyond the screen in a parallel camera, orthostereoscopic configuration will have VSR gradients on spherical retinae appropriate to its distance due to the imaging geometry. For a toed-in camera system, all surfaces in the scene will have additional vertical disparity gradients due to the keystoning. These increased vertical disparity gradients would indicate a nearer convergence distance or a nearer surface thus the distance of the far surface should be underestimated and concave curvature introduced. The distance underestimation would be compounded by rescaling of disparity for the near distance which should compress the depth range in the scene.

What about partial toe-in? For example, let us say we toed in on a target at 3 m and displayed it at 1.0 m with the centers of the image aligned? Would the vertical disparities in the image indicate a more distant surface, perhaps even one at 3 m (this would be the case if viewed in a haploscope)? A look at the pattern of vertical screen disparities in this case, however, shows that they are appropriate for a surface that is nearer than the 3 m surface, and in fact nearer than the screen if the half images are aligned on the screen. Thus when the vertical screen disparities are compounded by the inherent vertical retinal disparities introduced by viewing the screen, the toe-in induced distortion actually indicates a nearer surface rather than the further surface desired. We will see below that vertical disparity manipulations can produce the impression of a further surface but the required transformation is opposite to the one introduced by camera toe-in.

Do the toed-in configurations improve depth and size scaling?

Vertical disparities have been shown to be effective in the scaling of depth, shape and size from disparity.[9,21] When the cameras are toed-in the vertical disparities indicate a nearer surface. Therefore, camera toe-in should cause micropsia (or apparent shrinking of linear size) appropriate for the nearer distance. Similarly, depth from disparity should be scaled appropriate to a nearer surface and depth range should be compressed. Thus, if toe-in is used to converge an otherwise orthostereoscopic rig, then image size and depth should be compressed. Vertical disparity cues to distance are most effective in a large field of view display and the curvature, size and depth effects are most pronounced in these types of displays.[9,21]

In the orthostereoscopic case with parallel cameras,

there are no vertical screen disparities and the vertical disparities in the retinal images are appropriate for the screen distance and no size or depth distortions due to vertical disparity are predicted. Vertical disparities in the retinal (but not display) images can thus help obtain veridical stereoscopic perception.

I use computer graphics or image processing to render stereoscopic images. Can I use VSR to give an impression of different distances? If so how?

Incorporating elements that carry vertical disparity information (for example with horizontal edges) can lead to more veridical depth perception[8] and in this simple sense vertical disparity cues can assist in the development of effective stereoscopic displays. It is not certain that manipulating vertical disparity independent of vergence would be of use to content creators, but it is possible. In the lab we do this to look at the effects of vertical disparity gradients and to manipulate the effects of vertical disparities with vergence held constant.

We have seen that toe-in convergence introduces a vertical disparity cue that indicates that a surface is nearer than other cues indicate. This will scale stereoscopic depth, shape and size appropriately, particularly for large displays. To make the surface appear further away the opposite transformation is required to reduce the vertical disparity gradients in the retinal image—this essentially entails "toe-out" of the cameras. VSR manipulations, intentional or due to camera toe-in, exacerbate cue conflict in the display as the distance estimate obtained from the vertical disparities will conflict with accommodation, vergence, and other cues to distance.

FUSION OF VERTICAL DISPARITY

In many treatments of the camera convergence problem it is noted that the vertical disparities introduced by toed-in camera convergence may interfere with the ability to fuse the images and cause visual discomfort.[24] Certainly, vertical fusional range is known to be less than horizontal fusional range[23] making it likely that vertical disparities could be problematic. Tolerance to vertical disparities depends on several factors including size of the display, and the presence of reference surfaces.

When a stereoscopic image pair has an overall vertical misalignment, such as arises with vertical camera misalignment, viewers can compensate with vertical vergence and sensory fusional mechanisms. Vertical vergence is a disjunctive eye movement where the left and right eyes move in opposite directions vertically (vertical misalignment can also often be partially compensated by tilting the head with respect to the display). Vertical disparities are integrated over a fairly large region of space to form the stimulus to vertical vergence.[25] Larger displays increase the vertical vergence response and the vertical fusional range. Thus we predict that vertical disparities will be better tolerated in large displays. In agreement with this Speranza and Wilcox[26] found up to 30 minutes of arc of vertical disparity could be tolerated in a stereoscopic IMAX™ film without significant viewer discomfort. However, convergence via camera toe-in gives local variations in vertical disparity and thus images of objects in the display have spatially varying vertical disparities. Thus, averaging retinal vertical disparities over a region of space should be less effective in compensating for vertical disparity due to camera toe-in compared to overall vertical camera misalignment. Furthermore, any vertical vergence to fuse one portion of the display will increase vertical disparity in other parts of the display.

The ability to fuse a vertically disparate image is reduced when nearby stimuli have different vertical disparities, particularly if the target and background are similar in depth.[27] In many display applications the frame of the display is visible and serves as a frame of reference. In other applications such as augmented reality and enhanced vision displays the stereoscopic imagery may be imposed upon other imagery. Presence of these competing stereoscopic images will be expected to reduce the tolerance to vertical disparity due to camera convergence.[27] This indicates that vertical disparity distortions should be particularly disruptive in augmented reality displays where the stereoscopic image is superimposed on other real or synthetic imagery and parallel cameras or image rectification should be used.

ADAPTATION AND SENSORY INTEGRATION OF TOE-IN INDUCED VERTICAL DISPARITY

The human visual system relies on a variety of monocular and binocular cues to judge distance and relative depth in a scene. The effects of toe-in induced horizontal and vertical disparities on depth and distance perception discussed above will be reduced when viewing a scene rich in these cues. The extent of the perceptual distortion depends on perceptual biases and the relative effectiveness of the various cues. For example, Bradshaw and Rogers[21] performed an experiment using dot displays to study size and depth scaling as a function of distance indicated by vertical disparities and vergence. They argued that use of vertical disparity information to drive size and depth constancy requires measuring the relevant disparity gradients over a fairly large retinal area whereas vergence signals, correlated with egocentric distance, could be obtained during binocular viewing of a point source of light. Accordingly, when displays were small, subjects responded as if they were scaling the stimulus appropriate for the distance indicated by vergence; when displays were large subjects responded as if they were scaling the stimulus appropriate for the distance indicated by vertical disparity. When other cues reliably indicate a different distance than toe-in induced vertical disparities the effect of the latter on depth and size perception may be small. However, latent, even imperceptible, cue conflicts are believed to be a causal factor in simulator sickness symptoms such as eye strain and nausea.[5]

When sensory conflict is persistent, the visual system shows remarkable ability to adapt or recalibrate. Following prolonged viewing of a test stimulus that appears curved due to keystone-type vertical disparity transformations a nominally flat stimulus appears curved in the opposite direction. Duke and Wilcox[28] have claimed this adaptation is driven by

the curvature in depth induced rather than by the vertical disparities directly. In general, such an aftereffect can reflect "habituation" or "fatigue" of mechanisms sensitive to the adapting pattern, or from a recalibration of the vertical disparity signal, or a change in the relative weighting of cues driving depth constancy. At the present time it is unclear which of these adaptive changes can be produced by prolonged exposure to keystone patterns of vertical disparity.

The effects of vertical disparities induced by toe-in convergence also depends on context and may differ depending on the type of task being performed by the subject and whether they involve size constancy, depth constancy, absolute distance judgements or other spatial judgements. For example, Wei et al.[29] reported that full-field vertical disparities are not used to derive the distance dependent gain term for the linear vestibulo-ocular reflex, a reflexive eye movement that compensates for head movements, under conditions where vertical disparities drive depth constancy.

CONCLUSIONS

In conclusion, we concur with conventional wisdom that horizontal image translation is theoretically preferred to toe-in convergence with parallel stereoscopic displays. Toed-in camera convergence is a convenient and often used technique that is often well-tolerated[24] despite the fact that it theoretically and empirically results in geometric distortion of stereoscopic space. The distortion of stereoscopic space should be more apparent in fused or augmented reality displays where the real world serves as a reference to judge the disparity distortion introduced by the toe-in technique. In these cases, and for near viewing when the distortions are large, the distortions may be ameliorated through camera rectification techniques[15,30] if resampling of the images is practical.

It has been asserted by others that, since camera convergence through toe-in introduces vertical disparities into the stereoscopic imagery it should give rise to more natural or accurate distance perception than the parallel camera configuration. We have argued in this paper that these assertions are theoretically unfounded although vertical disparity gradients are an effective cue for depth and size constancy that could be used by creators of stereoscopic content. The geometrical distortions predicted from the artifactual horizontal disparities created by camera toe-in may be countered by opposite distortions created from the vertical disparities. However, when displayed on a single projector or monitor display the vertical disparity gradients introduced by unrectified, toed-in cameras do not correspond to the gradients experienced by a real user viewing a scene at the camera convergence distance. This is because the keystoning due to the camera toe-in is superimposed upon the natural vertical disparity pattern *at the eyes*.

Our analysis and data[27] implies that stereoscopic display/camera systems that fuse or superimpose multiple stereoscopic images from a number of sensors should be more susceptible to toe-in induced fusion and depth-distortion problems than displays that present a single stereoscopic image stream. Analysis of toe-in induced vertical disparity reinforces the recommendation that rectification of the stereoscopic imagery should be considered for fused stereoscopic systems such as augmented reality displays or enhanced vision systems that require toed-in cameras to view targets at short distances.

ACKNOWLEDGMENTS

The support of the Ontario Centres of Excellence and NSERC Canada is gratefully acknowledged. An abbreviated version of this paper was presented at IST/SPIE Electronic Imaging 2004 [R. Allison, Proc. SPIE **5291**, 167–178 (2004)].

REFERENCES

[1] A. Woods, T. Docherty, and R. Koch, Proc. SPIE **1915**, 36 (1993).
[2] D. B. Diner and D. H. Fender, *Human Engineering in Stereoscopic Viewing Devices* (Plenum Press, New York and London, 1993).
[3] L. Lipton, *Foundations of the Stereoscopic Cinema* (Van Nostrand–Reinhold, New York, 1982).
[4] Z. Wartell, L. F. Hodges, and W. Ribarsky, IEEE Trans. Vis. Comput. Graph. **8**(2), 129 (2002).
[5] J. P. Wann, S. Rushton, and M. Monwilliams, Vision Res. **35**(19), 2731 (1995).
[6] I. P. Howard and B. J. Rogers, *Depth Perception* (I. Porteous, Toronto, 2002).
[7] L. Lipton, *The Stereographics Developers Handbook* (Stereographics Corp., San Rafael, CA, 1997).
[8] H. von Helmholtz, *Physiological Optics*, English translation by J. P. C. Southall from the 3rd German edition of *Handbuch der Physiologischen Optik*, Vos, Hamburg (Dover, New York, 1962).
[9] B. J. Rogers and M. F. Bradshaw, Nature (London) **361**, 253 (1993).
[10] J. Garding, J. Porrill, J. E. W. Mayhew, and J. P. Frisby, Vision Res. **35**(5), 703 (1995).
[11] M. Siegel and S. Nagata, IEEE Trans. Circuits Syst. Video Technol. **10**(3), 387 (2000).
[12] A. State, J. Ackerman, G. Hirota, J. Lee, and H. Fuchs, *Proc. International Symposium on Augmented Reality (ISAR) 2001* (IEEE, Piscataway, NJ, 2001) pp. 137–146.
[13] V. S. Grinberg, G. Podnar, and M. W. Siegel, Proc. SPIE **2177**, 56 (1994).
[14] A. State, K. Keller, and H. Fuchs, *Proc. International Symposium on Mixed and Augmented Reality (ISMAR) 2005* (IEEE, Piscataway, NJ, 2005) pp. 28–31.
[15] N. Dodgson, Proc. SPIE **3295**, 100 (1998).
[16] S. Takagi, S. Yamazaki, Y. Saito, and N. Taniguchi, *Proc IEEE & ACM ISAR 2000* (IEEE, Piscataway, NJ, 2000) pp. 68–77.
[17] B. Gillam and B. Lawergren, Percept. Psychophys. **34**(2), 121 (1983).
[18] I. P. Howard, Psychonomic Monograph Supplements **3**, 201 (1970).
[19] B. Gillam, D. Chambers, and B. Lawergren, Percept. Psychophys. **44**, 473 (1988).
[20] J. E. W. Mayhew and H. C. Longuet-Higgins, Nature (London) **297**, 376 (1982).
[21] M. F. Bradshaw, A. Glennerster, and B. J. Rogers, Vision Res. **36**(9), 1255 (1996).
[22] L. Perez-Bayas, Proc. SPIE **4297**, 251 (2001).
[23] K. N. Ogle, *Researches in Binocular Vision* (Hafner, New York, 1964).
[24] L. B. Stelmach, W. J. Tam, F. Speranza, R. Renaud, and T. Martin, Proc. SPIE **5006**, 269 (2003).
[25] I. P. Howard, X. Fang, R. S. Allison, and J. E. Zacher, Exp. Brain Res. **130**(2), 124 (2000).
[26] F. Speranza and L. Wilcox, Proc. SPIE **4660**, 18 (2002).
[27] R. S. Allison, I. P. Howard, and X. Fang, Vision Res. **40**(21), 2985 (2000).
[28] P. A. Duke and L. M. Wilcox, Vision Res. **43**(2), 135 (2003).
[29] M. Wei, G. C. DeAngelis, and D. E. Angelaki, J. Neurosci. **23**, 8340 (2003).
[30] O. Faugeras and Q. Luong, *The Geometry of Multiple Images* (MIT Press, Cambridge, MA, 2001).