3D discomfort from vertical and torsional disparities in natural images

Christopher W. Tyler[1], Lora T. Likova[1], Kalin Atanassov[2], Vikas Ramachandra[2] and Sergio Goma[2]

1. Smith-Kettlewell Eye Research Institute, San Francisco.
2. Qualcomm, Inc, San Diego

ABSTRACT

The two major aspects of camera misalignment that cause visual discomfort when viewing images on a 3D display are vertical and torsional disparities. While vertical disparities are uniform throughout the image, torsional rotations introduce a range of disparities that depend on the location in the image. The goal of this study was to determine the discomfort ranges for the kinds of natural image that people are likely to take with 3D cameras rather than the artificial line and dot stimuli typically used for laboratory studies. We therefore assessed visual discomfort on a five-point scale from 'none' to 'severe' for artificial misalignment disparities applied to a set of full-resolution images of indoor scenes.

For viewing times of 2 s, discomfort ratings for **vertical disparity** in both 2D and 3D images rose rapidly toward the discomfort level of 4 ('severe') by about 60 arcmin of vertical disparity. Discomfort ratings for **torsional disparity** in the same image rose only gradually, reaching only the discomfort level of 3 ('strong') by about 50 deg of torsional disparity. These data were modeled with a second-order hyperbolic compression function incorporating a term for the basic discomfort of the 3D display in the absence of any misalignments through a Minkowski norm. These fits showed that, at a criterion discomfort level of 2 ('moderate'), acceptable levels of vertical disparity were about 15 arcmin. The corresponding values for the torsional disparity were about 30 deg of relative orientation.

Keywords: 3D discomfort Vertical disparity Torsional disparity 3D display Stereoscopic Misalignment

INTRODUCTION

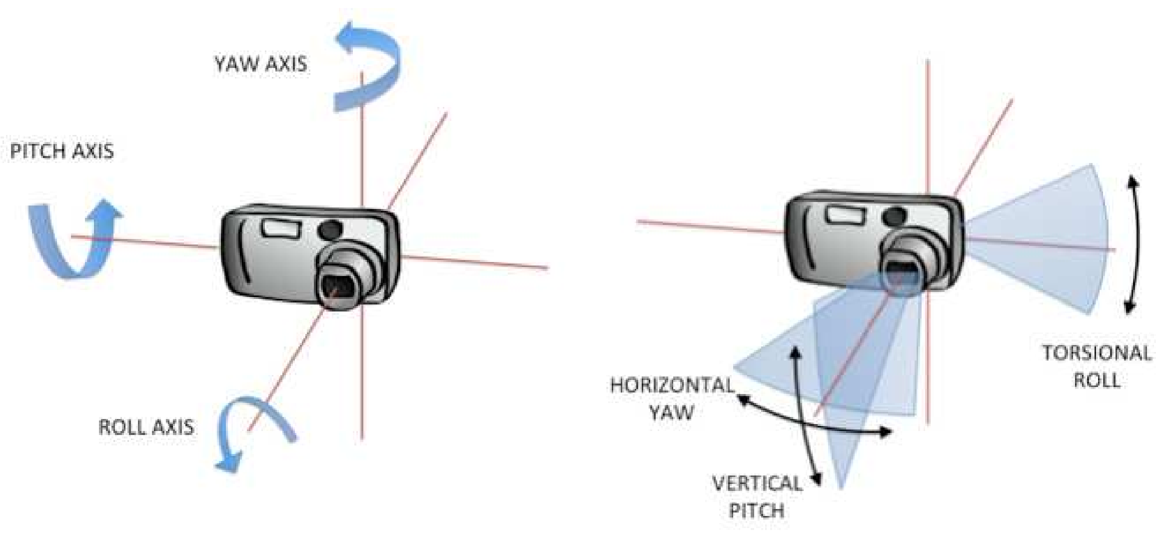
In the present era of expansion in 3D display technology, it is critical to have a good handle on the factors that cause discomfort in the viewing of 3D displays, particularly for the kinds of images that consumers are likely to shoot with their 3D devices. Overall, the issue of 3D image quality may be seen as encompassing four different categories that may cause viewing discomfort:

- **Camera misalignment issues** (translational disparities, torsional disparity, mean tilt, toe-in and other differential shear factors)
- **Optical distortions** (lens aberrations, lens displacement, relative focus)
- **Disparity registration issues** (pixellation, depth of focus, temporal synchronization)
- **Post-processing distortions** introduced during attempted compensation for camera misalignments

An important factor in the manufacture of devices for 3D image acquisition is the accuracy of alignment between the two cameras required to capture the respective images, which are subject to a variety of manufacturing inaccuracies. The precision of camera misalignment required to avoid visual discomfort when the resulting images are viewed in 3D is therefore the focus present report. Among the camera misalignment issues, there are diverse impacts on 3D image quality:

- **Horizontal misalignments**
- **Vertical misalignments**
- **Torsional misalignments**
- **Size and keystoning disparity fields**

The physical misalignments derive from errors in the relative positioning of the cameras in space. These errors are best viewed as combinations of rotations on three primary axes, as depicted in Fig. 1A: the pitch and yaw axes horizontally and vertically through the picture plane, and the roll axis coinciding with the optic axis through the center of the lens. Rotational errors is aligning a camera around the pitch and yaw axes produce vertical and horizontal image shifts, respectively (Fig. 1B), while rotational errors around the roll axis produce image rotations about the optical center. In stereoscopic systems, the requirement is to align a pair of cameras with respect to these three axes, with a relative displacement along the (horizontal) pitch axis to generate a horizontal disparity between the two images. Errors in the alignment between the two cameras then generate relative displacements between the images from the two cameras that are seen by the eyes of the viewer as three kinds of ***binocular disparity***. Errors in the rotational alignment between the two cameras may thus produce a combination of the three kinds of relative image displacement depicted in Fig. 1B – horizontal disparity, vertical disparity and torsional disparity.



A B

Fig. 1A. Depiction of the three axes of camera misalignment error (the pitch, roll and yaw axes). B. Depiction of the three directions of image displacement resulting from alignment error around the three axes (horizontal, vertical and torsional displacements, respectively).

Considered in terms of the form of the disparity fields caused by the image shifts, the misalignment errors depicted in Fig. 1B cause uniform affine and projective transforms across the visual field (as opposed to the local horizontal disparities generated by the array of objects in the world when viewed by two imaging arrays (such as the two eyes), which are an inherent property of the 3D projection. Non-rotational displacements of the two cameras from the intended locations will cause uniform horizontal and vertical image disparities for the respective camera misalignments, while displacements along the optic axes will generate small scaling differences in the image size between the two eyes. For roll rotation around the optical axes of the two cameras, the differential shift between the images is a torsional, or curl, disparity field around the center of the image (with no concomitant scaling differences). For misalignments in the form of rotations around the pitch and yaw axes in the image plane, the disparity field includes higher-order projective shear components of keystoning, which is a projective (non-affine) transformation in which the local image scale varies according to the differential distance to the scene from the two cameras.

Of this variety of image displacements, the two major aspects that cause visual discomfort when viewing the output on a display are the ***vertical*** and ***torsional*** image disparities between the two eyes. While vertical disparities are uniform throughout the image, torsional differences between the eyes introduce a range of local disparities that depend on the location in the image. At the center of the image, torsional disparities scale toward zero, which may be an advantageous configuration in that the human retina has its highest resolution at the center of gaze (which is typically near the center of the image). Along the horizontal meridian, the torsional disparities consist of a gradient of vertical disparity, whose discomfort factor may be expected to derive from the vertical disparity effects, although the weighting with respect to eccentricity will depend on a variety of factors. At the vertical meridian, torsional disparities become pure horizontal disparities, which locally should not be expected to produce much discomfort, but will introduce an overall depth slant of the image that will fade toward upright with their angle away from the vertical meridian. This horizontal disparity field corresponds to a sinusoidal warping of the depth image that will be inconsistent with the structure of, say, a textured vertical wall, and the resulting cue conflict may produce noticeable visual discomfort. These are some of the reasons that torsional disparity was considered an important factor to evaluate in the present study. Although torsional differences may in principle be compensated by torsional counter-rotation of the eyes, this tendency will be counteracted by the non-rotation of the frame of the viewing screen, and is not expected to have a strong compensatory effect unless the field of view is very large.[1]

The mean of the horizontal disparities from horizontal misalignments is a factor that is readily compensated by convergence/divergence shifts of the eyes without causing discomfort for the likely range of the misalignments. The overall and differential image scalings that accompany these misalignments are generally second-order (i.e., small) relative to the main effects, and have not been included in the present study. The vertical disparity components are not so readily compensated by vertical realignments of the two eyes, and result in significant levels of diplopia and interocular rivalry between the discrepant images to the two eyes.[2] For these reasons, this report focuses on the discomfort caused by uniform vertical and torsional disparities.

EXPERIMENTAL METHODS

Image types

The goal of the study was to determine the discomfort ranges for the kinds of natural image that people are likely to take with 3D camera devices rather than the artificial line and dot stimuli that are typically used for laboratory studies. We therefore applied artificial misalignment disparities to a set of full-resolution images of indoor scenes, with depth ranges from 2m to 10m from the cameras, which had a the average human-eye separation of 6 cm.

To determine whether there are interactions with the presence of horizontal disparity in the image, we can simply replace the stereo-image pair with a pair of identical images made by replicating one of the eye's images, to form a set of flat-image (2D) pairs for both vertical and torsional disparities. If the discomfort level is entirely controlled by the artifactual vertical and torsional disparities, there should be no significant difference between the ratings for the 2D and 3D presentations with and without the operative horizontal disparities. If, however, the perceived depth from the horizontal disparities tends to mask the visibility of the artifactual disparities, the discomfort levels should be rated as lower in the 3D condition. If, on the other hand, the horizontal disparity draws attention to the features with artifactual vertical and torsional disparities, it may increase the discomfort ratings in the 3D compared with the 2D condition.

Disparity ranges

For the 100 x 55 cm display viewed at 180 cm, the display size was 32 x 17.5 deg. The image resolution was 1024 x 768 pixels and each pixel calibrates to a visual angle of 1.85 arcmin. Pretesting suggested the need to cover a range of vertical disparities up to 4 deg of visual angle (or about ¼ of the picture height). For torsional disparity, where the images are rotated about the center, the aim was to cover about the same range of vertical disparities half way between the center and the edge of the image, or about 30 deg of torsional disparity. To cover the range effectively, the disparities were incremented in logarithmic steps of doublings at the low end and approximate half-doublings at the high end in both vertical and torsional disparity. Thus, rounding up at the low end, to cover the 4 deg of vertical disparity, we used disparities of 0, 1, 2, 4, 6, 8, 12, 16, 24 and 32 pixels. For the torsional disparities, the metric is rotational angle, and the values employed were 0, 1, 2, 4, 8, 12, 16, 24, 32 and 48 deg of torsional disparity.

Procedure

The participant viewed the images in random order with a presentation time of 2 s. This time was chosen to be short to reflect the typical shot in a 3D movie, but long enough to allow and assessment of the discomfort effect. The discomfort was rated on a five-level rating scale: 0 – no discomfort, 1 – mild, 2 – moderate, 3 – strong, 4 – severe discomfort. Once a choice was made, the participants had to confirm each choice with a second keystroke in order to minimize keystroke errors or attentional blinks.

Participants

The study was run on a group of 9 participants with normal or corrected-to-normal visual acuity and normal stereoscopic vision, with a broad age range of 21-67.

RESULTS

Vertical disparity ratings

The data for the average visual discomfort level rated by nine normal subjects for the vertical disparity conditions are plotted in Fig. 2. The discomfort ratings for the 2D images (blue diamonds) start close to zero and rise rapidly toward the 'severe' discomfort level of 4 by about 60 arcmin of vertical disparity. Interestingly, the discomfort ratings for the 3D images (red squares) did not start at zero, but at close to the 'mild' discomfort level of 1 and then rose at a similar rate. This 'mild' discomfort level is produced by the interocular leakage in the manufactured display, and hence is present even at a level of zero vertical disparity (i.e., for a perfect pair of encoded stereoscopic images). The datasets had tight correlations 0.94 and 0.97 for the 2D and 3D viewing conditions, respectively (excluding the datum points for the largest disparity, which reached saturation in the discomfort rating scale).

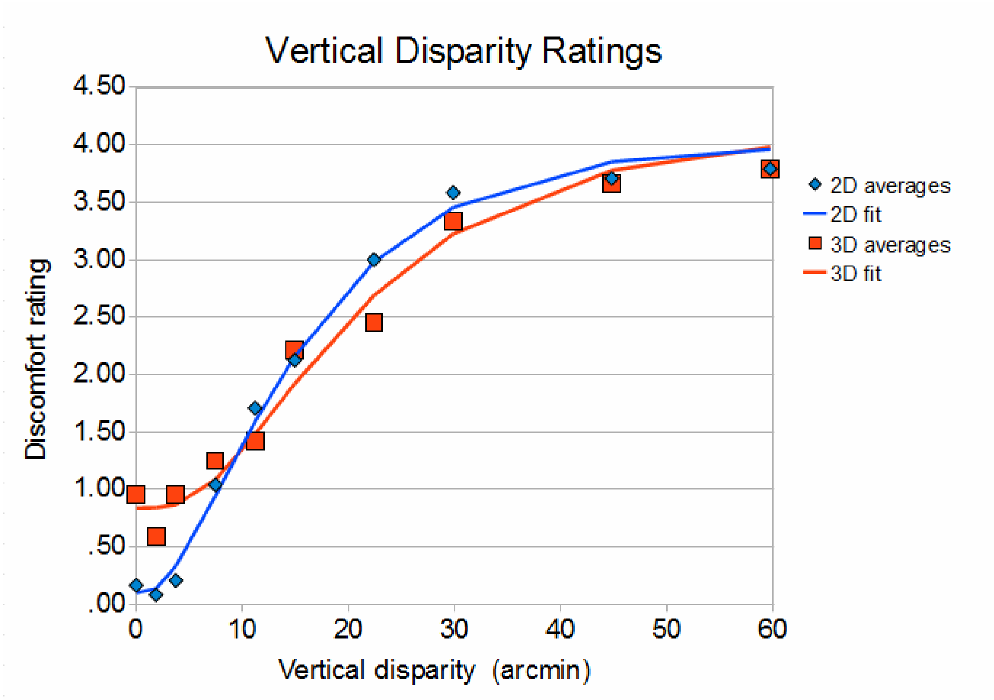


Fig. 2. Discomfort ratings for the relative vertical displacement of the images to the two eyes, generating a vertical disparity. Data are the averages for 9 subjects. Blue diamonds: 2D images; red squares: 3D images. Curves are the fits of the Disparity Energy Model of eq 1.

Model fit

The data are fit by a model discomfort function based on the disparity energy model, consisting of an energy function with a hyperbolic compression, to which is added a constant minimum discomfort level corresponding to the interocular leakage factor for the scene disparity in the 3D display. The equation for this model of the stereo discomfort rating (SDR) based on the disparity energy model, consisting of an energy function $\mathbf{E}_v$ of the vertical disparity d_v with a second-order hyperbolic compression:

$$\mathbf{E}_v = 1/(1+(k_v.\mathbf{d}_v)^2) \qquad (1)$$

where $\mathbf{d}_v$ is the vertical disparity, and k_v is the disparity scaling constant.

The exponent of the power function was taken as 2 (corresponding to a disparity energy model prior to the compression). The disparity is combined with a constant minimum discomfort level for the interocular leakage factor R_0 by a Minkowski norm operator with an adjustable exponent, p. When $p = 2$, this becomes the Euclidean norm:

$$\mathbf{SDR} = (\ (R/\mathbf{E}_v)^p + R_0^{\ p}\ )^{1/p} \qquad (2)$$

where R is the rating scaling constant.

The model gave excellent fits to the two datasets, with chi square residuals of 1.03 and 0.49, respectively, implying that it captured effectively all of the consistent variance in the average data in each case. The model disparity scaling parameters were 21 and 24 arcmin for the 2D and 3D conditions, respectively, with a base rating level of 0.8 (mild discomfort).

If we take the 'moderate' discomfort level as the criterion for the maximum acceptable level of visual discomfort in the 3D display, the criterion SDR levels for vertical disparity were 14 and 17 arcmin for the 2D and 3D displays, respectively. Even if we limit the criterion to a severe discomfort level of 3.5, the criterion SDR levels are 32 and 38 arcmin for the 2D and 3D displays, respectively.

Torsional disparity ratings

The SDR values for the torsional disparity conditions are plotted in Fig. 3. The discomfort ratings for the 2D images (blue diamonds) start close to zero and rise gradually, reaching only the 'strong' discomfort level of 3 by about 50 deg of torsional disparity. The discomfort ratings for the 3D images (red squares) start at close to the 'mild' discomfort level of 1 and rise at a similar rate. As in the vertical disparity condition, the 3D display produced an irreducible level of mild discomfort even in the absence of torsional disparity. The fits to the two datasets had tight correlations 0.99 and 0.95 for the 2D and 3D viewing conditions, respectively.

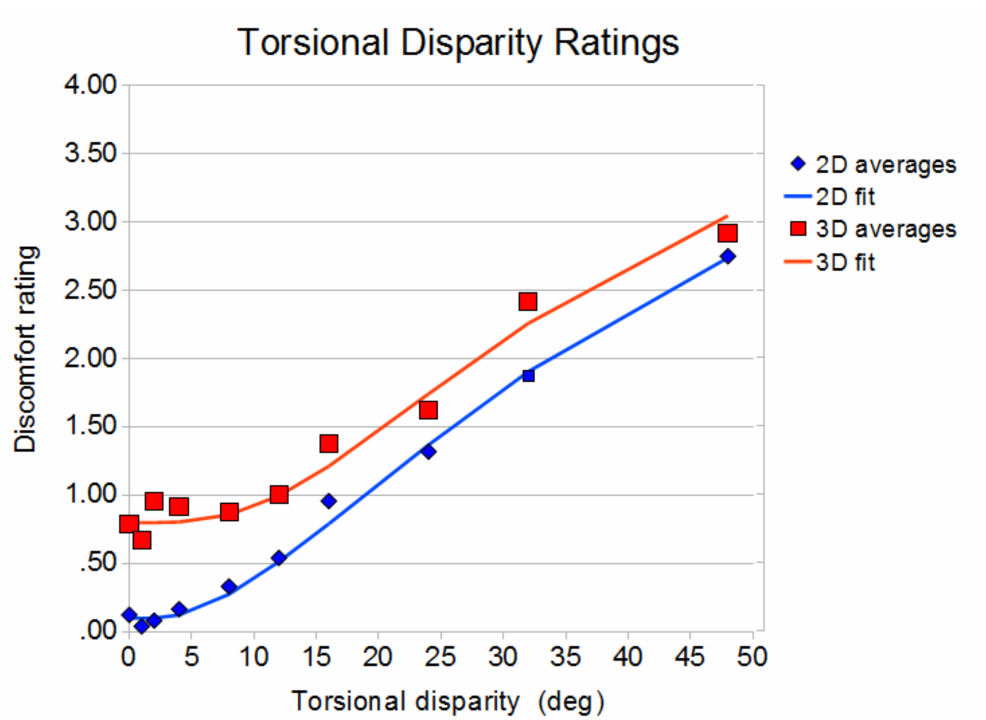


Fig. 3. Stereoscopic discomfort ratings (SDR) for the relative rotation of the images to the two eyes around the optic axes, generating a torsional disparity. Data are the averages for 9 subjects. Blue diamonds: 2D images; red squares: 3D images. Curves are the fits of the compressive Disparity Energy Model for the discomfort ratings of eq 1, specified by replacing d_v in eq 1 with a term for torsional disparity, d_θ.

The compressive disparity energy model for the discomfort ratings is specified by replacing d_v in eq 1 with a term for torsional disparity, d_θ. Although measured separately for the two conditions, the zero-misalignment disparity is successfully modeled with the same parameter for the two datasets. This function again gave excellent fits to the two SDR datasets, with chi square residuals of 2.36 and 2.51, respectively. Although these residual variances are significantly higher than in the case of the vertical disparity, they again imply that the model is a fully adequate description of the data in each case. The model parameters were 27 and 24 deg of torsional disparity between the two images for the 2D and 3D conditions, respectively.

Taking the 'moderate' discomfort level as the criterion for the maximum acceptable level of visual discomfort in the 3D display, the criterion levels of torsional disparity were 33 and 27 deg for the 2D and 3D displays, respectively.

DISCUSSION

Model fit

To convert the data to criterion levels, we could simply interpolate between the measured datum points, but it is more informative to fit the data with a model function that captures information about the underlying mechanisms generating the discomfort rating. One common model of disparity processing in recent years is the Disparity Energy Model[3-6], which, among other features, predicts that the strength of the disparity signal grows with the square of physical disparity, then saturates with some compression function.

We modeled this process with a second-order hyperbolic compression function for the prediction of discomfort as a function of the degree of misalignment disparity (eq 1). The model also incorporates a term for the basic discomfort of the 3D display in the absence of any misalignments. This term is combined with the misalignment disparity discomfort through a form of vector summation known as the Minkowski norm, which governs the transition from the display discomfort to the misalignment disparity discomfort. When the Minkowski exponent is 2, this combination rule becomes the standard Euclidean vector summation.

Criterion discomfort levels

At the criterion discomfort level of 2 (moderate discomfort), the criterion levels of vertical disparity were 14 and 17 arcmin for the 2D and 3D displays, respectively. Thus, as a rule of thumb, vertical disparity between the two eyes needs to be held below about 15 arcmin, or a quarter of a degree, to be within an acceptable range of discomfort. Similarly, for the torsional disparity, the criterion levels were 33 and 27 deg of image rotation, respectively. Thus, the human visual system seems to be relatively insensitive to torsional disparity, and holding it below about 30 deg seems to be an adequate rule of thumb for the acceptable amount of torsional disparity.

Even at the higher criterion level approaching severe discomfort, the vertical disparity limits were 32 and 38 arcmin, still only just over half a degree of vertical disparity allowable before reaching this level of discomfort. It is thus critical to keep the manufacturing tolerance well within a half a degree of vertical disparity in order to avoid user dissatisfaction with any device incorporating a pair of stereo cameras. The 30 deg allowance for torsional disparity is less critical, since this should well within the available manufacturing tolerances for the orientational alignments of pairs of cameras on any stereoscopic image capture device.

Inherent 3D display discomfort

It is noteworthy that viewing the 3D images in the absence of any misalignment disparity produced a measurable level of discomfort relative to the 2D viewing conditions (identical in the two eyes). It was evident from debriefing the participants that this irreducible level of 3D viewing discomfort was attributable to a display deficiency rather than to the presence of disparity *per se*. Looking around the 3D world or viewing through a window of 30 deg visual angle does not generally cause discomfort in most viewers.

In the display available for the study, and in all other 3D displays that we have encountered in a survey of 3D video displays (9 makes), the manufacturers have not succeeded in eliminating visible crosstalk between the images to the two eyes. Thus, in addition to the intended image of any disparate feature in the scene, one can often notice ghost images from the other eye on either side of the fused binocular image. These ghost images both detract from the perceived depth of the scene (since they form zero disparity pairings with the image for the intended eye), and cause discomfort by virtue of the triple images seen for every feature that has disparity, generating a kind of visual jitter and masking adjacent features in the scene.

In our assessment, the interocular crosstalk or leakage artifact in the disparate images was the precipitating factor for discomfort reported in the images with zero or negligible misalignment disparity (18 images). We cannot validate this assessment with certainty because the available displays do not allow the presentation of 3D images *without* crosstalk, but the absence of discomfort (essentially zero SDR) in the corresponding 2D ratings verifies that the discomfort in the 3D displays at zero disparity was not due to any other factor in the display configuration or the images selected for the study. One such factor, for example, was a slight interocular flicker due to the frame-sequential presentation to the two eyes, especially noticeable when they eyes jump around to inspect the display. However, this interocular flicker was also present in the 2D presentations, which were viewed under the 3D conditions but with identical images to the two eyes, so that they were rated at essentially zero discomfort by the subjects. Thus, the discomfort ratings for the zero misalignment conditions must have been due to some property of the disparities introduced in the 3D images, which we propose to have derived from the interocular crosstalk artifact. (The other possibility, that it derived from the stereoscopic disparity information *per se*, seems implausible because the world is replete with natural disparity information that rarely causes discomfort; the artificial factor of the interocular crosstalk seems a far more likely candidate.)

Disparity ranges

The above limits are designed to meet present day standards for 3D display quality. In the longer term, we may specify the limits that are required to provide fully artifact-free 3D displays that are free of any visual discomfort. Based on the present sample, this level of precision would be met if the vertical disparities were held below about 5 arcmin and the torsional disparities below about 12 deg of rotation.

We further note that even this level of precision is based on a typical sample, and that some individuals with extreme sensitivity to these misalignment disparities may require even finer precision in the alignment specification. It would require a larger scale study to determine the range of variation of sensitivity in the population and specify the proportion of the population that would be satisfied with each level of alignment precision.

Finally, although the contrast of the cross-talk was not measured in the present 3D display, it is worth noting that the human visual system is highly sensitive to the presence of low-contrast images, and that image structure of all kinds will be found in natural images. To completely eliminate visible cross-talk from such image structure in a 3D display, it is necessary to reduce its contrast to below about 0.1%[7] which is lower by about a factor of 10 than the crosstalk levels currently available on home displays. This requirement is not of concern in the design of 3D acquisition technology (since there is no crosstalk between pairs of 3D cameras), but is critical for the design of the 3D displays on which the results are viewed, including the viewfinders built into 3D acquisition devices. This very low level of crosstalk is therefore the ultimate goal of the 3D displays of the future.

REFERENCES

[1] Tyler CW, The horopter and binocular fusion. In, Vision and Visual Disorders. Vol 9, Binocular Vision. Regan D (Ed), Macmillan: New York, pp. 19-3 (1991).

[2] Kertesz AE, Jones RW. Human cyclofusional response.Vision Res. 10:891-6 (1970).

[3] Fleet DJ, Wagner H, Heeger DJ, Neural encoding of binocular disparity: Energy models, position shifts and phase shifts. Vision Res 36:1839-57 (1996).

[4] Ohzawa I, DeAngelis GC, Freeman RD, Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. Science 249: 1037-1041 (1990).

[5] Ohzawa I, Mechanisms of stereoscopic vision: the disparity energy model. Current Opinion Neurobiol 8: 509-515 (1998).

[6] Stürzl W, Hoffmann U, Mallot HA, Vergence control and disparity estimation with energy neurons: Theory and implementation. Artificial Neural Networks - ICANN 2002, 142-143 (2002).

[7] Van Nes FL, Bouman MA, Spatial modulation transfer in the human eye. J Opt Soc Amer 57:401-406 (1967).