# A novel stereoscopic display technique with improved spatial and temporal properties

Paul V. Johnson[1], Joohwan Kim[2], Martin S. Banks[1,2]
[1]UC Berkeley / UCSF Graduate Program in Bioengineering; [2]University of California, Berkeley, Vision Science Program, 360 Minor Hall, Berkeley CA, USA

## ABSTRACT

Common stereoscopic 3D (S3D) displays utilize either spatial or temporal interlacing to send different images to each eye. Temporal interlacing sends content to the left and right eyes alternatingly in time, and is prone to artifacts such as flicker, unsmooth motion, and depth distortion. Spatial interlacing sends even pixel rows to one eye and odd rows to the other eye, and has a lower effective spatial resolution than temporal interlacing unless the viewing distance is large. We propose a spatiotemporal hybrid protocol that interlaces the left- and right-eye views spatially, but the rows corresponding to each eye alternate every frame. We performed psychophysical experiments to compare this novel stereoscopic display protocol to existing methods in terms of spatial and temporal properties. Using a haploscope to simulate the three protocols, we determined perceptual thresholds for flicker, motion artifacts, and depth distortion, and we measured the effective spatial resolution. Spatial resolution is improved, flicker and motion artifacts are reduced, and depth distortion is eliminated. These results suggest that the hybrid protocol maintains the benefits of spatial and temporal interlacing while eliminating the artifacts, thus creating a more realistic viewing experience.

**Keywords**: stereoscopic 3D, flicker, depth distortion, psychophysics, perceptual artifacts

## 1. INTRODUCTION

Commercial S3D displays, with the exception of auto-stereoscopic displays, utilize either temporal interlacing or spatial interlacing to present different images to the two eyes, but both methods are prone to different types of perceptual artifacts that detract from viewing experience. Temporally interlaced displays utilize active shutter glasses to alternate left- and right-eye views in time, and are prone to visible flicker, unsmooth motion appearance, and depth distortions. Spatial interlacing, on the other hand, uses passive polarized glasses to send even pixel rows to one eye and odd pixel rows to the other eye, resulting in a lower effective spatial resolution at typical viewing distances. We propose a novel technique that is a spatiotemporal hybrid of these two methods. In our proposed hybrid protocol, the left- and right-eye views are interlaced spatially, but the rows corresponding to each eye alternate every frame.

Flicker visibility is well predicted by the amplitude and frequency of the Fourier fundamental of the luminance-varying signal from a display[2,4,11]. The duty cycle is the percentage of the frame time for which an image is presented; temporally interlaced S3D displays require duty cycles of 0.5 or less, while spatially interlaced displays have duty cycles of 1. Shorter duty cycles create greater amplitudes of the fundamental frequency, thus one expects more visible flicker to occur with temporal interlacing as opposed to spatial interlacing. In the hybrid protocol, flicker should be minimal because each eye is receiving content simultaneously for the entirety of the frame, creating an effective duty cycle of 1.

Capture rate, the number of unique images displayed per second to each eye, is considered the best predictor of motion artifacts[4]. The visibility of motion artifacts is primarily dictated by the monocular images because there is little if any effect of the phase of stimulation between the two eyes[2,4]. Eye movements are an important determinant of the appearance of motion on displays. If the observer fixates a stationary point on the screen, a moving object jumps across the retina in discrete steps, but each step is stationary for the duration of that step. The size of the shifts on the retina from one frame to the next depends on object speed. If this shift is too large, motion appears jerky or unsmooth, an effect known as *judder*. Viewers often track a moving object by making smooth-pursuit eye movements. With real objects in the natural environment, an accurate tracking movement makes the object's image stationary on the retina. With digitally displayed objects, the movement has a different effect. If the eye tracks at the time-average speed of the object, the

Stereoscopic Displays and Applications XXV, edited by Andrew J. Woods, Nicolas S. Holliman, Gregg E. Favalora, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 9011, 901105 · © 2014 SPIE-IS&T
CCC code: 0277-786X/14/$18 · doi: 10.1117/12.2040516

image is smeared across the retina for the duration of each sample-and-hold presentation. Coupled with the temporal integration of the eye, this causes image smearing which creates *motion blur*. Another effect called *banding* can occur with multi-flash presentations. In banding, repeated presentation of edges in an image creates the appearance of shifted edges that look like ghost images rather than blurred images.

Depending on how content is captured and displayed using the hybrid protocol, motion artifacts could be reduced compared to temporal interlacing. There are two possible methods to capture content for the hybrid protocol. The first, full capture, takes full resolution images for each eye, and needs two sub-frames to present them (Figure 1, panel 3). The second, partial capture, takes half resolution images but only needs one sub-frame to present it (Figure 1, panel 4). Using partial capture, content can be captured at twice the rate compared to temporal interlacing, theoretically resulting in smoother looking motion. At any particular frame the viewer will see a lower resolution image, but since the next frame "fills in" the missing information due to the rapid row switching, there should be a minimal decrease in resolution.

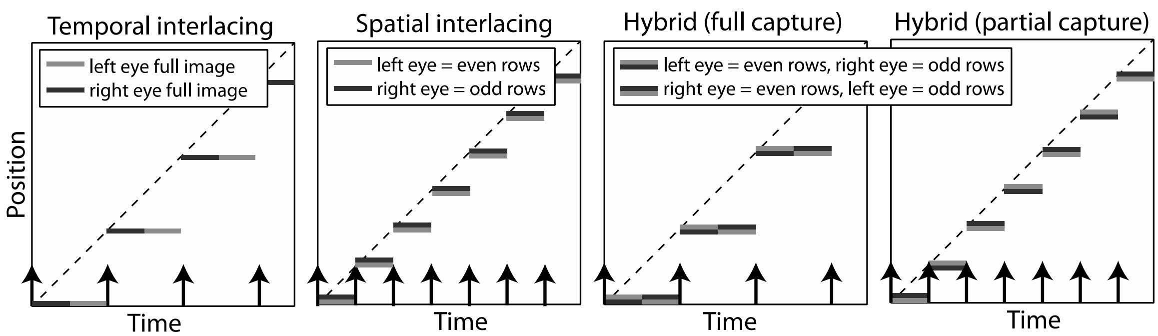


Figure 1. Four stereoscopic 3D display protocols. Arrows indicate the times at which content is captured. Note that spatial interlacing and hybrid (partial capture) allow for twice the capture rate of temporal interlacing, which should create smoother looking motion.

In temporal interlacing of 3D images the right and left eyes do not view images simultaneously, even though the camera system would typically capture the images simultaneously. This means that the second eye sees an image that lags the correct timestamp. When there is movement in the scene, the visual system can interpret these temporal lags as disparity and the depth in the scene will be distorted[1,4,9]. Consider the case of an object moving horizontally (Figure 2, left, dotted black line). The position of the object is captured with right and left cameras simultaneously at the times marked by black arrows. When the images are presented alternatingly on the screen, the left image (gray) is consistently delayed. If the visual system were to create matches between left- and right-eye views, the estimated disparity would depend on which reference image is chosen (Figure 2, right). Notice that every other disparity estimate is shifted upward from the desired disparity. At sufficiently high frame rates, the visual system will filter the discrete disparity estimates and achieve a continuous percept in between the discrete estimates. This causes horizontally moving objects to appear at incorrect depths, an illusion known as the Mach-Dvorak effect. The hybrid and spatial interlacing protocols should have minimal depth distortion because content is displayed simultaneously and there is no ambiguity as to how the visual system will match left- and right-eye views.

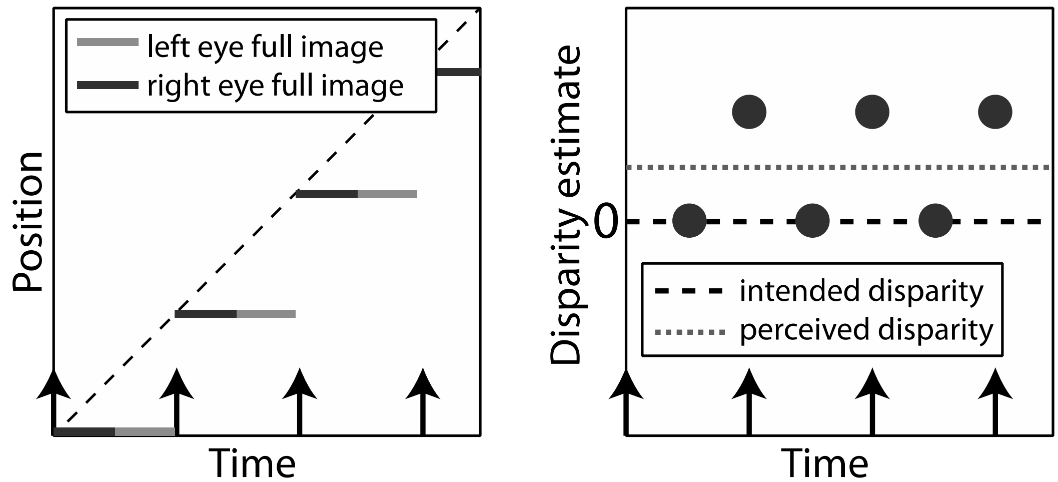


Figure 2. Left: Temporally interlaced stereo presentation. Right and left eye images are captured simultaneously but displayed in alternation. Right: The visual system attempts to match left and right eye images, but there is inherent ambiguity in the match, resulting in false matches 50% of the time. The perceived disparity is shifted from the intended disparity.

In spatial interlacing, each eye receives only a half resolution image in the vertical direction, resulting in a lower effective resolution than temporal interlacing at typical viewing distances. Similar to how the smoothness of motion is predicted by monocular images, effective spatial resolution is limited by what each individual eye sees independently[6]. For static images, hybrid interlacing should have improved spatial resolution relative to spatial interlacing because each eye receives a full resolution image, albeit over the course of two frames.

We demonstrate that this novel hybrid display protocol combines the better properties of temporal and spatial interlacing. In this study we compare spatial, temporal, and hybrid interlacing in terms of spatial resolution, flicker, motion artifacts, and depth distortion. We used a mirror stereoscope to simulate all three interlacing protocols. To the best of our knowledge, this hybrid interlacing technique has not been characterized. We demonstrate using psychophysical experiments that it has significant improvements over existing methods and should provide a better viewing experience.

## 2. EXPERIMENT 1: MOTION ARTIFACTS

### 2.1 Apparatus

Motion artifact experiments were carried out on a 2-monitor CRT mirror haploscope. A DATAPixx (VPixx Technologies) was used to precisely synchronize the two displays. SwitchResX was used to control the frame rate of the displays up to 200Hz.

### 2.2 Methods

Viewing distance was set so that pixels subtended 1 arcmin. We tested six subjects who had normal visual acuity and stereovision. We measured motion artifacts by presenting a series of moving 1° bright squares on a dark background. We found the stimulus speed above which motion artifacts–edge banding, motion blur, or judder–were perceived. Two eye movement conditions were used: a tracking case where subjects tracked the stimulus, and a fixation case where subjects fixated a cross and the stimulus moved below it. See Figure 3. Psychometric functions were fit and thresholds determined using psignifit[3], a statistical fitting software.

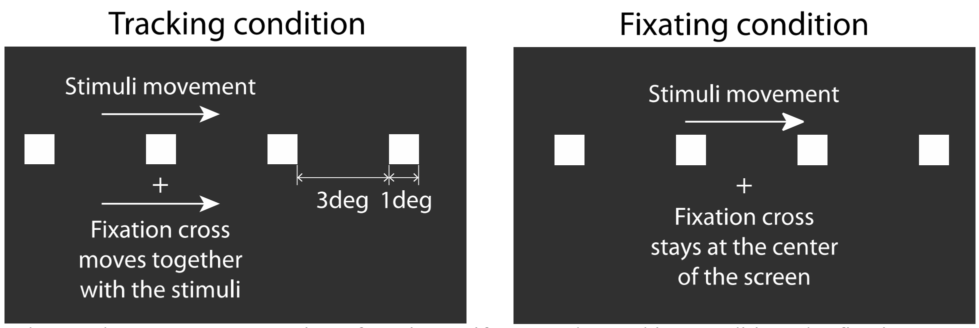


Figure 3. Stimulus used to measure perception of motion artifacts. In the tracking condition, the fixation target moved with the same velocity as the squares across the center of the display. In the fixating condition, the fixation target remained stationary.

### 2.3 Results

Results confirm that motion artifacts are best predicted by capture rate. For spatial interlacing and hybrid partial capture, the refresh rate and capture rate match because there are no sub-frames to present. For temporal interlacing and hybrid full capture, the refresh rate must be twice the capture rate. When we plot artifact threshold as a function of refresh rate, we see a clear benefit of hybrid partial capture compared to temporal interlacing, in the fixating case (see Figure 4). In the tracking case, this difference is slightly less pronounced due to a large degree of variability between subjects. In temporal interlacing, the spatial offset between left- and right- eye views is interpreted as disparity and the edges appear crisp, but in hybrid interlacing there is no way for the visual system to interpret the spatial offset as a disparity, so the edge appears blurry or jagged. We believe this is why the hybrid full capture protocol performed the worst, despite having the same capture rate as temporal interlacing.

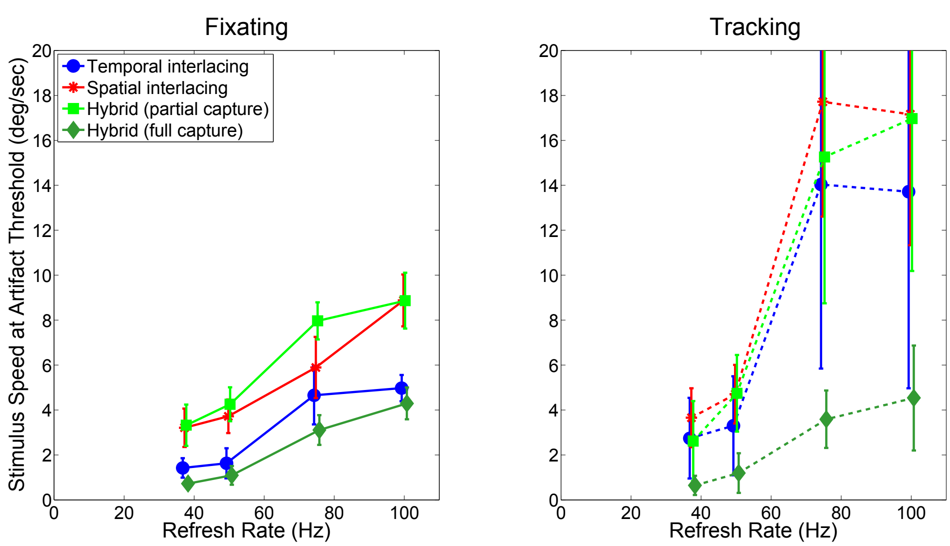


Figure 4. Motion artifacts in 4 display protocols. The stimulus speed at the artifact threshold is plotted as a function of refresh rate. Greater values on the ordinate axis indicate smoother motion. For the fixating case, motion looks markedly worse in the hybrid full capture and temporal interlacing protocols, as predicted. For the tracking case, the hybrid full capture had the most pronounced motion artifacts.

## 3. EXPERIMENT 2: FLICKER

### 3.1 Apparatus

Flicker experiments were done on a 1-monitor CRT mirror haploscope because the pixel size could be easily manipulated on such a setup.

### 3.2 Methods

We determined the critical flicker frequency for each display protocol by presenting a simple bright stimulus subtending 1° on a dark background at several different frame rates and having the viewer report whether they perceived visible fluctuations in luminance. In one condition subjects foveated the stimulus, and in another they fixated a cross and the stimulus appeared 4° below. The pixel pitch was varied between 0.5 arcmin and 2.4 arcmin. Psychometric functions were fit and thresholds determined using psignifit.

### 3.3 Results

Results from flicker measurements demonstrate that temporal interlacing had noticeably greater flicker than spatial interlacing, which had virtually no flicker at any refresh rate. The critical flicker frequency was 41Hz for temporal interlacing, while spatial interlacing had no visible flicker. Hybrid interlacing had no visible flicker when pixels subtended small angles, less than ~1 arcmin, or when the stimulus was in the periphery. When pixels were large the critical flicker frequency reached as high as 65Hz when pixels subtended 2.4 arcmin.

Flicker results are consistent with predictions. If we consider the contrast sensitivity function of the human visual system, our sensitivity to spatial frequencies drops precipitously beyond 30 cycles/deg[5,8]. A pixel pitch of 2 arcmin corresponds to 15 cycles/deg, to which we are highly sensitive. The rapid switching that occurs between even and odd rows in the hybrid protocol is therefore much more noticeable in this case, and we see perceptible flicker. Note that the recommended viewing distance for HD television is 3 picture heights, corresponding to 1 arcmin. Hybrid interlacing should have negligible flicker at this viewing distance.

The visual system is generally more sensitive to flicker in the periphery, but contrast sensitivity is also substantially reduced. The critical flicker frequency for temporal interlacing was unchanged between foveal and peripheral viewing, likely because critical flicker frequency does not change very much within 5° of the fovea[10]. Our hybrid result is consistent with contrast sensitivity in peripheral vision being extremely poor.

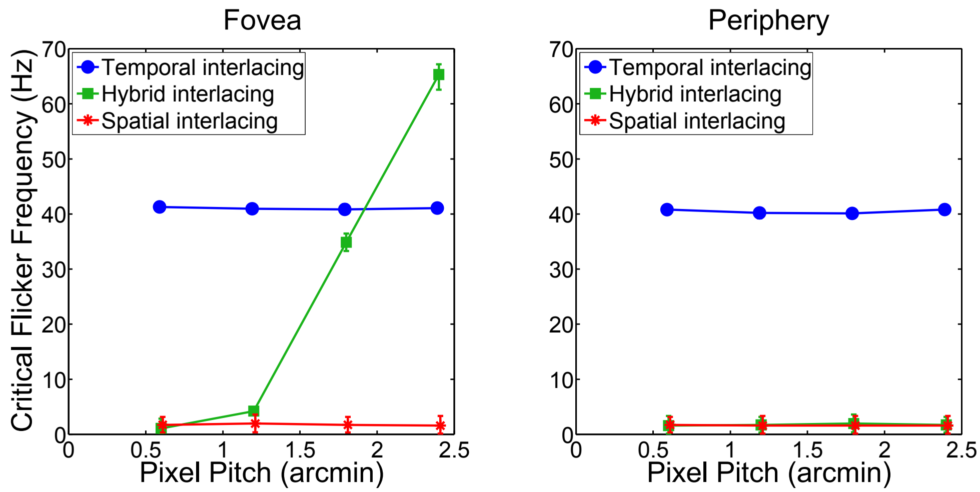


Figure 5. Critical flicker frequency as a function of pixel pitch. Plotted above are flicker thresholds for spatial, temporal, and hybrid interlacing, pooled across all subjects. For temporal and spatial interlacing, thresholds do not change as a function of pixel size. For hybrid interlacing, flicker becomes noticeable when pixels are large but is absent when pixels are small.

## 4. EXPERIMENT 3: SPATIAL RESOLUTION

### 4.1 Apparatus

To measure effective spatial resolution for the different S3D protocols, the same 1-monitor mirror haploscope was used as in Experiment 1.

### 4.2 Methods

We determined effective spatial resolution with a "tumbling E" task where viewers indicate the orientation of a letter E presented stereoscopically with several different sizes. See Figure 6. Three viewing distances were used so that we could better ascertain the limitations of the display regardless of individuals' visual acuity. Pixel pitch varied between 0.5 arcmin and 2 arcmin. Psychometric functions were fit and thresholds determined using psignifit.

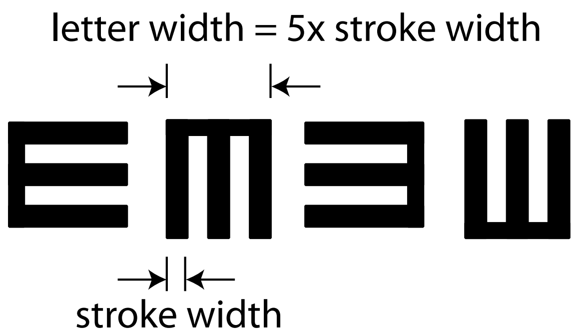


Figure 6. Stimulus used to measure effective spatial resolution.

### 4.3 Results

See Figure 7 for a plot of threshold viewing angle against pixel pitch. The threshold viewing angle was based on the stroke width of the stimulus at a size when subjects responded correctly 62.5% of the time. When pixels subtend a relatively large angle of 2 arcmin, hybrid interlacing performs similarly to temporal interlacing. Both temporal and hybrid interlacing have better effective resolution than spatial interlacing (paired t-test, $p<0.02$). When pixels are indistinguishable, around 0.5 arcmin, all three display protocols are equivalent because the limiting factor becomes human

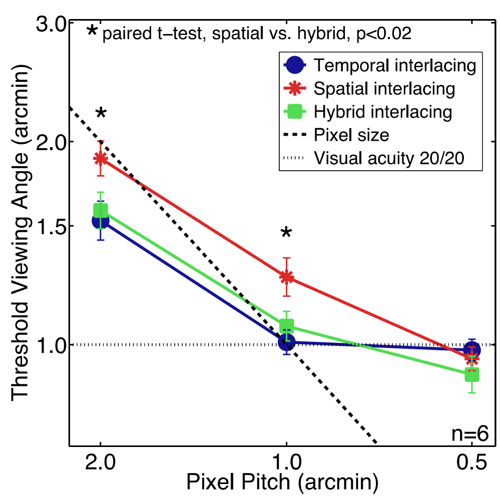


Figure 7. Effective resolution of three display protocols. Threshold viewing angle, or the minimum stroke width resolvable at threshold performance level, is plotted as a function of pixel pitch. When pixels are large (2 arcmin), temporal and hybrid protocols have improved resolution compared to spatial interlacing (paired t-test, $p<0.02$). When pixels are small (0.5 arcmin), there are no

significant differences, as the limit to performance is human visual acuity at this viewing distance, rather than anything to do with the display.

## 5. EXPERIMENT 4: DEPTH DISTORTION

### 5.1 Apparatus

The apparatus was the same as the 2-monitor mirror haploscope used in Experiment 1.

### 5.2 Methods

We measured depth distortion by presenting two stimuli moving horizontally in opposite directions and having users adjust the disparity until they appeared at the same depth. The stimuli consisted of several 1° bright squares separated by 3° (see Figure 8). Blur was added in the horizontal direction to reduce the appearance of motion artifacts and make the task more effective. Psychometric functions were fit and thresholds determined using psignifit.

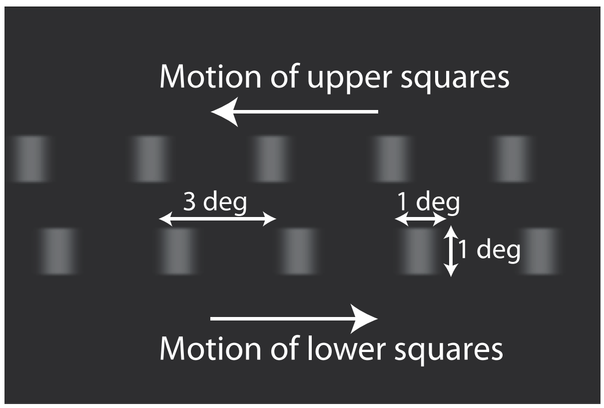


Figure 8. Illustration of the experimental stimulus used to measure depth distortion. Two groups of bars moved horizontally in opposite directions. If there was any depth distortion, one group appeared to be closer while the other appeared farther than the display surface. The task was to choose which group appeared closer (2 alternative forced choice).

### 5.3 Results

Temporal interlacing resulted in significant depth distortion, while spatial and hybrid interlacing had virtually no depth distortion (see Figure 9). Our prediction for depth distortion is based on interocular delay:

$$\Delta = s\delta, \tag{1}$$

where $s$ is the speed in degrees per second and $\delta$ is the interocular delay in seconds. For hybrid and spatial interlacing, presentation is effectively simultaneous so the interocular delay is 0 sec. For temporal interlacing, interocular delay is dependent on the refresh rate of the display: in this case, 100Hz.

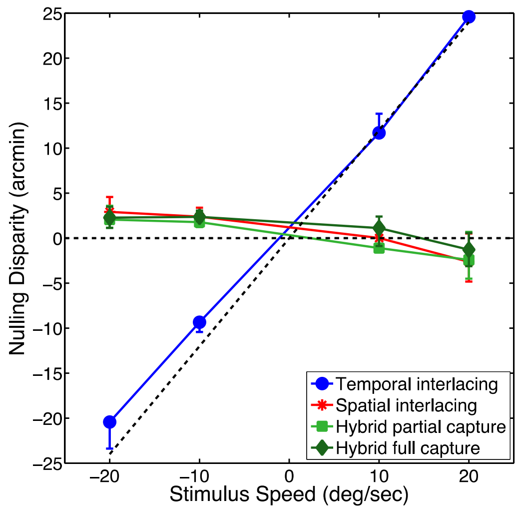


Figure 9. Plotted is the nulling disparity necessary to make the right- and left-moving objects appear at the same depth for a given stimulus speed. The diagonal dotted black line is the prediction based on equation 1. Spatial and hybrid interlacing appear to have little depth distortion, and temporal interlacing has depth distortion in line with predictions.

## 6. CONCLUSION

Depth distortion and flicker were significantly reduced with the hybrid protocol compared to temporal interlacing. Since capture rate is the best predictor of motion artifacts, hybrid motion was equivalent to temporal interlacing, unless we present only half the information at each frame and effectively double the capture rate. In this case, hybrid motion looks substantially smoother than temporal interlacing under certain conditions. The effective spatial resolution of hybrid interlacing was improved compared to spatial interlacing. The results suggest that hybrid interlacing combines the best qualities of both spatial and temporal interlacing and thereby provides a more realistic perceptual experience.

## ACKNOWLEDGMENT

This research was supported by NIH research grant R01EY012851.

## REFERENCES

[1] Burr, D. C. and Ross, J., "How does binocular delay give information about depth?," Vision Research 19(5), 523-532 (1979).

[2] Cavonius, C. R., "Binocular interactions in flicker," Quarterly Journal of Experimental Psychology 31(2), 273-280 (1979).

[3] Fründ, I., Haenel, N. V., and Wichmann, F. A., "Inference for psychometric functions in the presence of nonstationary behavior," Journal of Vision, 11(6) (2011).

[4] Hoffman, D. M., Karasev, V. I., and Banks, M. S., "Temporal presentation protocols in stereoscopic displays: Flicker visibility, perceived motion, and perceived depth," Journal of the Society for Information Display 19(3), 271-297 (2011).

[5] Kelly, D. H., "Motion and vision. II. Stabilized spatio-temporal threshold surface," JOSA 69(10), 1340-1349 (1979).

[6] Kim, J. S., and Banks, M. S., "Effective Spatial Resolution of Temporally and Spatially Interlaced Stereo 3D Televisions," SID Symposium Digest of Technical Papers, Blackwell Publishing Ltd., 43(1), 879-882 (2012).

[7] Kim, J. S., Johnson, P. V., and Banks, M. S., "The visibility of temporal artifacts in stereo 3D displays," 8th Asia-Pacific Conference on Vision, 62 (2012).

[8] Mannos, James, and Sakrison, D., "The effects of a visual fidelity criterion of the encoding of images," Information Theory, IEEE Transactions on, 20(4), 525-536 (1974).

[9] Read, J. C. and Cumming, B. G, "The stroboscopic Pulfrich effect is not evidence for the joint encoding of motion and depth," Journal of Vision, 5(5) (2005).

[10] Tyler, C. W., "Analysis of visual modulation sensitivity. III. Meridional variations in peripheral flicker sensitivity," JOSA A 4(8), 1612-1619 (1987).

[11] Watson, A. B., Ahumada, A. J., and Farrell, J. E., "Window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays," Journal of the Optical Society of America A 3(3), 300-307 (1986).