

UC Berkeley

UC Berkeley Previously Published Works

Title

Stereoscopic 3D display technique using spatiotemporal interlacing has improved spatial and temporal properties

Permalink

<https://escholarship.org/uc/item/7mx0q6r1>

Journal

Optics Express, 23(7)

ISSN

1094-4087

Authors

Johnson, Paul V
Kim, Joohwan
Banks, Martin S

Publication Date

2015-04-06

DOI

10.1364/oe.23.009252

Peer reviewed

Stereoscopic 3D display technique using spatiotemporal interlacing has improved spatial and temporal properties

Paul V. Johnson,^{1*} Joohwan Kim,² and Martin S. Banks^{1,2}

¹UC Berkeley - UCSF Graduate Program in Bioengineering, Berkeley, CA, USA

²University of California, Berkeley, Vision Science Program, Berkeley CA, USA

*pvjohn98@gmail.com

Abstract: Stereoscopic 3D (S3D) displays use spatial or temporal interlacing to send different images to the two eyes. Temporal interlacing delivers images to the left and right eyes alternately in time; it has high effective spatial resolution but is prone to temporal artifacts. Spatial interlacing delivers even pixel rows to one eye and odd rows to the other eye simultaneously; it is subject to spatial limitations such as reduced spatial resolution. We propose a spatiotemporal-interlacing protocol that interlaces the left- and right-eye views spatially, but with the rows being delivered to each eye alternating with each frame. We performed psychophysical experiments and found that flicker, motion artifacts, and depth distortion are substantially reduced relative to the temporal-interlacing protocol, and spatial resolution is better than in the spatial-interlacing protocol. Thus, the spatiotemporal-interlacing protocol retains the benefits of spatial and temporal interlacing while minimizing or even eliminating the drawbacks.

©2015 Optical Society of America

OCIS codes: (330.1400) Vision - binocular and stereopsis; (330.6130) Spatial resolution; (330.5510) Psychophysics; (120.2040) Displays; (100.2000) Digital image processing.

References and links

1. D. M. Hoffman, V. I. Karasev, and M. S. Banks, "Temporal presentation protocols in stereoscopic displays: Flicker visibility, perceived motion, and perceived depth," *J. Soc. Inf. Disp.* **19**(3), 271–297 (2011).
2. J. S. Kim and M. S. Banks, "Effective Spatial Resolution of Temporally and Spatially Interlaced Stereo 3D Televisions," *SID Symp. Dig. Tec.* **43**(1), 879–882 (2012).
3. J. Hakala, P. Oittinen, and J. Häkkinen, "Depth artifacts caused by spatial interlacing in stereoscopic 3D displays," *ACM Trans. Appl. Percept.* in press.
4. S. Sechrist, "Display Week 2011 Review: 3-D," *Inform. Display* **27**(7–8), 16–18 (2011).
5. D. H. Kelly, "Visual contrast sensitivity," *Opt. Acta (Lond.)* **24**(2), 107–129 (1977).
6. P. V. Johnson, J. S. Kim, D. M. Hoffman, A. Vargas, and M. S. Banks, "Motion artifacts in 240Hz stereoscopic 3D displays," *J. Soc. Inf. Disp.* in press.
7. P. J. Bex, G. K. Edgar, and A. T. Smith, "Multiple images appear when motion energy detection fails," *J. Exp. Psychol.* **21**(2), 231–238 (1995).
8. A. B. Watson, A. J. Ahumada, Jr., and J. E. Farrell, "Window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays," *J. Opt. Soc. Am. A* **3**(3), 300–307 (1986).
9. D. G. Stork and D. S. Falk, "Temporal impulse responses from flicker sensitivities," *J. Opt. Soc. Am. A* **4**(6), 1130–1135 (1987).
10. D. H. Brainard, "The Psychophysics Toolbox," *Spat. Vis.* **10**(4), 433–436 (1997).
11. D. G. Pelli, "The VideoToolbox software for visual psychophysics: Transforming numbers into movies," *Spat. Vis.* **10**(4), 437–442 (1997).
12. F. A. Wichmann and N. J. Hill, "The psychometric function: I. Fitting, sampling, and goodness of fit," *Percept. Psychophys.* **63**(8), 1293–1313 (2001).
13. F. A. Wichmann and N. J. Hill, "The psychometric function: II. Bootstrap-based confidence intervals and sampling," *Percept. Psychophys.* **63**(8), 1314–1329 (2001).
14. I. Fründ, N. V. Haenel, and F. A. Wichmann, "Inference for psychometric functions in the presence of nonstationary behavior," *J. Vis.* **11**(6), 16 (2011).

15. D. M. Hoffman, P. V. Johnson, J. S. Kim, A. Vargas, and M. S. Banks, "240Hz OLED technology properties that can enable improved image quality," J. Soc. Inf. Disp. in press.
16. C. R. Cavonius, "Binocular interactions in flicker," Q. J. Exp. Psychol. **31**(2), 273–280 (1979).
17. D. H. Kelly, "Motion and vision. II. Stabilized spatio-temporal threshold surface," J. Opt. Soc. Am. **69**(10), 1340–1349 (1979).
18. J. J. Koenderink, M. A. Bouman, A. B. de Mesquita, and S. Slappendel, "Perimetry of contrast detection thresholds of moving spatial sine wave patterns. II. The far peripheral visual field (eccentricity 0–50)," J. Opt. Soc. Am. **68**(6), 850 (1978).
19. ITU-R Recommendation BT, 709–5: "Parameter values for the HDTV standards for production and international programme exchange," International Telecommunication Union, Geneva, Switzerland (2002).
20. ITU-R Recommendation BT, 2022: "General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays," International Telecommunication Union, Geneva, Switzerland (2012).
21. ITU-R Recommendation BT, 2020–1: "Parameter values for ultra-high definition television systems for production and international programme exchange," International Telecommunication Union, Geneva, Switzerland (2014).
22. J. Davis, Y.-H. Hsieh, and H.-C. Lee, "Humans perceive flicker artifacts at 500Hz," Sci Rep **5**, 7861 (2015), doi:10.1038/srep07861.
23. R. M. Soneira, "3D TV display technology shoot-out," http://www.displaymate.com/3D_TV_ShootOut_1.htm (2011).
24. S. de Witt, "Active/Passive 3D Myth Revisited," <https://stijndewitt.wordpress.com/2012/03/03/active-vs-passive-3d-myth-revisited/> (2012).
25. D. Didyk, E. Eisemann, T. Ritschel, K. Myszkowski, and H.-P. Seidel, "Apparent display resolution enhancement for moving images," ACM T. Graphic **29**(4), 1–8 (2010).
26. D. C. Burr and J. Ross, "How does binocular delay give information about depth?" Vision Res. **19**(5), 523–532 (1979).
27. J. C. Read and B. G. Cumming, "The stroboscopic Pulfrich effect is not evidence for the joint encoding of motion and depth," J. Vis. **5**(5), 417–434 (2005).
28. D. Kane, P. Guan, and M. S. Banks, "The limits of human stereopsis in space and time," J. Neurosci. **34**(4), 1397–1408 (2014).

1. Introduction

Stereoscopic 3D (S3D) displays send slightly different images to the two eyes, creating binocular disparity, which yields an enhanced sensation of depth relative to conventional displays. Nearly all S3D displays use temporal interlacing or spatial interlacing to present disparate images to the left and right eyes. Temporal interlacing delivers the left- and right-eye views alternately in time. This is often accomplished by using liquid-crystal shutter glasses that alternately transmit and block the images to the eyes in synchrony with the display. Thus, only one eye receives light at a given moment, but it receives all the pixels. This protocol is schematized in the first panel of Figs. 1 and 2. Spatial interlacing delivers even pixel rows to one eye and odd pixel rows to the other eye simultaneously. This is typically done using a film-patterned retarder on the display that polarizes the emitted light in opposite directions row by row. The polarization can be linear or circular. The viewer wears passive eyewear that transmits alternate rows to the two eyes. Thus, both eyes receive light at any given moment, but each receives only half the pixels. This protocol is schematized in the second panel of Figs. 1 and 2.

The two methods have different shortcomings from a perceptual standpoint. Temporal interlacing is prone to temporal artifacts such as flicker, unsmooth motion appearance, and distortions of perceived depth [1]. Spatial interlacing results in lower spatial resolution at typical viewing distances [2] and can also cause distortions of perceived depth [3]. We sought a technique that would combine the better features of the two protocols—spatial resolution with temporal interlacing and temporal performance with spatial interlacing—while minimizing their shortcomings. In the proposed spatiotemporal-interlacing protocol, the left- and right-eye views are interlaced spatially, but the rows presented to each eye alternate temporally. For brevity, we will henceforth refer to the proposed technique as the *hybrid protocol*.

To describe the protocols we tested, it is useful to define some terms clearly. A *display frame* is the minimal time during which the assignment of a pixel value is maintained. A new

assignment can occur either to update the image content or to interlace for stereo presentation. Different presentation techniques can require different numbers of display frames to present images to the two eyes. For example, temporal interlacing requires two display frames because it presents one eye's view at one time and the other eye's view at another. Spatial interlacing requires only one display frame because it shows both eyes' views simultaneously. *Display frame rate* is the number of display frames per unit time. This is the native frame rate of the display. *Capture rate* is the number of unique captured (or generated) images per unit time, and is strictly less than or equal to the display frame rate. *Presentation rate* is the number of images (unique or not) presented per unit time. In multi-flash procedures, the presentation rate is the capture rate multiplied by the number of flashes. For example, in the popular triple-flash protocol used by RealD for cinema, the capture rate is 24Hz for each eye but each captured image is displayed three times within a frame for a presentation rate per eye of 72Hz.

There are two possible methods to capture and present content with the hybrid protocol: *dual frame* and *single frame*. In the dual-frame protocol, the captured data are presented over two display frames. In the first display frame, the odd rows of the left-eye's image data are displayed in odd rows on the screen and are seen by the left eye, and the even rows of the right-eye's data are displayed in even rows on the screen and seen by the right eye. In the second sub-frame, the even rows of the left-eye's image data are displayed in even rows on the screen and are seen by the left eye, and the odd rows of the right-eye's data are displayed in odd rows and seen by the right eye. Because each display frame presents half the pixel rows of each eye's view, the protocol can present all the captured data. The dual-frame hybrid technique is schematized in the third panel of Figs. 1 and 2. In the single-frame protocol, the captured data are presented on one display frame and updated on every successive frame. In one frame, the odd rows of the left-eye's image data are displayed in odd rows on the screen and are seen by the left eye, and the even rows of the right-eye's image data are displayed in even rows on the screen and seen by the right eye. In the next frame, new image data are shown, but now the even rows of the left-eye data are displayed in even rows on the screen to be seen by the left eye, and the odd rows of the right-eye's data are displayed in odd screen rows to be seen by the right eye. The single-frame hybrid protocol therefore shows only half of the captured data on each frame, but the capture rate is twice that of dual-frame hybrid. This technique is schematized in the fourth panel of Figs. 1 and 2. We compared the four techniques for a fixed display frame rate, so in Figs. 1 and 2 the image data in spatial interlacing and single-frame hybrid are updated at twice the rate as in temporal interlacing and dual-frame hybrid.

RealD and Samsung developed a presentation technique [4] similar to the hybrid technique proposed here. Their technique uses two display frames to present S3D image data. Pixel rows are divided into eight blocks across the screen. During the first frame, odd-numbered blocks (1st, 3rd, 5th, and 7th from the top) present the left-eye view and even-numbered blocks (2nd, 4th, 6th, and 8th) the right-eye view. The views swap eyes for the second frame. Hence the difference between the RealD/Samsung technique and the one we propose is how pixel rows get assigned to left- and right-eye views. The RealD/Samsung method spatially alternates between left- and right-eye views every 135 rows (if there are 1080 pixel rows as in HDTV) while our technique does it every other row. At the recommended viewing distance for HDTV, pixel rows subtend 1 arcmin, so the blocks in the RealD/Samsung technique would subtend 135 arcmin, yielding a fundamental spatial frequency of 0.2 cycles/deg (cpd). The blocks in our technique subtend 1 arcmin for a fundamental frequency of 30 cpd. The visual system is much more sensitive to spatiotemporal variations at 0.2 cpd than to such variations at 30 cpd [5], so our technique should provide substantially better image quality and substantially fewer temporal artifacts than the RealD/Samsung technique.

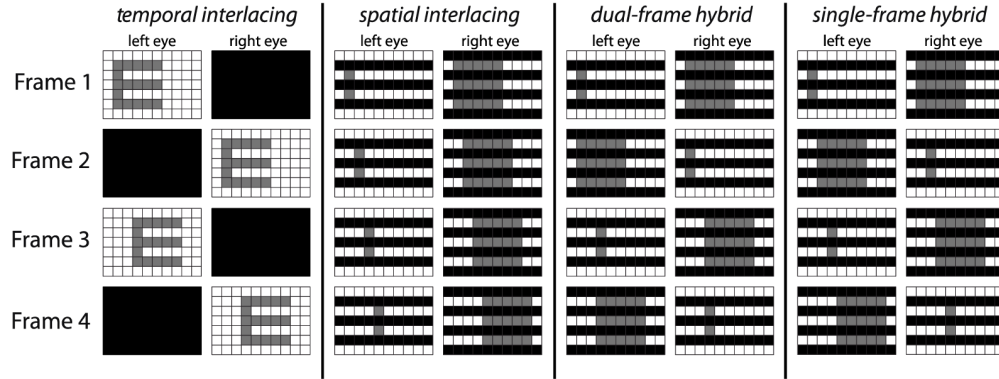


Fig. 1. S3D display protocols. From left to right, the protocols schematized are temporal interlacing, spatial interlacing, dual-frame hybrid, and single-frame hybrid. To schematize the protocols, we show the images seen by the left and right eyes in two columns for each protocol. Time proceeds from top to bottom. The grid pattern in each panel represents pixels. The stimulus being displayed is a black letter “E” with a height and width of 5 pixels. The stimulus is moving rightward by one pixel per frame such that by frame 5, the E has moved four pixels rightward in all protocols. Black represents pixels that are not displayed to an eye at a given time. In the temporal interlacing and dual-frame hybrid protocols, two display frames are required to show the data captured at one time to both eyes. In the spatial-interlacing and single-frame hybrid protocols, updated image data are shown on every display frame, so the E moves from its previous location with every display frame.

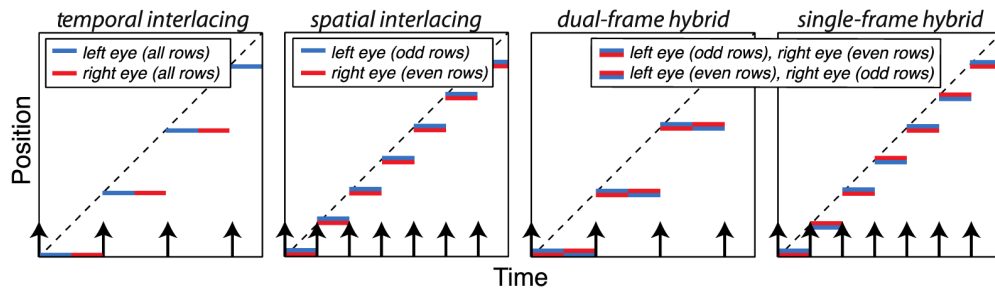


Fig. 2. S3D display protocols schematized in space-time plots. From left to right, the protocols schematized are temporal interlacing, spatial interlacing, dual-frame hybrid, and single-frame hybrid. Each panel plots position on the screen as a function of time for a stimulus moving at constant speed. The dashed lines represent the object’s motion in the real world. The blue and red lines represent the motion on a digital display, blue for images seen by the left eye and red for images seen by the right eye. We assumed a display with a fixed frame rate. Black arrows indicate the times at which content was captured. With a fixed frame-rate display, spatial interlacing and single-frame hybrid allow for presentation of twice the capture rate compared to temporal interlacing and dual-frame hybrid.

We investigated motion artifacts, flicker, spatial resolution, and depth distortions in four protocols: temporal interlacing, spatial interlacing, dual-frame hybrid, and single-frame hybrid. We found that the proposed hybrid protocol—specifically, the single-frame protocol—has the better properties of temporal and spatial interlacing. Specifically, it has the effective spatial resolution of a temporal-interlaced display while avoiding flicker, motion artifacts, and depth distortions that occur with temporal interlacing.

2. Experiment 1: Motion artifacts

Motion artifacts include *judder* (jerky or unsmooth motion appearance), *motion blur* (apparent smearing in the direction of stimulus motion), and *banding* (appearance of multiple edges in the direction of stimulus motion). The best predictors of motion artifacts are capture rate and the speed of a moving object: Artifacts become more visible with decreasing capture

rate and increasing speed [1,6]. Motion appearance also depends on whether the viewer holds fixation stationary as an object moves by or makes a smooth eye movement to track the object. With stationary fixation, the object jumps across the retina by some distance in every frame, the jump size being proportional to object speed. Bex and colleagues [7] described a largest spatial displacement beyond which temporal aliasing occurs. When the aliases fall within the range of visible spatial and temporal frequencies, judder occurs [1,8]. With tracking eye movements, the image of a real moving object would be stationary on the retina. When viewing simulated movement on a digital display, the eye tracks at the time-average speed of the object, so the 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 motion blur. Longer duty cycles and faster object motions create more visible blur [6]. Banding occurs with multi-flash presentations: repeated presentation of edges creates the appearance of shifted edges that look like ghost images rather than blurred images.

In the first experiment, we determined the visibility of motion artifacts for the four display protocols, for different object speeds, and when the viewer held fixation stationary or tracked the object.

2.1 Subjects

Six subjects, ages 22 to 32 years, participated. All had normal or corrected-to-normal visual acuity and stereo acuity. Two were authors; the others were not aware of the experimental hypotheses. In all experiments, appropriate consent and debriefing were done according to the Declarations of Helsinki.

2.2 Apparatus

Psychophysical experiments were carried out on a two-display mirror stereoscope. The displays were CRTs (Iiyama HM204DT). A DATAPixx data-acquisition and graphics toolbox (VPixx Technologies) was used to synchronize the two displays precisely. A software package, SwitchResX (www.madrau.com), was used to control the frame rate and resolution of the displays. The resolution of each display was 1200×900 , so pixels subtended 1 arcmin at the viewing distance of 115 cm. The CRT frame rate was either 100Hz or 75Hz as needed to simulate different capture and presentation rates. Different duty cycles were simulated by presenting one or more CRT frames.

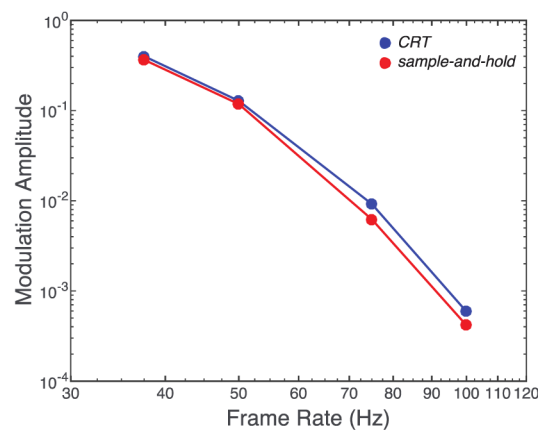


Fig. 3. Calculated luminance modulation for CRT and sample-and-hold displays. Peak-to-trough luminance modulation after filtering by the human temporal impulse-response function is plotted as a function of display frame rate. Blue represents modulation for the CRTs and red represents modulation for a sample-and-hold display with instantaneous on and off responses.

We simulated low display frame rates by repeating CRT frames within a simulated display frame. For instance, we simulated a display frame rate of 50Hz using a CRT refresh rate of 100Hz and repeating each CRT frame twice before updating. The refresh rate of the CRTs was high, so temporal filtering in early vision should make the stimulus effectively the same as an actual sample-and-hold presentation. We checked this by conducting a simulation. We first measured the impulse-response function of the CRTs. We then created sequences of impulses used to simulate the sample-and-hold presentations and convolved them with the temporal impulse-response function of the human visual system [9]. The stimulus was a uniform white field. In the simulation, we calculated the temporal modulation of luminance for the CRT and sample-and-hold display. Figure 3 plots those modulations as a function of display frame rate and shows that they were very similar. Thus, our means of simulating a sample-and-hold display was valid.

2.3 Methods

We used the stereoscope to send the appropriate content to each eye at each moment in time. By so doing, we could simulate the four S3D display protocols (Figs. 1 and 2). We measured the visibility of motion artifacts by presenting a series of moving 1° bright squares separated by 3° on an otherwise dark background (Fig. 4). Stimulus duration was 1 sec. Stimuli were presented binocularly with zero disparity. We used MATLAB with the Psychophysics Toolbox extension to render and display all content [10,11]. We adjusted luminances so that stimulus contrast was equivalent in the four display protocols. We presented two eye-movement conditions: a tracking condition, in which subjects made smooth movements to track the stimulus, and a non-tracking condition, in which fixation was stationary as the stimulus moved by. In the tracking condition, a fixation cross was presented to one side 0.5 sec before the stimulus appeared. The cross moved across screen with the stimulus to aid tracking (Fig. 4, left). In the non-tracking condition, a fixation cross was presented at screen center 0.5 sec before the onset of the stimulus. Then the stimulus moved adjacent to the stationary cross (Fig. 4, right). Stimulus motion was either horizontal (left to right, or right to left) or vertical (top to bottom, or bottom to top).

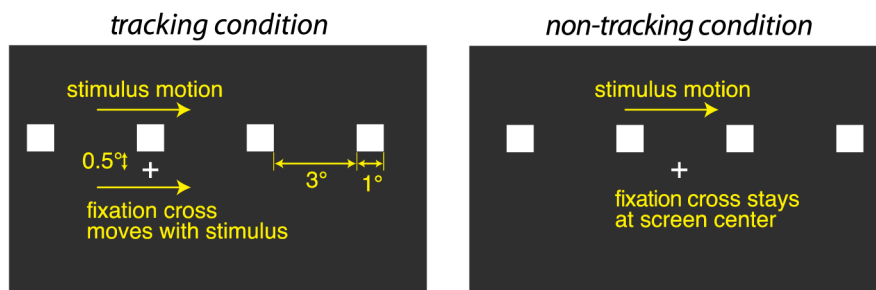


Fig. 4. Stimulus used to measure visibility of motion artifacts. In the tracking condition, the fixation target moved with the same velocity as the squares. In the non-tracking condition, the fixation target remained stationary as the squares moved by. In both cases, the fixation target appeared 0.5 sec before stimulus onset.

Subjects indicated after each trial whether they had seen motion artifacts or not, regardless of the type of artifact (e.g., edge banding, blur, or judder). This was a yes-no, single-presentation method. A 1-up/1-down adaptive staircase procedure adjusted the speed of the stimulus to estimate the value that just yielded motion artifacts. Twenty trials were presented for each staircase. Staircases were randomly interleaved within an experimental session. Maximum speed was $20^\circ/\text{sec}$. We fit the data with a cumulative Gaussian (with lower and upper asymptotes of 0 and 1) whose parameters were determined with a maximum-likelihood criterion [12–14]. Henceforth we report the stimulus speed at which the Gaussian crossed 0.5,

which is an estimate of the speed at which motion artifacts were perceived on half the trials. When we averaged across subjects, we did so by pooling the psychometric data from all subjects and then fitting those data with one cumulative Gaussian function.

The experiment consisted of ~1280 trials per subject: 4 display protocols \times 4 capture rates \times 2 eye-movement conditions \times 2 directions \times 20 trials. The data for horizontal and vertical motion were very similar, so we combined the data from the two directions yielding 40 trials per psychometric function fitting. It took about 1 hour for each subject to complete the experiment.

2.4 Results and discussion

Figure 5 plots the data from the non-tracking condition. Each panel shows the object speed at which observers reported motion artifacts on half the trials as a function of display frame rate. Different colors represent the data from different protocols. There were clear differences across observers in the speeds at which they reported artifacts, but they all exhibited the same effects across protocols. Differences between subjects could be due to differences in their sensitivity to artifacts, as well as differences in their response criterion. Of greatest interest is how the different protocols fared in terms of artifact visibility. The results show, as expected, that temporal interlacing is more prone to motion artifacts than spatial interlacing [1,6]. Artifacts with the dual-frame hybrid protocol were similar to those with temporal interlacing, while artifacts with the single-frame hybrid protocol were similar to those with spatial interlacing. Thus, the single-frame version of the hybrid technique is relatively unsusceptible to motion artifacts.

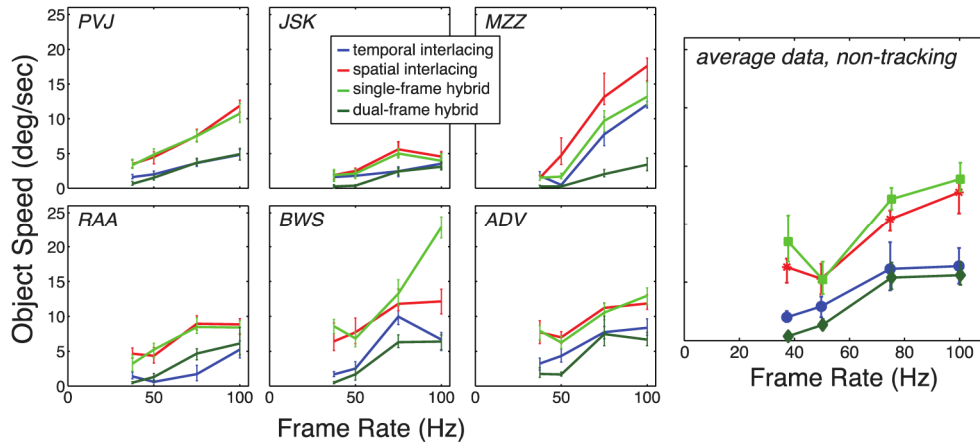


Fig. 5. Visibility of motion artifacts for the four protocols in the non-tracking condition. The six panels on the left show the data from individual observers. The large panel on the right shows the data averaged across observers. Each panel plots the stimulus speed at which artifacts were reported on half the trials as a function of display frame rate. Blue, red, bright green, and dark green represent the results for temporal interlacing, spatial interlacing, single-frame hybrid, and dual-frame hybrid, respectively. Error bars represent 95% confidence intervals. Temporal interlacing and dual-frame hybrid require two display frames per capture period, while spatial interlacing and single-frame hybrid require only one frame. For a given display frame rate, spatial interlacing and single-frame hybrid can therefore present twice the capture rate, allowing for smoother motion appearance.

From previous work, we expect capture rate and object speed to be the primary determinants of motion artifacts [1,6]. Specifically, whenever the ratio S/R_C (where S is speed and R_C is capture rate) exceeds a critical value, artifacts should become visible. Figure 6 plots the average data in Fig. 5 as a function of capture rate. Plotted this way, the speed at which artifacts became visible is very similar across protocols. The dashed line is $S/R_C = 0.136$, obtained from fitting a straight line to the data. The data show that artifacts are more likely to

be visible whenever that ratio exceeded 0.136. These results are quite similar to those of Hoffman et al. [1] and Johnson et al. [6] who observed critical ratios of 0.1 to 0.2.

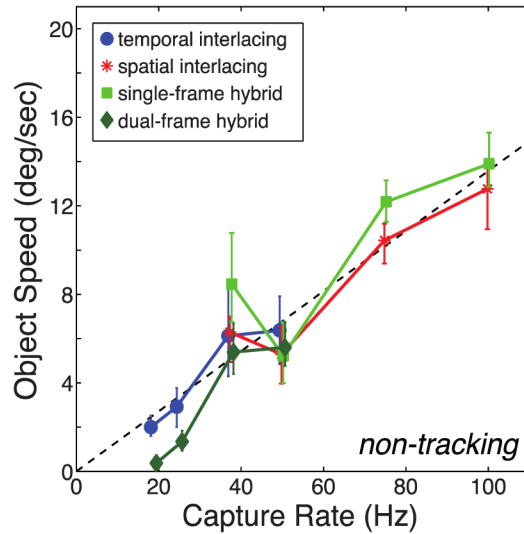


Fig. 6. Capture rate and the visibility of motion artifacts for the non-tracking condition. The speed at which motion artifacts were reported on half the trials is plotted as a function of capture rate. The data are the same as the right panel in Fig. 4, but plotted on a different abscissa. Temporal interlacing and dual-frame hybrid require two display frames per capture period, so the maximum possible capture rate was 50Hz for those protocols. Blue, red, bright green, and dark green represent the results for temporal interlacing, spatial interlacing, single-frame hybrid, and dual-frame hybrid, respectively. Error bars represent 95% confidence intervals. The dashed line represents the equation $S/R_C = 0.136$.

The results from the tracking condition are shown in Fig. 7. As you can see, artifacts became less visible with tracking at the highest capture rates of 75 and 100Hz (that is, higher speeds were required to produce them than in the non-tracking condition) in the spatial interlacing, temporal interlacing, and single-frame hybrid protocols. Observers reported that tracking also changed the type of artifact seen. With tracking, motion blur became the most frequent at higher frame rates; with stationary fixation, most artifacts were judder and edge banding. The visibility of motion blur should depend on how much each stimulus presentation is smeared across the retina: specifically, on how far the stimulus moves across the retina during a presentation. The displacement on the retina is proportional to object speed (because the speed of the eye movement is determined by that speed) and the “on” time of a single presentation (the time the stimulus is illuminated for one eye during the presentation of one content frame): that is, $D = ST$, where D is displacement, S is speed, and T is presentation time. We examined whether the product ST is the determinant of artifact visibility, which would be consistent with the hypothesis that motion blur was the primary artifact in the tracking condition. Figure 8 re-plots the data from Fig. 7 as a function of presentation time in log-log coordinates. Presentation time T is the reciprocal of the display frame rate for the temporal- and spatial-interlacing protocols and the single-frame hybrid protocol. Presentation time is twice the reciprocal of the display frame rate for the dual-frame hybrid protocol because two frames are presented to each eye before the content is updated. In generating Fig. 8, we discarded the data in Fig. 7 for which no speed was found that produced motion artifacts.

Table 1 shows the actual presentation times on the CRT (in parentheses) along with the presentation times being simulated for a sample-and-hold LCD. The shorter times for the CRT still allow a reasonable simulation for an LCD because of temporal filtering by the

visual system. The impulse response of the CRT had a duration of ~ 0.001 sec (1msec). We repeated frames to create longer presentation times; the numbers in those cases are time from the onset of the first impulse to the offset of the last one.

Table 1. Presentation times for the four protocols.

Display frame rate (Hz)	100	75	50	38
CRT frame rate used (Hz)	100	75	100	75
Temporal interlacing (sec)	0.01 (0.001)	0.013 (0.001)	0.02 (0.011)	0.027 (0.014)
Spatial interlacing (sec)	0.01 (0.001)	0.013 (0.001)	0.02 (0.011)	0.027 (0.014)
Dual-frame hybrid (sec)	0.02 (0.011)	0.027 (0.014)	0.04 (0.031)	0.053 (0.041)
Single-frame hybrid (sec)	0.01 (0.001)	0.013 (0.001)	0.02 (0.011)	0.027 (0.014)

If a certain retinal displacement D_c is required to create visible motion blur, the data should be predicted by the equation $S = D_c/T$, which is equivalent to $\log S \propto -\log T$. As you can see, this equation provides a good fit to the data with short presentation times, but not with long presentation times. We believe the poor fit at long presentation times is due to flicker caused by small eye movements becoming visible. To test this, we did a follow-up experiment on three of the six observers. We asked them to indicate after each trial whether they had seen motion blur or not and thus to ignore other artifacts when making their responses. The open circles represent those data, which were indeed well predicted by the equation $S = D_c/T$. This confirms that motion blur is determined by how much a single presentation displaces across the retina.

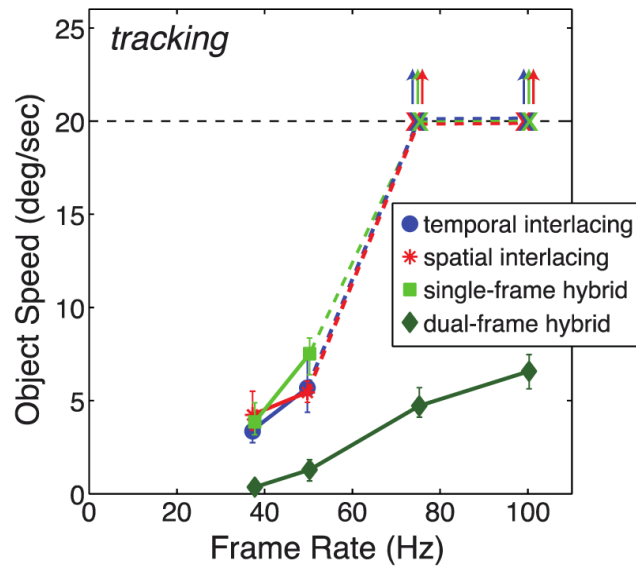


Fig. 7. Visibility of motion artifacts with tracking eye movements. The speed at which motion artifacts were reported on half the trials is plotted as a function of display frame rate. Data have been averaged across subjects. Different colors represent the data from different protocols. Error bars represent 95% confidence intervals. The horizontal dashed line represents the maximum speed tested, so data points (X's) plotted on that line indicate conditions in which no artifacts were reported at any tested speed.

We conclude that motion artifacts in the single-frame hybrid protocol are no more visible than in the spatial-interlacing protocol whether the viewer is tracking the stimulus or not. Therefore, the single-frame version of the hybrid protocol is as unsusceptible to motion artifacts as conventional spatial interlacing and is less susceptible than temporal interlacing.

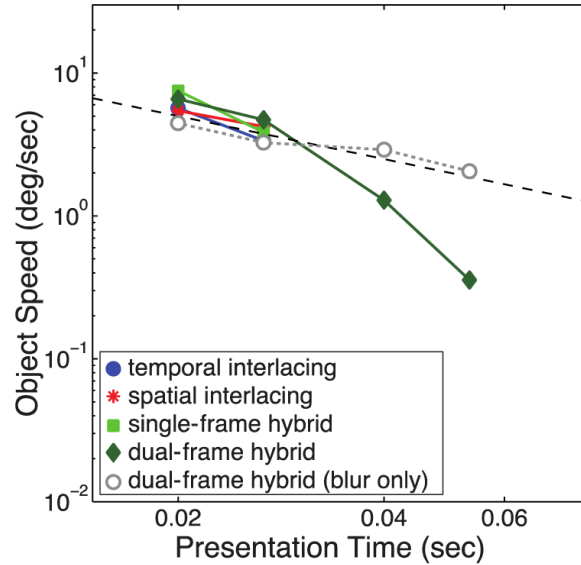


Fig. 8. Motion artifacts and presentation time. The speed at which motion artifacts were reported on half the trials is plotted as a function of presentation time. The solid symbols and lines represent the data from Fig. 7 with the exception of conditions in which no artifacts were reported. The open symbols and dotted line represents data in the follow-up experiment in which observers were asked to report motion blur only. The dashed line represents the prediction that motion blur is seen whenever displacement across the retina exceeds a critical value: $ST > D_c$.

3. Experiment 2: Flicker

Visible flicker is defined as perceived fluctuations in the brightness of a stimulus due to the digital display of the stimulus. Presentation rate has been shown to be the major determinant of flicker visibility [1,6,15]. The threshold for temporally interlaced S3D displays is $\sim 40\text{Hz}$ [1]. The threshold value is well predicted by the amplitude and frequency of the Fourier fundamental of the luminance-varying signal from the display [1,8,16]. The temporal frequency of the Fourier fundamental differs across protocols when the display frame rate is the same. For temporal interlacing, the fundamental frequency is half the display frame rate because the display has to alternate between the two-eyes' views. For the two hybrid protocols, it is also half the display frame rate, but the phase is shifted between the even and odd pixel rows. For spatial interlacing the fundamental frequency is the same as the display frame rate. Duty cycle (the time the stimulus is illuminated in one eye divided by the "on" plus "off" time in the same eye) also plays an important role. Shorter duty cycles create greater amplitudes of the fundamental frequency, so one expects flicker to be more visible as duty cycle is decreased. Temporal-interlacing displays cannot present duty cycles greater than 0.5 because each eye receives a black frame for at least half the time while the display sends light to the other eye. Spatial-interlacing displays, on the other hand, can have duty cycles as great as 1 because content is presented simultaneously to both eyes. In the hybrid protocols, duty cycle is 0.5 for each pixel row. Spatial interlacing should be the least susceptible to flicker because it has a higher fundamental frequency and a longer duty cycle than the other protocols. Temporal interlacing and the two hybrid protocols have the same fundamental frequency and duty cycle, but their spatial characteristics are different. Any image presented using only odd (or even) pixel rows has dark stripes on even (or odd) pixel rows. Thus the spatial frequency associated with the temporal alternation is much higher than with temporal interlacing. Flicker is less visible at high than at low spatial frequencies so flicker should be

much less visible in the hybrid and spatial-interlacing methods than in the temporal-interlacing protocol.

3.1 Subjects

Four subjects, ages 22 to 32 years, participated. All had normal or corrected-to-normal visual acuity and stereo acuity. Two were authors; the others were not aware of the experimental hypotheses.

3.2 Apparatus

The experiments were conducted on one display (ViewSonic G225f) seen binocularly via a mirror stereoscope. We used a single-display stereoscope for this experiment because it allowed us to increase viewing distance to 213cm, which in turn allowed us to make the height of the alternating blocks smaller. SwitchResX was used to control the frame rate and resolution.

3.3 Methods

We determined the display frame rate at which flicker was just visible for each protocol. The stimulus was a stationary bright $1 \times 1^\circ$ square on a dark background. It was presented binocularly with zero disparity for 1sec. Luminance and contrast were equivalent for the tested protocols. The resolution of the display was 1066×800 , yielding a pixel size of 0.6 arcmin. Flicker visibility is strongly dependent on spatial frequency [17]. To determine how the spatial frequency of the alternating blocks affected flicker visibility, we varied block height. The heights h were 0.6, 1.2, 1.8, 2.4, 3.6, 6.0, 8.4, and 14.4 arcmin. These correspond respectively to spatial frequencies per eye of 50, 25, 16.7, 12.5, 8.3, 5, 3.6, and 2.1 cpd. The block height in the RealD/Samsung protocol viewed at the recommended viewing distance for HDTV was 135 arcmin, corresponding to a spatial frequency of 0.2 cpd. The interaction of flicker visibility and spatial frequency is also highly dependent on retinal eccentricity [18]. To examine the effect of eccentricity, we presented stimuli in two retinal positions: on the fovea and 4° below the fovea. The CRT was able to use various frame rates (80, 90, 100, 120, 130, 140, or 150Hz), allowing us to simulate a large set of display frame rates: 6, 8, 10, 20, 30, 40, 50, 60, 80, 90, 100, 120, 140, and 150Hz. After every trial, subjects indicated whether they had seen flicker or not.

The experiment consisted of ~6240 trials per subject: 3 display protocols \times 8 block sizes \times 2 retinal locations \times 13 presentation rates \times 10 trials (130 trials per psychometric function). About 3 hours were required for each subject to complete the experiment. There was only one hybrid protocol because the dual- and single-frame protocols are identical when the stimulus is stationary. We used the method of constant stimuli to vary presentation rate in order to estimate the rate at which flicker was reported on half the trials. As in Experiment 1, we fit a cumulative Gaussian to the resulting psychometric data using a maximum-likelihood criterion and used the 50% point on that function as the estimate of the rate that produced just-visible flicker. All conditions were randomly interleaved. To average across subjects, we pooled the psychometric data from all subjects and then fit those data with one cumulative Gaussian function, as in Experiment 1.

3.4 Results and discussion

Figure 9 plots the display frame rate that produced just-visible flicker as a function of block height. Block height was irrelevant to the temporal- and spatial-interlacing protocols, so there could be only one estimated rate at which flicker was seen for each of those protocols: ~82Hz for temporal interlacing and none for spatial (i.e., flicker was not seen even at the lowest presented rate). The frame rate of 82Hz corresponds to monocular presentation with a fundamental frequency of 41Hz, which agrees with the previously reported flicker thresholds (Hoffman et al., 2011). Thus, as expected, flicker was substantially more visible with

temporal interlacing than with spatial interlacing. The hybrid results are more interesting. With this protocol, blocks are alternately illuminated, so when flicker is perceived, it is at the scale of individual blocks. As we said earlier, sensitivity to high temporal frequencies decreases with increasing spatial frequency [5], and that decrease occurs at lower spatial frequencies in the periphery than in the fovea [18]. For these reasons, we expected that flicker would become more visible as block height increased and that the height producing flicker would be greater in the periphery than in the fovea. This is precisely what we observed. In the fovea, flicker was not seen when block height was 0.6 arcmin and then became more visible as height increased from 1.2 to 2.4 arcmin. In the periphery, flicker was not perceived when block height was 0.6-2.4 arcmin and then became visible at greater heights. Thus, the hybrid protocol is relatively unsusceptible to flicker provided that the size of each alternating block of pixels is smaller than 2 arcmin.

The recommended viewing distance for HD television is 3.1 times the picture height and for UHD is 1.55 times picture height [19–21]. Both displays at the recommended distances yield pixels subtending 1 arcmin. We observed negligible flicker with the hybrid protocol when block height was less than 2 arcmin, so an implementation of this protocol should produce essentially no visible flicker when viewed at the recommended distance or farther.

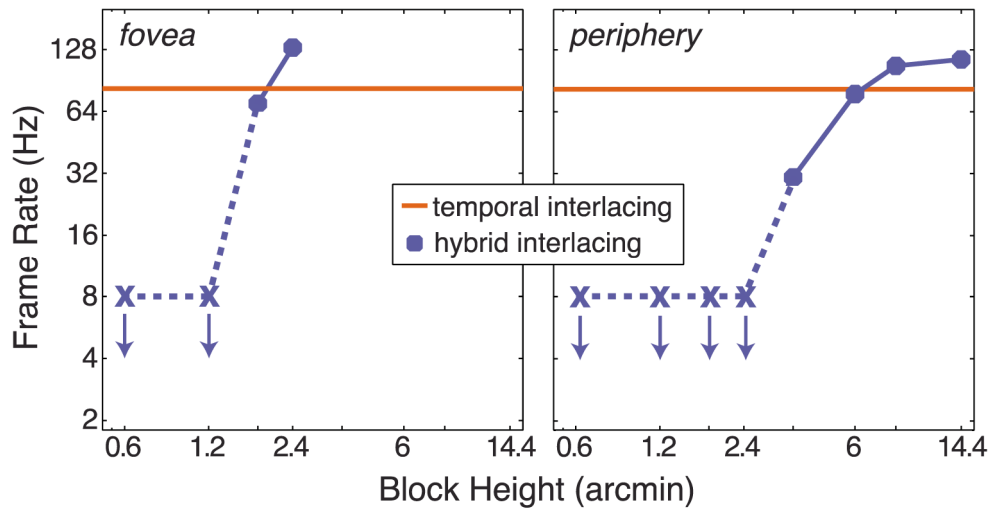


Fig. 9. Flicker visibility for different protocols and block heights. Each panel shows the display frame rate at which flicker was reported on half the trials as a function of the height of the blocks of pixels that were alternated. The left panel shows the data when the stimulus was on the fovea and the right panel the data when the stimulus was 4° below the fovea. Data have been averaged across observers. Orange and blue represent the data with the temporal-interlacing and hybrid protocols, respectively. Flicker was never visible with the spatial-interlacing protocol even at the lowest tested display frame rate of 8 Hz, so we do not plot those data. Flicker was also never visible with the hybrid protocol at the lowest tested display frame rate when blocks were small; those points are represented by X's. Error bars represent 95% confidence intervals; some are too small to be visible.

It is interesting that the hybrid protocol produced more visible flicker than temporal interlacing when the block heights were 2.4 arcmin and larger in the fovea and 6 arcmin and larger in the periphery. We believe this increased visibility is caused by small eye movements, or microsaccades, that dart back and forth across the boundaries between alternating blocks. Davis and colleagues [22] demonstrated that these small movements can cause visible flicker even when the display's alternation rate is very high. They divided the screen of a display with a very high frame rate into left and right halves. The luminance of each half was modulated at very high temporal frequencies, but in opposite phases. Subjects saw flicker at the boundary between the two halves, even when the modulation rate was as

high as 500Hz. Davis and colleagues argued persuasively that the perceived flicker was due to high-frequency horizontal eye movements across the vertical alternation boundary causing different parts of the retina to be exposed to modulation rates that were occasionally much lower than 500Hz. We believe the same effect underlies flicker visibility with the hybrid approach when pixels are sufficiently large.

From our results, we believe that the RealD/Samsung hybrid display created very noticeable flicker because the heights of the alternating blocks were much greater than the heights for which our subjects reported flicker at even high frame rates.

4. Experiment 3: Spatial resolution

If images are presented in every pixel row to an eye, the spatial frequency due to the rows is 30cpd at the recommended viewing distances for HD and UHD-TV [19–21]. At 30cpd, the rows would be barely visible. With spatial interlacing, images are presented in every other row to an eye, so the spatial frequency is 15cpd per eye making the rows more visible monocularly. There are claims that the visual system can fuse two monocular images like those in spatial interlacing to form a binocular image with no missing rows [23,24]. If these binocular-fusion claims are correct, the effective spatial resolution of a spatial-interlacing display would be the same as a temporal-interlacing display that presents all rows to each eye. If these claims are incorrect, however, one would have to double the viewing distance with spatial interlacing to make the rows roughly equally visible compared to temporal interlacing. We measured the spatial resolution of different protocols to see if effective resolution is indeed reduced in spatial interlacing and to see if the hybrid protocols provide greater effective resolution than spatial interlacing.

Kim and Banks [2] measured effective spatial resolution with spatial and temporal interlacing at different viewing distances. They found that viewers' ability to discern fine detail was reduced with spatial interlacing provided that the viewing distance was not too great. They observed the resolution difference with both monocular and binocular viewing suggesting that the binocular-fusion claim is incorrect. Hybrid interlacing should provide greater effective resolution than spatial interlacing because each eye receives a full-resolution image, albeit over the course of two frames. If no object motion is present, the visual system can average over the two frames to gain high resolution. If motion occurs, however, the gain in resolution may depend on the direction and speed of motion.

The effective spatial resolution of any display is affected by viewing distance. At long distances, where pixels are too small to be resolved by the visual system, resolution becomes "eye limited" [2]. The size and arrangement of pixels will therefore not matter to the viewer's ability to see fine detail. At shorter distances, where pixels can be resolved by the visual system, resolution becomes "display limited" [2] and then the size and arrangement of pixels affects the viewer's ability to see fine detail.

We examined the effective spatial resolution of the four protocols illustrated in Figs. 1 and 2 and also determined how vertical and horizontal motion influences the outcome.

4.1 Methods & apparatus

The same six subjects participated as in Experiment 1. The one-display stereoscope from Experiment 2 was used in order to enable a long viewing distance of 213cm. Recall that a pixel subtends 1arcmin at the recommended viewing distances for HD- and UHD-TV. We determined the effective spatial resolution of the four display protocols with a "tumbling E" task. In this task, observers report which of four orientations of the letter E was presented (Fig. 10). The size of the letter was varied to find the just-identifiable size. The letter was black on an otherwise white background. The stimuli were presented stereoscopically with a disparity of 0. Display resolution was 1280 × 960 at 213cm yields a pixel size of 0.5 arcmin.

We simulated changes in viewing distance by simulating pixels of different sizes—one pixel for a simulated pixel size of 0.5 arcmin, 2 × 2 pixels for 1 arcmin, and 4 × 4 for 2

arcmin—and having subjects view from a fixed distance of 213cm. We did this instead of actually changing viewing distance so that we could randomly interleave all experimental conditions. We verified that our method for simulating the effect of viewing distance was valid by conducting a control experiment. In the control experiment, we used the same Tumbling-E task to measure letter-acuity thresholds in three participants for the temporal-interlacing, spatial-interlacing, and hybrid protocols. The letters did not move so the two versions of the hybrid protocol were identical. We measured acuity when the viewing distance was actually varied (distances of 213, 106, and 53cm; 1×1 pixels) and when changes in distance were simulated (distance was fixed at 213cm, and pixels were 1×1 , 2×2 , and 4×4). The measured acuities did not differ systematically (expressed in angular units), which shows that our method of simulating different viewing distances was valid. The control experiment was done with the single-display setup.

Frame rate was 120Hz. We presented static and moving stimuli; motion was vertical or horizontal at a speed of $3^\circ/\text{sec}$. On the motion trials, a fixation cross was presented eccentrically 0.5 sec before stimulus onset in order to inform the subject of the direction of the upcoming movement. The stimulus then crossed screen center and the viewer tracked it with their eyes. Stimuli were presented for 0.6 sec and the viewer responded up, down, left, or right to indicate the perceived orientation of the letter. The task was therefore a 4-alternative, forced-choice task. No feedback about the correctness of each response was provided.

The experiment consisted of 10,368 trials per subject: 4 display protocols \times 3 pixel sizes \times 3 movements \times 4 orientations \times 9 letter sizes \times 8 trials. About 3 hours were required for each subject to complete the experiment. We used the method of constant stimuli to vary letter size. We fit the resulting psychometric data with a cumulative Gaussian using a maximum-likelihood criterion. The lower asymptote was located at 25%, and the acuity threshold for each condition was defined as the letter size for which orientation was identified correctly on 62.5% of the trials. A total of 288 trials went into each fit: 864 when we combined across motion direction. To average across subjects, we pooled the psychometric data from all subjects and then fit those data with one cumulative Gaussian function, as in Experiments 1 and 2.

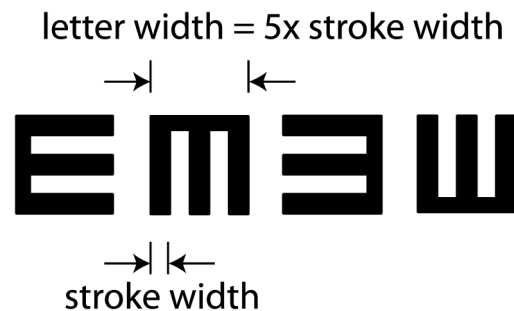


Fig. 10. Stimuli used to measure spatial resolution. The height and width of the letter was always five times the stroke width. Thus, when letter size was manipulated, the stroke width changed as well as the letter height and width. On each trial, subjects indicated which of the four orientations was presented.

4.2 Results and discussion

The results are shown in Fig. 11. The six panels on the left show the data from the individual observers. The data have been averaged across the three motion conditions. The panel on the right shows those data averaged across observers. The horizontal and diagonal dashed lines represent the expected resolution thresholds for eye-limited and display-limited conditions, respectively. As you can see, thresholds increased as the simulated pixels became larger (i.e., as the simulated viewing distance became shorter) following the eye-limited and display-

limited predictions fairly well. Importantly, resolution differed across protocols. Clearly, spatial interlacing had poorer effective resolution than temporal or hybrid interlacing in the display-limited regime (i.e., where the pixels were 1-2 arcmin). With smaller pixel sizes in the eye-limited regime (0.5 arcmin), the four protocols had very similar effective resolutions.

We next examined the influence of motion on effective spatial resolution. Figure 12 shows the data, averaged across observers, for stationary, horizontally moving, and vertically moving stimuli. With no motion, resolution with spatial interlacing was poorer than with the temporal-interlacing or the hybrid protocols at the shorter simulated viewing distances. With motion present, resolution with the spatial-interlacing and the dual-frame hybrid protocols was quite dependent on the direction of motion. When it was horizontal, resolution with those two protocols was worse than resolution with the temporal-interlacing and single-frame hybrid protocols. When motion was vertical, however, resolution with the spatial-interlacing and dual-frame hybrid protocols improved substantially. The improvement with vertical motion and the lack of improvement with horizontal motion both make sense. The problem with the spatial-interlacing and dual-frame hybrid protocols is that potentially useful data are not shown to a given eye in every presentation. By moving the stimulus vertically, all parts of the letter can be presented to each eye over time, so performance improves. When the stimulus moves horizontally, the missing data are not presented at any time, so performance does not improve. The use of motion to create higher effective resolution has been examined extensively in computer graphics [25].

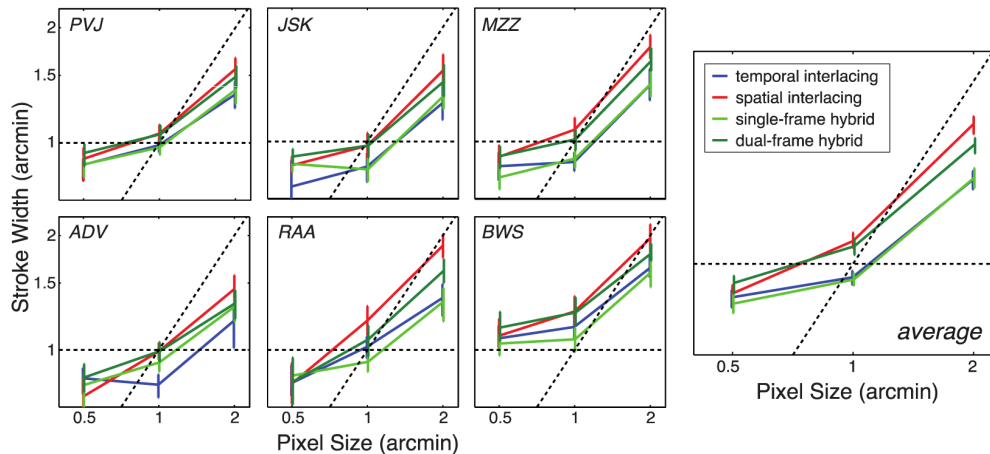


Fig. 11. Spatial resolution for different protocols and simulated viewing distances. The six panels on the left plot the data from individual observers, averaged across the three motion conditions. Each panel plots the letter stroke width for which the observer identified letter orientation correctly on 62.5% of the trials. Different colors represent the data from different protocols. Error bars represent 95% confidence intervals. The horizontal and diagonal dashed lines represent the expected values for eye-limited and display-limited acuities, respectively, on a conventional 2D display. The right panel shows the data averaged across subjects.

We conclude that the single-frame version of the hybrid protocol has better spatial resolution than the spatial-interlacing protocol. Indeed, the effective resolution of the proposed protocol is on par with temporal interlacing.

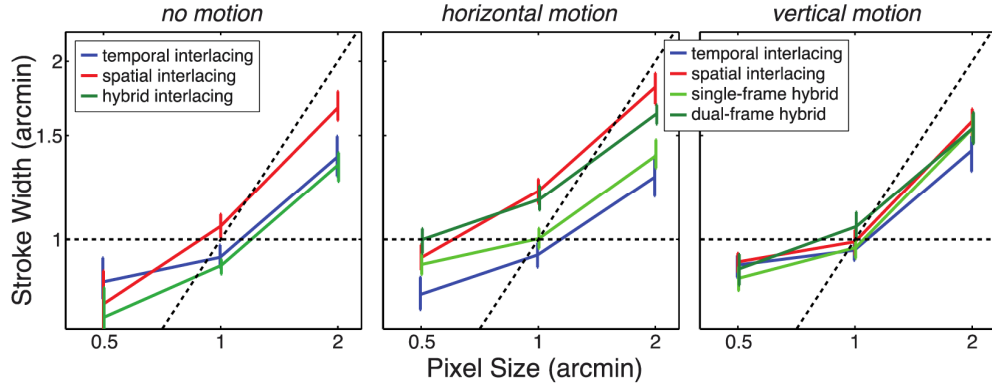


Fig. 12. Effect of motion on spatial resolution. Data are averaged across subjects. The left, middle, and right panels show the data for no motion, horizontal motion, and vertical motion, respectively. Different colors represent the data from different protocols. Error bars represent 95% confidence intervals. The horizontal dashed line represents the expected resolution thresholds in the eye-limited regime and the diagonal dashed line the expected thresholds in the display-limited regime.

5. Experiment 4: Depth distortion

In temporal-interlacing S3D displays, the left- and right-eye views are presented in alternation. This means that the second eye sees an image later than the first eye even though the two eyes' contents were captured at the same time. When there is movement in the scene, the visual system interprets the temporal lag as a spatial disparity and perceived depth becomes distorted [1,26,27]. Consider an object moving horizontally and presented on a temporal-interlacing display (left panel, Fig. 13). The position of the object is captured with left and right cameras simultaneously at the times marked by black arrows. When the images are presented in alternation, the right one is delayed. The visual system has to match left- and right-eye images to compute disparity, but it is unclear how to make the matches because none occur at the same time. If a given left-eye image were matched with the subsequent right-eye image (green arrow in left panel of Fig. 13), the estimated spatial disparity would be correct (green dots in right panel). But if that same left-eye image were matched with the preceding right-eye image (purple arrow in left panel), the estimated disparity would be incorrect (purple dots in right panel). The brain has no way to know which match is correct because they both have the same inter-ocular time difference. The most reasonable strategy then is to average the two estimates creating a disparity estimate halfway in-between [1,27,28]. The induced spatial disparity is:

$$\Delta = s\tau \quad (1)$$

where s is object speed and τ is the inter-ocular offset of successive presentations. At most frame rates, the viewer does in fact perceive the moving object at a depth consistent with the average disparity estimate [1,27]. For a rightward-moving stimulus with the left-eye image presented before the right-eye image (as in the left panel of Fig. 13), the time-average estimate is shifted toward crossed (near) disparity, so the object is perceived as closer than intended. For a leftward-moving stimulus, the time-average estimate is shifted toward uncrossed (far) disparity, so the object is seen as farther than intended. This type of depth distortion should not occur with spatial-interlacing and hybrid displays because they present content simultaneously to the two eyes, yielding no ambiguity about which image in the right eye to match with a given image in the left eye.

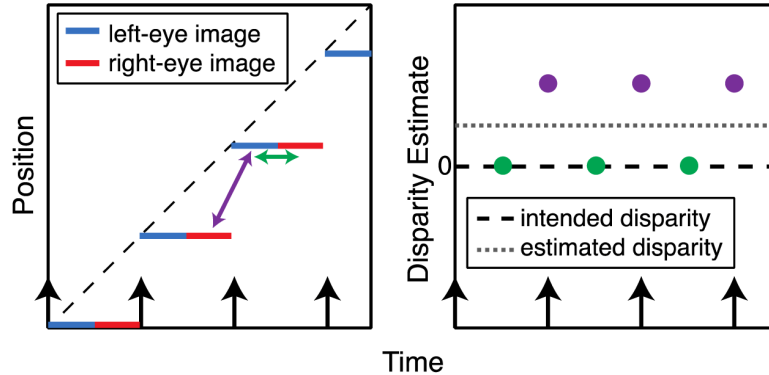


Fig. 13. Depth distortion due to temporal interlacing. Left: Temporal-interlacing S3D presentation of an object moving at constant speed. Left- and right-eye images are captured simultaneously, but displayed in alternation. A given left-eye image could be matched binocularly with a subsequent right-eye image (green arrow) or a preceding right-eye image (purple arrow). Right: Disparity estimates due to different left- and right-eye matches. The matches between left-eye images and subsequent right-eye images yield correct disparity estimates (green dots), but the matches between left-eye images and preceding right-eye images yield incorrect estimates (purple dots). Averaging the correct and incorrect disparities yields an estimate halfway between the two (dotted line, which is given by Eq. (1)).

Another type of depth distortion occurs in spatial-interlacing displays when the viewing distance is short enough for the pixel rows to be resolved monocularly. Because the left- and right-eye views are offset vertically by one pixel, the eyes make a vertical vergence eye movement to binocularly fuse the rows (bright rows aligned in the two eyes and dark rows also aligned). The vertical eye movement causes a change in the horizontal disparity at the retinas for off-vertical and off-horizontal edges, so those edges appear at unintended depths [3]. Interestingly, some spatial-interlacing displays eliminate this effect by presenting data rows alternately [23,24]. Odd rows on the display are seen by one eye and even rows by the other. But the data presented to odd rows alternate between odd and even and the data presented to even rows alternate between even and odd. The alternation rate is sufficiently high for the alternating images to be temporally averaged by the visual system. This vertical-averaging algorithm eliminates the depth distortion [3], but at the cost of reduced spatial resolution.

The proposed hybrid technique alternates the delivery of even and odd rows to the two eyes, so there is no consistent stimulus to drive vertical vergence to an unintended value. Thus, depth distortions due to the vertical offsets in spatial-interlacing displays should not occur with this technique. We did not test this possibility because Hakala et al. [3] have already shown that this type of distortion occurs with spatial interlacing and there is no reason to believe that it should occur with the hybrid protocols proposed here.

5.1 Methods & apparatus

The same subjects and apparatus were used as in Experiment 1. The frame rate was 100Hz and the resolution of each display was 1200×900 . We measured the perceived depth of moving objects in all four display protocols. We did this by presenting two sets of bright 1° rectangles moving horizontally in opposite directions against a dark background (Fig. 14). On half the trials, the upper rectangles moved leftward and the lower ones rightward and on the other half of the trials, the motion directions were the opposite. Speed varied from -20 to $20^\circ/\text{sec}$. The vertical edges of the rectangles were blurred slightly to reduce the salience of motion artifacts; this made the task easier to perform. Subjects were not instructed about fixation. After each trial, they indicated whether the top or bottom rectangles appeared closer (yes-no, single-presentation task). Based on the response, spatial disparity was added to the

stimulus for the next trial according to a 1-down/1-up staircase procedure. Disparity was added to both the top and bottom rectangles, but with opposite sign. The goal was to find the disparity that had to be added to the moving stimulus in order to eliminate the depth distortion (i.e., make the top and bottom rectangles appear to be at the same depth). We call this added disparity the *nulling disparity*; it is a quantitative measure of the direction and size of the depth distortion.

The experiment consisted of about 320 trials per subject: 4 display protocols \times 4 speeds \times 20 trials. About 1 hour was required for each subject to complete the experiment. We determined psychometric functions using the method we described earlier.

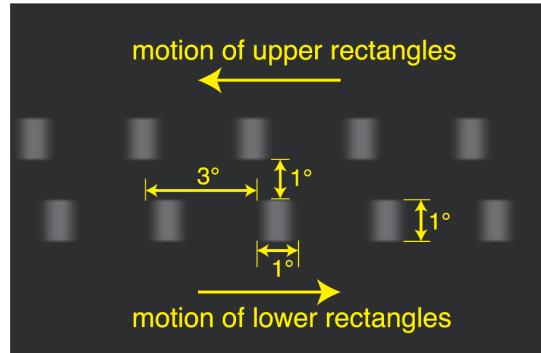


Fig. 14. The stimulus used to measure depth distortion. Two groups of rectangles moved horizontally in opposite directions. Sometimes the upper rectangles moved to the left and the lower rectangles moved to the right, as shown above, and sometimes the direction of motion was reversed. On a given trial, the upper or lower group may have appeared closer than the other.

5.2 Results and discussion

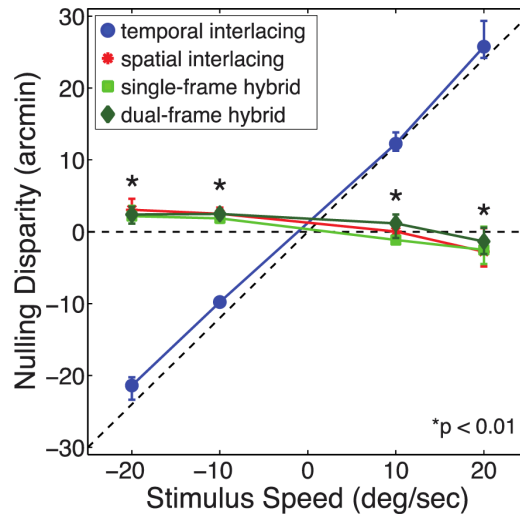


Fig. 15. Depth distortions for different protocols. The nulling spatial disparity is plotted as a function of speed. Different colors represent the results from the different protocols. The data have been averaged across subjects. The diagonal dashed line represents the predictions of Eq. (1) once multiplied by two because there were always two distortions in opposite directions, one for the rightward-moving stimulus group and one for the leftward-moving group. Error bars denote 95% confidence intervals. Asterisks indicate speeds at which the spatial-interlacing and hybrid protocols yielded significantly less distortion than the temporal-interlacing protocol (paired t-tests, $p < 0.01$).

6. Discussion

We found that the single-frame hybrid protocol maintains the benefits of both temporal and spatial interlacing, while eliminating the drawbacks. Specifically, motion appearance and flicker were substantially better than with temporal interlacing, depth distortion was eliminated, and spatial resolution was better than with spatial interlacing. Thus, spatiotemporal interlacing is an attractive solution for presenting stereoscopic content with minimal temporal and spatial artifacts. We next discuss the underlying causes of the effects we observed with the different protocols and how one might implement the single-frame hybrid technique.

6.1 Sampling, display, and viewing pipelines for different protocols

We have sufficient understanding of spatial and temporal filtering in the human visual system to make rigorous predictions about how different protocols, frame rates, duty cycles, and pixel sizes ought to affect flicker visibility, motion artifacts, and effective spatial resolution on a display. To this end, we modeled the pipeline from stimulus to display to viewing for the four protocols: temporal interlacing, spatial interlacing, and the two hybrid techniques.

The display of video content involves three dimensions (two in space and one in time), but we show the analysis for two dimensions only (one in space and one in time) for ease of visualization. Typically image data $i(x,t)$ are anti-aliased before being sent to the display, so we anti-aliased by convolving with a cubic interpolation function, $a(x,t)$. We then simulated how intensity varies over space and time when the image data are presented on a digital display. We sampled the anti-aliased image data with a comb function representing the spatiotemporal sampling of the display, where the samples are separated spatially by x_0 (pixel spacing) and temporally by t_0 (display frame time). The displayed intensities have finite spatial and temporal extent, which we represent with a spatiotemporal aperture function $p(x,t)$. The double asterisk represents two-dimensional convolution. In this example, the pixel fill factor is assumed to be 1 (meaning that the pixel width is equal to the inter-pixel separation), but the fill factor could have other values:

$$\left[[i(x,t) ** a(x,t)] s(x,t) \right] ** p(x,y) \quad (2)$$

$$\left[[i(x,t) ** a(x,t)] \text{comb} \left(\frac{x}{x_0}, \frac{t}{t_0} \right) \right] ** \text{rect} \left(\frac{x}{x_0}, \frac{t}{t_0} \right), \quad (3)$$

where rect is a scaled rectangle function with widths x_0 in space and t_0 in time, and x_0 and t_0 also represent the spatial and temporal separations of samples in the comb function. In the Fourier domain, the second equation becomes:

$$\left[[I(f_x, f_t) A(f_x, f_t)] ** \text{comb}(x_0 f_x, t_0 f_t) \right] \text{sinc}(x_0 f_x, t_0 f_t), \quad (4)$$

where f_x and f_t are spatial and temporal frequency, respectively, and the sinc function has zeros at $f_x = 1/x_0, 2/x_0$, etc. and at $f_t = 1/t_0, 2/t_0$, etc. In the hybrid protocols, the sampling function has a phase shift in x at the alternation rate. In the single-frame hybrid protocol, there is also a phase shift in time. With these phase shifts, there are different spatiotemporal sampling functions. For the single-frame hybrid protocol, it is:

$$\begin{aligned} & \text{comb} \left(\frac{x}{2x_0}, \frac{t}{2t_0} \right) \text{ for odd pixel rows,} \\ & \text{comb} \left(\frac{x+x_0}{2x_0}, \frac{t+t_0}{2t_0} \right) \text{ for even pixel rows,} \end{aligned} \quad (5)$$

and they alternate at the alternation rate. For the dual-frame hybrid protocol, the sampling function is:

$$\begin{aligned} & \text{comb}\left(\frac{x}{2x_0}, \frac{t}{2t_0}\right) \text{ for odd pixel rows,} \\ & \text{comb}\left(\frac{x+x_0}{2x_0}, \frac{t}{2t_0}\right) \text{ for even pixel rows.} \end{aligned} \quad (6)$$

Because one set of pixel rows is presented with a delay of one display frame, we need separate spatiotemporal aperture functions for the odd and even pixel rows:

$$\begin{aligned} & p\left(\frac{x}{x_0}, \frac{t}{t_0}\right) \text{ for odd pixel rows,} \\ & p\left(\frac{x}{x_0}, \frac{t-t_0}{t_0}\right) \text{ for even pixel rows.} \end{aligned} \quad (7)$$

In the simulations shown here, we assumed an illumination time equal to the display frame time and a pixel width and height equal to the pixel separation for the aperture functions. Other values could of course be assumed. We convolve the anti-aliased input with each of these sampling functions separately and then sum them to obtain an amplitude spectrum associated with each protocol. In the simulations shown here, we also assumed a display frame rate of 60Hz and a pixel size of 1 arcmin (because that corresponds to the recommendation for HD- and UHD-TV). Object speed was 1.08°/sec (65 pixels/sec). Other values could of course be assumed. We computed the amplitude spectra for one eye's image only because flicker and motion artifacts are determined primarily by monocular processing [1] and effective spatial resolution is also determined primarily by monocular processing [2].

The sequence of computations for a stimulus moving vertically across the screen at constant speed is shown in Fig. 16. (We replace x with y in the figure to remind the reader that the motion is vertical.) The first row is a space-time plot of the stimulus and convolution with the anti-aliasing kernel. The output of the convolution is the second row and is the data sent to the screen to be displayed. The third row shows the sampling functions associated with the four protocols: from left to right, temporal interlacing, spatial interlacing, dual-frame hybrid, and single-frame hybrid. The fourth row represents the outputs of the sampling in the four protocols. Those outputs are convolved with the spatiotemporal function displayed in the fifth row to produce the space-time sequences of finite pixels at finite time intervals shown in the sixth row. Note that dual-frame hybrid delays the captured information of even rows while single-frame hybrid already has the delay in the captured information and hence does not delay it in this step. Those sequences are subjected to Fourier transformation and the resulting spatiotemporal amplitude spectra are shown in the bottom row.

Figure 17 provides larger versions of the amplitude spectra for the four protocols. The spectra consist of a filtered version of the original signal (diagonal line through the origin) as well as spatiotemporal aliases. When aliases near the temporal-frequency axis are visible, viewers should see flicker. When aliases in other locations in frequency space are visible, viewers typically see judder. The human visual system is sensitive to only a small range of the spatiotemporal frequencies generated by digital displays. The sensitivity range is quantified by the spatiotemporal contrast sensitivity function (also called the *window of visibility* [8]). The window of visibility is represented by the orange diamonds in each panel of Fig. 17; the diamond shape is a reasonable approximation to the actual sensitivity function [17]. When the aliases fall within the visible range, flicker and motion artifacts should be visible. When they fall outside the visible range, the stimulus should appear unflickering and its motion should appear smooth. The advantages of hybrid interlacing are readily apparent in

the amplitude spectra. In particular, the single-frame hybrid should be relatively unsusceptible to flicker and motion artifacts because the aliases occur at higher frequencies than with the other protocols. This of course is what we observed experimentally.

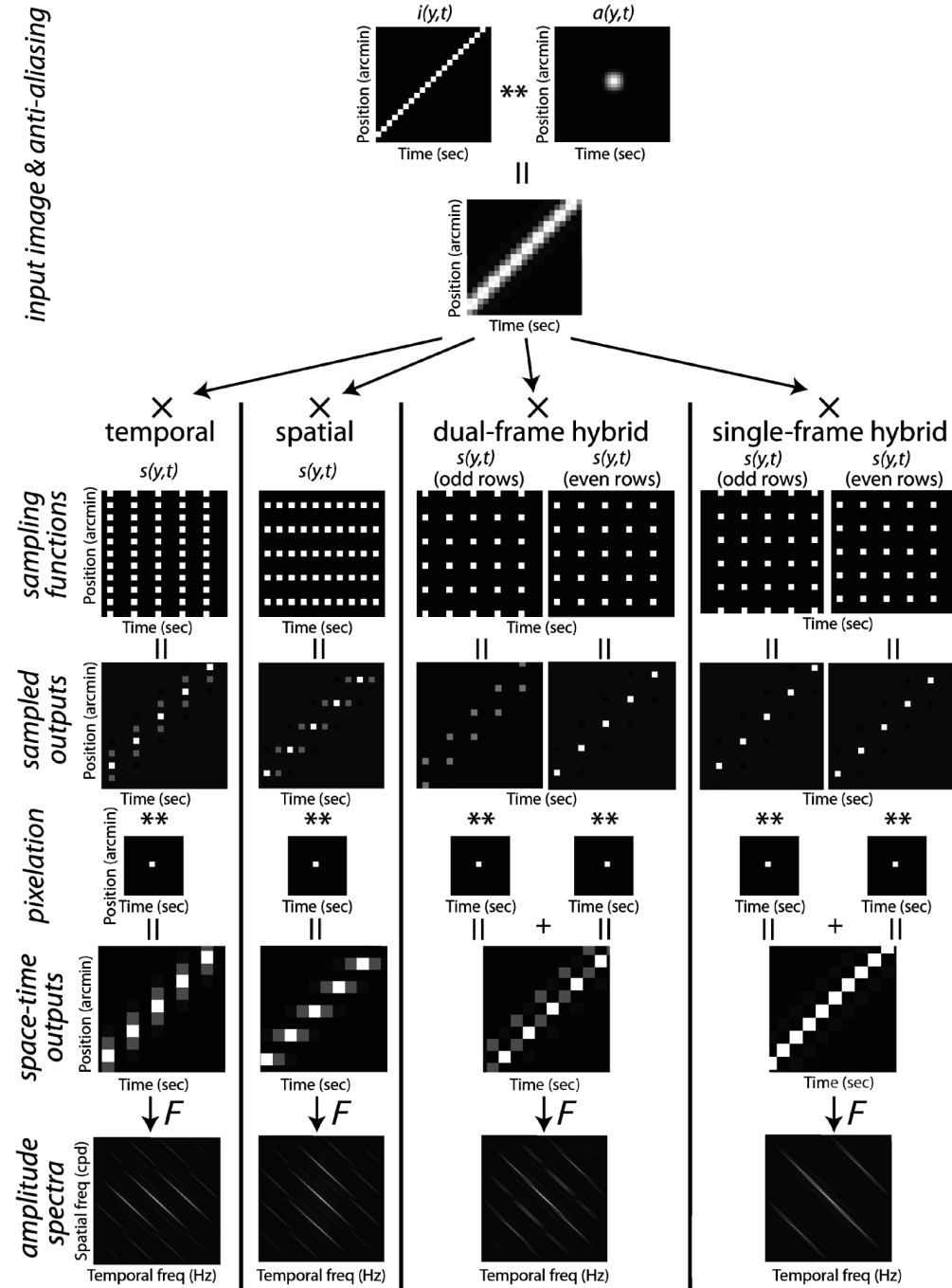


Fig. 16. The presentation of a moving stimulus on the four protocols in space-time and in the Fourier domain. See text for details.

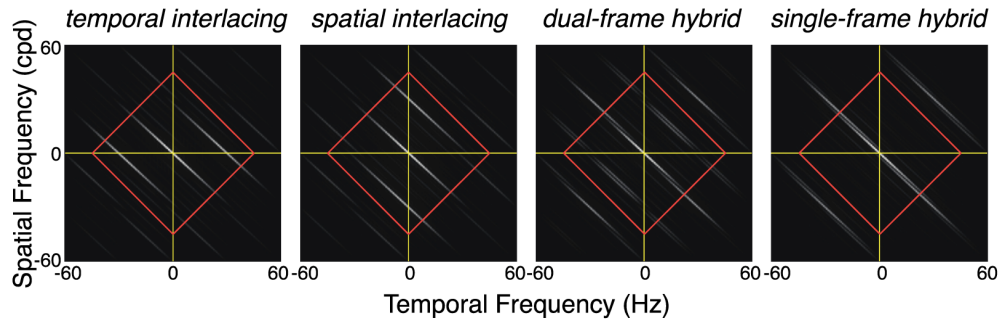


Fig. 17. Amplitude spectra for a stimulus moving vertically in the four interlacing protocols. These panels are based on magnified versions of the bottom row in Fig. 16. The spectra contain a filtered version of the original signal (diagonal line intersecting the origin) as well as aliases due to sampling. The orange diamonds represent the window of visibility, the range of spatial and temporal frequencies that are visible to a typical human observers. Aliases within the window of visibility can cause visible artifacts. In the case of temporal interlacing, aliases have large amplitudes in the temporal frequency direction, indicating the possibility of temporal artifacts, (e.g., flicker and motion artifacts). In spatial interlacing, aliases have large amplitudes in the spatial frequency direction, indicating a possible loss of spatial resolution. The single-frame hybrid has no aliases within the window of visibility, suggesting that aliases will not be visible.

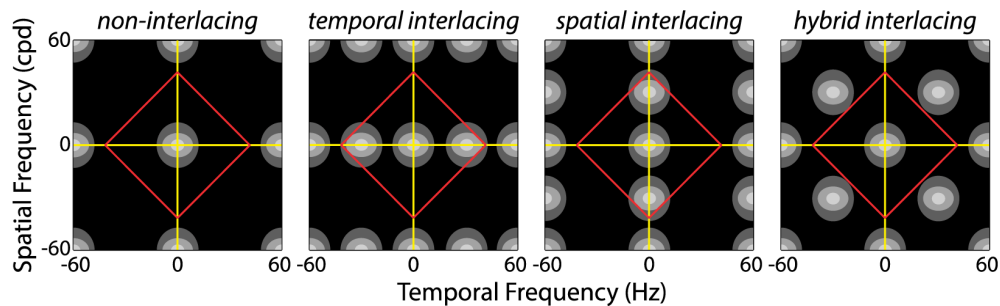


Fig. 18. Amplitude spectra for different interlacing protocols. The display frame rate is 60Hz and pixel size is 1arcmin. The orange diamonds represent the window of visibility. The leftmost panel shows the signal presented on a non-interlacing display. The central pattern is the original signal, and its aliases repeat with a period of 60Hz in temporal frequency and 60cpd in spatial frequency. When temporal interlacing is used, the aliases occur at lower temporal frequencies (multiples of 30Hz). Similarly, when spatial interlacing is used, aliases occur at a lower spatial frequencies (multiples of 30cpd). Hybrid interlacing produces aliases at multiples of 30Hz in temporal frequency and 30cpd in spatial frequency, but they are located diagonally in frequency space. As a consequence, the aliases in hybrid interlacing are farther from the window of visibility, making them much less visible. In this cartoon, we omitted the effect of pixilation for simplicity.

This analysis of the pipeline also helps one understand the determinants of effective spatial resolution with different protocols and display parameters. In Fig. 18, we present a hypothetical spatiotemporal stimulus with a low-pass amplitude spectrum; such a spectrum (represented by concentric circles at (0,0)) is characteristic of most natural images. If the stimulus is presented on a non-interlacing display (first panel), aliases appear at every 60Hz in temporal frequency and 60cpd in spatial frequency. Temporal interlacing (second panel) loses half the frames for each eye's view, resulting in additional aliases at temporal frequencies of -30 and 30 Hz. Spatial interlacing (third panel) drops half the pixel rows in each eye's view, yielding additional aliases at spatial frequencies of -30 and 30 cpd. The single-frame hybrid protocol (fourth panel) produces aliases that they are located diagonally in frequency space. This is because the sampling function in the space-time domain is a set of

impulse functions positioned diagonally. Thus, the aliases created by hybrid interlacing are farther (by a factor of $\sqrt{2}$) from the origin. The orange diamond in each panel represents the window of visibility. Because of its diamond shape the aliases in the single-frame hybrid protocol are even less likely to be visible than in temporal and spatial interlacing. Thus, higher spatial frequencies can be seen by the viewer without intrusion by the aliases created in spatial interlacing. The prediction that the hybrid protocol should have higher spatial resolution than spatial interlacing was, of course, confirmed by our experimental measurements.

6.2 Implementation of spatiotemporal interlacing

There are at least two ways to implement the hybrid protocol in a stereoscopic display. The first requires that the viewer have active eyewear that alternates left- and right-eye views; the second involves active polarization switching at the display and thereby allows one to use passive eyewear.

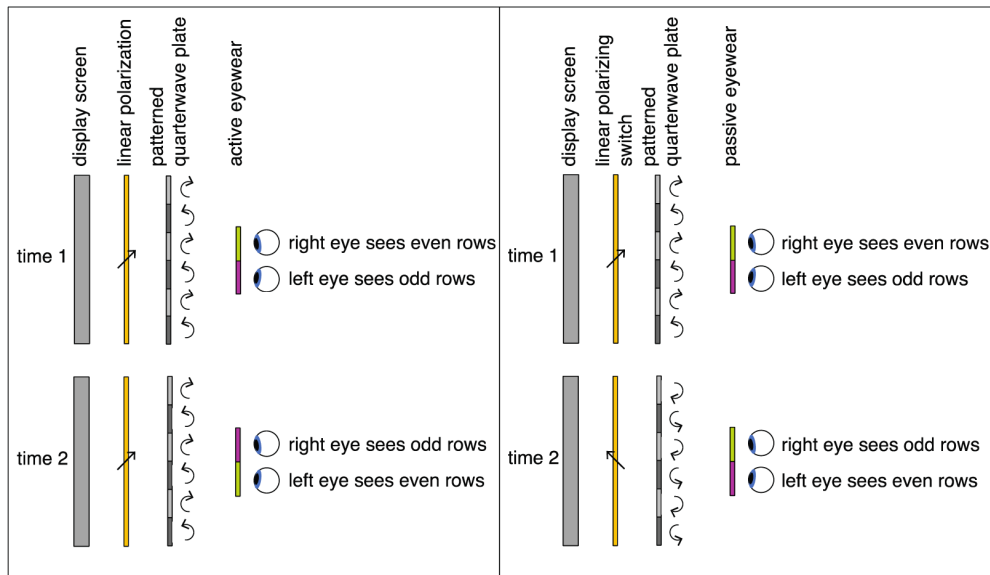


Fig. 19. Two ways to implement the hybrid protocol. The schematic on the left shows how one can implement the protocol with active eyewear. The display delivers light to a passive linear polarizer, which delivers linearly polarized light to a patterned quarterwave plate. Half of the elements in the patterned plate are oriented vertically and the other half horizontally. As a consequence, the light transmitted through half of the elements is polarized clockwise and the light transmitted through the other half is polarized counter-clockwise. The eyewear lens in front of the right eye transmits light polarized counter-clockwise and the lens in front of the left eye transmits light polarized clockwise. With each frame, the state of the eyewear lenses is reversed in synchrony with the images presented on the display. The schematic on the right shows how one can implement the hybrid protocol with passive eyewear. The display delivers light to an active linear polarizer, which in turn delivers linearly polarized light to the patterned quarterwave plate. Again half of the plate elements are oriented vertically and the other half horizontally. Thus, light transmitted through the first half is polarized clockwise and light transmitted through the other half is polarized counter-clockwise. The eyewear lens in front of the right eye always transmits light polarized clockwise and the lens in front of the left eye always transmits light polarized counter-clockwise. On each frame, the linear polarizer switches to transmit light polarized at the orthogonal angle relative to the previous frame. With that switch, the polarization direction of each element in the quarterwave plate reverses.

The first implementation is schematized in the left panel of Fig. 19. The display sends light through a linear polarizing stage (yellow), which then transmits to a patterned quarterwave plate (gray). The quarterwave plate yields circular polarization that is clockwise

in half the elements and counter-clockwise in the other half. The viewer wears active eyewear that alternates between the modes: one in which the clockwise elements are seen by the left eye and the counter-clockwise elements by the right eye (time 1) and one in which the clockwise elements are seen by the right eye and the counter-clockwise elements by the left eye (time 2).

It is undesirable in many applications to have active eyewear, so we designed an implementation that uses passive eyewear. This implementation is schematized in the right panel of Fig. 19. The display sends light through a linear polarizing stage that switches between polarization angles of $+45^\circ$ and -45° . When the linear stage is at $+45^\circ$, the patterned quarterwave plate produces clockwise polarization in the odd rows and counter-clockwise in the even rows. When the linear stage is at -45° , the quarterwave plate produces clockwise polarization in the even rows and counter-clockwise in the odd rows. The passive eyewear transmits clockwise polarization to the left eye and counter-clockwise to the right eye.

Both implementations would yield spatiotemporal interlacing as we have simulated in the experimental work presented here. The design of the quarterwave plate (e.g., row by row, checkerboard, etc.) determines the spatial pattern of the alternating blocks on the display. An interesting research question is what spatial pattern is most effective perceptually: a row-by-row pattern as we emulated here, a checkerboard pattern, or something else?

7. Conclusion

In this paper, we propose an S3D presentation technique with spatiotemporal interlacing. Our psychophysical experiments demonstrate that spatiotemporal hybrid interlacing maintains the better properties of both spatial and temporal interlacing. The hybrid technique has better spatial properties than spatial interlacing and better temporal properties than temporal interlacing. We developed a computational model that illustrates how different protocols ought to affect flicker, motion artifacts, and spatial resolution. The results from the model are consistent with the experimental results. We also provided a description of how this display might be implemented using currently available technology. This display technique should provide a better viewing experience than existing methods.

Acknowledgment

This research was supported by NIH research grant R01EY012851.