UC San Diego
UC San Diego Previously Published Works

Title

Apparent Sharpness of 3D Video When One Eye's View is More Blurry

Permalink

https://escholarship.org/uc/item/5bb8h1hp

Journal

i-Perception, 4(6)

ISSN

2041-6695

Authors

Robinson, Alan

Jain, Ankit

Scott, Mathew

et al.

Publication Date

2013-09-01

DOI

10.1068/i0570

Copyright Information

Peer reviewed

Apparent sharpness of 3D video when one eye's view is more blurry

Alan Robinson
Departments of Psychology and Cognitive Science, UCSD, 9500 Gilman Drive, La Jolla, CA, USA;
e-mail: aerobins@cogsci.ucsd.edu

Ankit Jain
Department of Electrical and Computer Engineering, UCSD, 9500 Gilman Drive, La Jolla, CA, USA;
e-mail: ankitkj@ucsd.edu

Mathew Scott
Department of Psychology, UCSD, 9500 Gilman Drive, La Jolla, CA, USA; e-mail: kaliandloki@netzero.com

Don MacLeod
Department of Psychology, UCSD, 9500 Gilman Drive, La Jolla, CA, USA; e-mail: dmacleod@ucsd.edu

Truong Nguyen
Department of Electrical and Computer Engineering, UCSD, 9500 Gilman Drive, La Jolla, CA, USA;
e-mail: nguyent@ece.ucsd.edu

Abstract. When the images presented to each eye differ in sharpness, the fused percept remains relatively sharp. Here, we measure this effect by showing stereoscopic videos that have been blurred for one eye, or both eyes, and psychophysically determining when they appear equally sharp. For a range of blur magnitudes, the fused percept always appeared significantly sharper than the blurrier view. From these data, we investigate to what extent discarding high spatial frequencies from just one eye's view reduces the bandwidth necessary to transmit perceptually sharp 3D content. We conclude that relatively high-resolution video transmission has the most potential benefit from this method.

Keywords: stereo vision, blur, 3D video, mixed resolution, bandwidth.

1 Introduction

Julesz ([1971](#)) first observed that the input to each eye need not be equally sharp to produce a pleasing fused percept, suggesting that the perceived power spectrum is not simply the average of the spectrum for each eye. A quite blurry view in one eye and a sharp view in the other lead to a surprisingly sharp fused percept, with well-defined stereoscopic depth, as demonstrated in [Figure 1](#). This suggests that the lower spatial frequencies which are seen by both eyes can be used to extract disparity, and that the high spatial frequencies presented to one eye are integrated into the fused percept so that they appear in the proper 3D spatial location.

In this paper, we explore the application of this effect to stereoscopic video. Stereoscopic depth is produced by showing different views to the left and right eyes. Typically, this is encoded using a video that is twice as wide as its 2D version, thus requiring roughly twice as much bandwidth. This overhead would be reduced if one view is presented at the original resolution and the other is subsampled so that only the low spatial frequencies are preserved. There have been some experiments on this in the video compression literature, where this technique is called "binocular suppression" (because the blurry view is perceptually suppressed by the sharp one), though the name mixed resolution is probably more accurate, because the blurry view, in effect, has a much lower resolution than the non-blurry view.

While fusing [Figure 1(b)](#) may produce a relatively pleasing percept, psychophysical experiments have shown that differential input to each eye has detrimental effects on stereoacuity. Halpern and Blake ([1988](#)) showed that stereoacuity for a narrow band target improved when contrast was increased equally for both eyes. Presumably, this is because the increased contrast makes the target easier to localize for each eye. Meanwhile, increasing the contrast in just one eye lowered stereoacuity, even though that should have reduced positional noise from that eye, resulting in better overall thresholds. Only when the target was near threshold did adding contrast to just one eye improve thresholds.
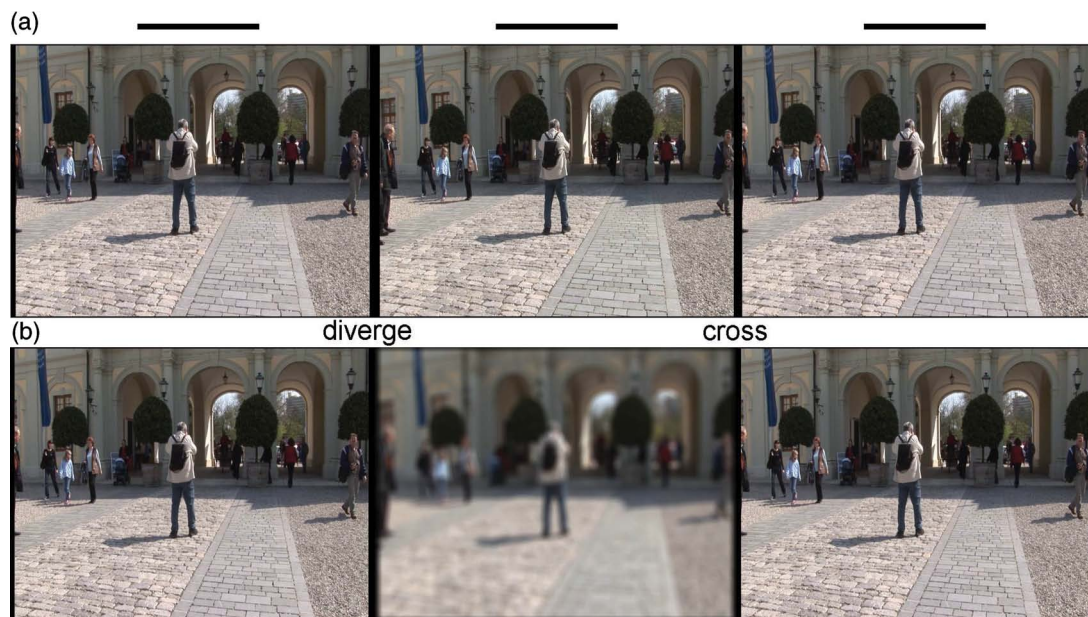
Figure 1. (a) Full-resolution stereoscopic pair, arranged for divergent viewing on the left, or crossed viewing on the right. (b) The same images, but with one eye's image blurred. When fused the scene should appear much sharper than the blurry eye's view alone, but slightly less sharp than (a). Because this effect depends on spatial frequency, the optimal viewing distance is when the horizontal bars subtend 1° of visual angle.

Schor and Heckman ([1989](#)) and Legge and Gu ([1989](#)) reported similar results, also using narrow band stimuli. Kontsevich and Tyler ([1994](#)) suggested that these results could be due to competition between eyes, with the higher contrast target inhibiting the lower contrast one, but Stevenson and Cormack ([2000](#)) have since shown that decrements in performance can be found with analogous two-target tasks performed monocularly, such as vernier acuity. Thus, it is possible that the findings above do not reflect a general principle for how differential information is combined between eyes. Furthermore, Hess, Liu, and Wang ([2003](#)) have shown that this effect disappears when the targets are broadband 1/f (fractal) noise. This is a sobering reminder that the results from simple stimuli often do not generalize to natural scenes.

While differences in contrast may or may not reduce stereoacuity, differences in blur between eyes have consistently been shown to have a negative effect when tested psychophysically. Westheimer and McKee ([1980](#)) used frosted lenses to blur targets and found that stereoacuity was worse when one eye was blurred than when both were blurred. Similar findings were reported by Wood ([1983](#)) and Simons ([1984](#)). Hess et al. ([2003](#)) conducted a comprehensive version of this experiment using 1/f fractal noise patterns processed with either high- or low-pass filters and found the same pattern, but only for low-pass filtering.

Given these findings, it is interesting that the depth in [Figure 1](#) does not appear significantly degraded. One reason for this is that while stereo thresholds are increased, making small differences hard to see, the large depth variations are still visible. Thus, it appears that these large variations alone are sufficient to drive a pleasing depth percept. Perhaps a second reason is that a natural scene provides a large number of other cues to depth, such as shape from shading and perspective, which could be used to enhance a noisy disparity cue (e.g. Robinson & Macleod, [submitted](#)).

While there is a large body of psychophysics literature on how stereoacuity suffers (or not) as a function of differential blur or contrast, there is very little work characterizing the degradation of other visual properties. Meegan, Stelmach, and Tam ([2001](#)) psychophysically measured the sharpness of mixed-resolution images of natural scenes by comparing them with images where blur was applied equally to each eye. They found that the mixed-resolution images were almost as sharp as their unblurred view. Schmidt ([1994](#)) tested Snellen visual acuity after blurring just one eye and found a little blur slightly reduced acuity, but additional blur made almost no difference. They tested only two subjects, however.

In contrast, the engineering literature has repeatedly explored the subjective appearance of stereoscopic content with differing levels of blur in each eye. Stelmach, Tam, Meegan, and Vincent (2000) had subjects rate the subjective quality, depth, and sharpness for mixed-resolution videos, relative to full-resolution versions that were shown for comparison on each trial. On a screen with 50 pixels/degree resolution, they found that depth quality was unaffected even when downsampling to 1/4th resolution and that sharpness quality ratings were only influenced minimally. It is perhaps surprising that the subjective depth changed so little given the results of Hess et al. (2003), but as we argued above, there are many cues to depth in a natural scene, which might subjectively restore the small depth variations that were lost due to the differing blur between eyes.

Aflaki, Hannuksela, Häkkinen, Lindroos, and Gabbouj (2010) used a very similar methodology, but on a display of about half the resolution (only 23 pixels/degree). They found that all mixed-resolution content was given low subjective quality ratings (even when just resampling by 1/2), unlike Stelmach et al. (2000). It is difficult to determine what the key difference is, though perhaps part of it is due to the ambiguity of comparing across subjective rating scales as well as the differing resolutions.

Chen, Bovik, and Cormack (2011) had subjects search for regions of blur applied to natural scenes. When the blur was in the same location in the left and right views, this was an easy task; it became much harder when the blur was in different locations. This was not true for other types of distortion (white noise and JPEG quantization), suggesting that the visual system is particularly good at integrating high and low spatial frequency information between eyes, and can even do so on a spatially localized basis.

Together, these papers show that people can at times be rather oblivious to large differences in sharpness between eyes. With one exception, however, none of these experiments measures the perceptual sharpness of the videos quantitatively. This makes it difficult to compare to the psychophysical results for stereoacuity and, furthermore, makes it difficult to determine how much bandwidth, if any, could be saved using the mixed-resolution technique versus other more traditional techniques for bandwidth reduction that also reduce sharpness (such as encoding both views at a lower overall resolution).

The exception is the work of Meegan et al. (2001), which measured perceptual sharpness of mixed-resolution content by finding the equivalent blur level when applied to both eyes equally. Their results, however, were reported after transforming each subject's data an unknown amount, an approach suited to the question of interest to that paper, but not to quantifying apparent sharpness. There are other reasons to conduct a new study as well. Computer hardware has advanced, allowing the presentation of much higher resolution content, and we can use better display equipment that does not suffer from crosstalk, flicker, or low luminance levels. Thus, we can derive a more accurate measure of the human visual system's ability to fuse high and low spatial frequencies from separate eyes. Not only will this be useful for applied work in 3D video compression, but it is also of interest from a basic vision science perspective.

2   Experiment 1

From previous published work, we expect that a sharp view and a blurry view (hereafter, *mixed-sharpness*) will fuse to produce a percept that is intermediate in apparent sharpness. In particular, we expect that the perceived energy at high spatial frequencies will be reduced. It would, however, take a great deal of subject hours to measure the perceived power spectrum using an adjustment paradigm. Instead, we created videos that varied in how much of their high spatial frequencies had been removed from both views (hereafter, *equal-sharpness*) and used a 2AFC staircase procedure to determine which of these most closely matched a given level of mixed-sharpness. In pilot work, we found that this produced videos that were a good perceptual match to the mixed-sharpness versions.

If we ask subjects just about the apparent sharpness of mixed-sharpness content, it is possible that we will miss other types of visual degradation. In previous work, our group (Jain, Bal, Robinson, MacLeod, & Nguyen, 2012) conducted an experiment where subjects were asked to separately rate videos on a 1–5 scale for sharpness, perceived depth, and overall quality. The mixed-sharpness test videos varied along several dimensions, including: the amount of blur applied, whether the blur was applied to just one eye through the entire video, or if it alternated between the left and right eyes on successive frames, and the frame rate. Even with all of these stimulus variables, overall quality and apparent sharpness produced essentially identical ratings. We conclude therefore that it is reasonably safe to measure sharpness alone, as we do in this experiment.

2.1 Methods

2.1.1 Participants

Eleven naive subjects participated, all with normal or corrected-to-normal acuity, and the ability to perceive stereoscopically defined depth.

2.1.2 Apparatus

Stimuli were presented on a 22-inch LaCie electron22blueIV Diamondtron CRT driven by an NVIDIA GeForce GT 545 GT video card at a refresh rate of 120 Hz, in an otherwise unlit room. A chinrest was used to maintain a viewing distance of 6.4 ft. A mirror stereoscope presented a separate image to each eye. Each image subtended $6° \times 9.6°$ ($W \times H$) with a resolution of 640×1024 pixels. Video playback was controlled using Matlab running the Psychophysics Toolbox, version 3 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) on a Windows XP computer.

Our setup was designed to allow very high resolution content (107 pixels/degree), zero crosstalk between eyes, and zero flicker, a combined set of features that cannot be had using either modern LCD shutter glasses or passive polarized 3D LCDs.

The same apparatus was used for both experiments.

2.1.3 Stimuli

We used two high-quality stereoscopic video clips, which we spatially downsampled[1] to exactly match the native resolution of the screen, and thus were quite sharp. They filled the horizontal dimension of the screen (6° per eye), but only about half of the vertical dimension (4.5°) (the rest of the screen was black). The clips were 2 seconds long, played at 30 Hz. One clip had low motion: a stationary camera view of people walking around a courtyard, where fewer than 20% of pixels changed per frame (a still from this sequence is shown in Figure 1). The other clip had higher motion: the camera panned to track a central figure, and at least 80% of pixels changed on each frame. The clips were excerpts taken from different parts of the same video sequence (downloaded from http://www.3dtv.at/movies/Oldtimers_en.aspx) and, thus, except for motion and scene content, were relatively similar to each other in terms of image features such as saturation and sharpness.

We removed high spatial frequencies in two ways: equal-blur (both views were equally blurred) and mixed-blur (only the right view was blurred[2]). We blurred by filtering with a circular kernel (disk, generated with *fspecial* in Matlab) of different diameters. This produces fewer visual artifacts than downsampling an image to a smaller size and then upsampling back. It also allowed us to generate a wider range of stimuli, since resampling is particularly prone to alias artifacts unless integer ratios (such as 1:2 or 1:4) are used. There also exist many different methods that could be used in the upsampling process, and we did not want to limit our results to a particular method and the artifacts it produces.

Because blurring is slow, we pre-generated all of our clips and then on each trial displayed the video closest to the desired blur level. We pre-rendered the equal-blur stimuli with the following filter diameters (listed in pixels to convey how much information they remove): 1 (no blur), 1.2, 1.4, 1.6, 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 18 (corresponding to 0.56–10.1 arcmin). Note that a diameter of 1 pixel causes no blur, since the resulting convolution (filtering) kernel is exactly 1 pixel in size. A non-integer diameter is approximated by changing the relative weighting of pixels at the edge of the kernel.

For the mixed-blur stimuli, the left view was always unblurred, while the right view blur diameter was either 4, 8, 12, or 16 pixels (2.2–9 arcmin). This produced a total of eight conditions (two levels of motion and four levels of mixed-blur).

For conciseness, we will use pixel units from here on; Mixed-Blur 4 therefore indicates the mixed-blur condition generated with a 4-pixel diameter filter.

2.1.4 Procedure

Each trial consisted of a pair of video clips, labeled Movie A and Movie B. The clips showed the same scene, and only differed in blur level. Subjects had to indicate which of the two clips appeared sharper.

---

[1]We used the Matlab 2011b imresize function with bicubic interpolation and antialiasing.

[2]Previous work has shown that it does not matter which view is blurred, even when eye dominance is considered (Aflaki et al., 2010; Jain et al., 2012; Meegan et al., 2001).

The sequence timing was [Movie A] [0.5s gray inter-stimulus interval] [Movie B] [1.5s inter-stimulus interval] … (Repeat sequence until subject responds A or B). Subjects were encouraged to watch the full sequence at most three times per trial, but no actual limit was imposed.

There were eight conditions, corresponding to each of the different mixed-blur stimuli. Each condition had a separate 1-up/1-down staircase, which determined the equal-blur video it would be compared to on each trial. Our staircase used a large initial stepsize to quickly find the region of subjective equality and then progressively smaller stepsizes to localize it more precisely. After 20 trials, these data were used to initialize a probability-tracking staircase that fits a psychometric curve to the data on each trial and then randomly selected a video with 90%, 60%, 40%, or 10% probability of being preferred over the mixed-blur video. A total of 38 trials were collected per condition. The order of conditions and videos (equal- or mixed-sharpness) was randomly varied between trials.

To give subjects some experience with the task, before the experiment we ran 25 training trials with the same method except that we swapped out the mixed-resolution videos for an unfiltered (full-resolution) video. Because of this, there was always an objectively correct response during training, so we could verify that subjects understood the experiment and could see the stimuli clearly.

Subjects were told to select the sharpest video of each pair, but were not told how the videos were manipulated. We did stress to them the importance of using both eyes, however, "so that they could see the videos in three dimensions."

2.2 Results

One subject showed a strong preference for mixed-sharpness video, no matter how blurry, suggesting they may have misunderstood the instructions; we did not include their data in the following analysis. All 10 remaining subjects demonstrated full understanding of the task by the end of training, and most could discriminate between no blur and the smallest blur filter used. We interviewed subjects after completing the experiment and none guessed that there was any difference in sharpness between one eye's view and the other, nor was there any difference in fatigue between eyes. We also verified that subjects perceived the mixed-sharpness videos as three- dimensional.

To determine the apparent sharpness of the mixed-sharpness videos, we fit a psychometric function to the 2AFC data. From this, we estimated the point of subjective equality (PSE)—the amount of binocular blur such that the binocular blur and the mixed blur were preferred equally. For each subject and mixed-blur stimulus, we did this by fitting a cumulative normal ogive to the data relating the fraction preferred to the amount of binocular blur. The ogives could vary both in PSE (position) and slope; the estimated values of PSE and slope were those that maximized the likelihood of the observed data, assuming binomial variability in the counts. To obtain standard errors for the PSE, we computed the relative likelihood of the data for a range of values of both PSE and slope, and found by integrating across this surface the marginal likelihood of the data as a function of the PSE. The estimated standard error was the standard deviation of a Gaussian function fit to that marginal distribution; this did not differ consequentially from the standard deviation of the marginal likelihood distribution taken directly.

The PSE for each condition, averaged across subjects, is shown in Figure 2(a). Very little consistent difference was observed between the high-motion and low-motion video clips, so we will report their average in the discussion that follows.

For all conditions, the apparent sharpness of the mixed-blur videos was much higher than would be expected from the blurry view alone, and significantly sharper than the average of the blurry and sharp view in most conditions (Figure 2b). The Mixed-Blur 4 condition PSE was 1.5-pixel equal-blur, (2.7 times sharper) a dramatic reduction in apparent blur. The next three larger mixed-blur conditions (8, 12, and 16) were much less sharp, but had quite similar sharpness to each other, with PSE's of 4.2, 4.7, and 5.2, respectively. This suggests that apparent sharpness is not a linear function of the sharp and blurred view's filter sizes. Instead, there seems to be two separate domains: one where the fused view is quite sharp, and the other where the fused view is fairly blurry. In the second domain, increasing the blur in one eye has little effect on the sharpness of the fused percept. Our data cannot speak to the exact transition point between domains, but it does suggest that the two extremes we tested would be the most useful from an applied perspective: Mixed-Blur 4 was quite sharp, and Mixed-Blur 16 was the most efficient, without being much less sharp than Mixed-Blur 8.

One question of interest is how variable the subjects were in their PSE responses. In particular, it would be useful to know if some subjects are particularly bothered by the mixed-blur video. The individual data from the 10 subjects analyzed above are shown in Figure 3, averaged over high and
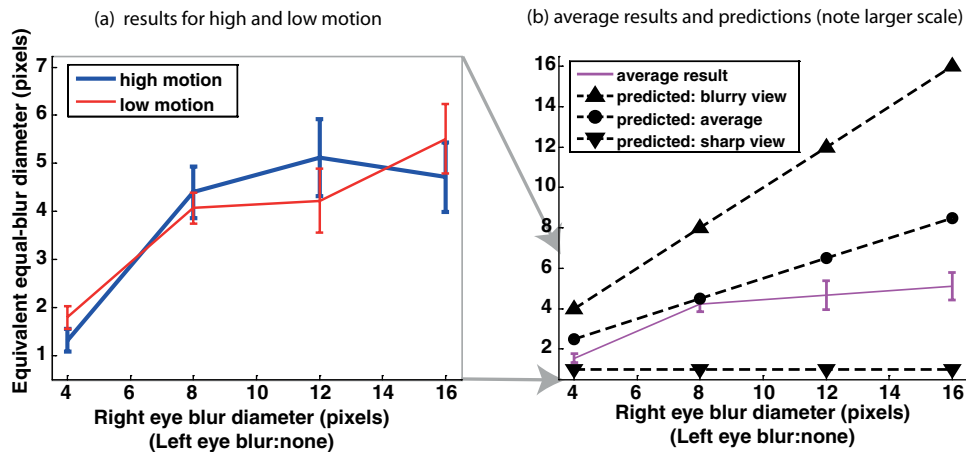
Figure 2. (a) Results of Experiment 1, averaged across subjects. Error bars denote ±1 standard error. (b) The same results, averaged over the two motion conditions and plotted at a smaller scale to allow space to contrast them against three predicted levels of sharpness: the sharpness if either the blurry view or sharp view alone determined the final percept, and the sharpness if the percept were the average of the sharpest and blurriest views.

low motion. Some significant variation between subjects can be seen, and there are clearly some out-liers who see mixed-blur stimuli as significantly less sharp than the prediction made by averaging the sharpness of the left and right images (black dotted line), though the reverse is also true. The level of variability is perhaps most tolerable in the Mixed-Blur 4 condition, however, where even the subject with the least sharp PSE still sees a quite sharp percept.

2.3 Discussion

Our results demonstrate that mixed-blur videos appear much sharper than their blurrier view. They do, however, appear less sharp than their unblurred view. Thus, the contrast (energy) at each spatial frequency is a weighted average of the contrast at that frequency in each eye. The weighting is heavier for the eye with higher contrast, somewhat resembling a max function. The difference in sharpness between our conditions, however, suggests that this weighting function is fairly complex. Much more time-consuming psychophysical methods, using artificial stimuli (e.g. gratings), will most likely be necessary to trace out its exact shape.

Next, we ask how well this generalizes to other spatial frequencies.

3 Experiment 2

Here, we use the same method and stimuli as Experiment 1, except that blur was applied to both left and right sides of the mixed-sharpness videos, so that the sharper of the two views is still noticeably
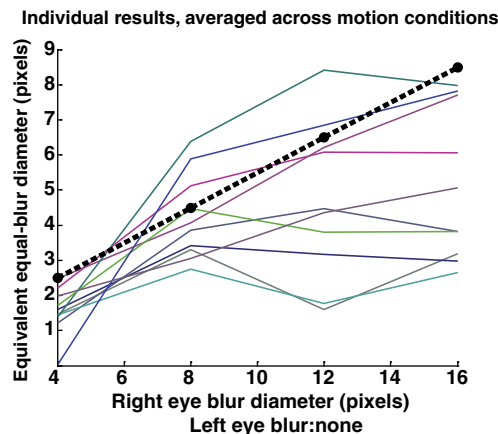

Figure 3. Results of Experiment 1 for each subject. Each line denotes an individual subject's data after averaging over motion conditions; the colors are chosen arbitrarily to help individuate each line. The dashed black line shows the predicted sharpness from averaging the left and right eye's views.

blurred. Will the final percept still be very close to the least blurred view? This tests the generality of our finding from Experiment 1 that a blur ratio of 1:4 (i.e. Mixed-Blur 4) is particularly advantageous. This also tests how well a mixed-sharpness encoding would be suited to a low-resolution display that could not transmit the high spatial frequencies delivered by the unblurred view in Experiment 1.

3.1 Methods

Ten new subjects participated. All had normal or corrected-to-normal acuity, the ability to perceive stereoscopically defined depth, and were naive to the experiment's purpose.

We used the same source videos as in Experiment 1. A new set of mixed-sharpness videos was generated. Four separate blur conditions were created by blurring the right eye's view by a 8-, 12-, or 16-, or 20-pixel (0.074°, …, 0.18°) filter. The left eye's view was blurred by a 4-pixel (0.037°) diameter filter for all conditions. This was done for both low- and high-motion videos, producing a total of eight conditions.

The procedure was otherwise identical to the first experiment.

3.2 Results and discussion

All subjects demonstrated full understanding of the task by the end of the training trials, and most could discriminate between no blur and the smallest blur filter used. We interviewed subjects after completing the experiment and none guessed that there was any difference in sharpness between one eye's view and the other, nor was there any difference in fatigue between eyes.

The difference between high- and low-motion conditions remained slight (Figure 4a), so just as in Experiment 1 we will discuss their average from here on. We found a striking similarity in perceived sharpness across all mixed-blur conditions, as shown in Figure 4(a). Increasing right eye blur from 8 to 20 pixels only increased the equivalent binocular blur from 5.6 to 6.8 pixels. All subjects saw the mixed-resolution content as sharper than its blurriest view (Figure 4b), but the effect was somewhat less impressive than in the first experiment. Indeed, for Mixed-Blur 8, the apparent sharpness was roughly equal to the average blur delivered to each eye.

To formally compare experiments, we will consider the two conditions from the two experiments where the ratio of left eye blur to right eye blur is 1:4. We will characterize the fused sharpness as a weighted average of the blur diameters of the sharp and blurry view, and then compare weights between experiments. For the first experiment (Figure 2), a combination of 1 and 4 pixels of blur produced a PSE of 1.5 pixels of blur; thus, the weight w applied to the more blurred image satisfies $1*(1 - w) + 4*w = 1.5$. For the second experiment, w satisfies $4*(1 - w) + 16*w = 6.6$. The respective weights come out to 0.16 and 0.22, suggesting that the fused percept was more similar to the sharpest view in the first experiment than in the second.

This, and the general flatness of the data from Experiment 2, suggest that the sharpness of the fused percept depends on more than the ratio of blur between left and right eyes. Possibly, the weighting has some dependence on the highest spatial frequencies in the image. Halpern and Blake (1988)
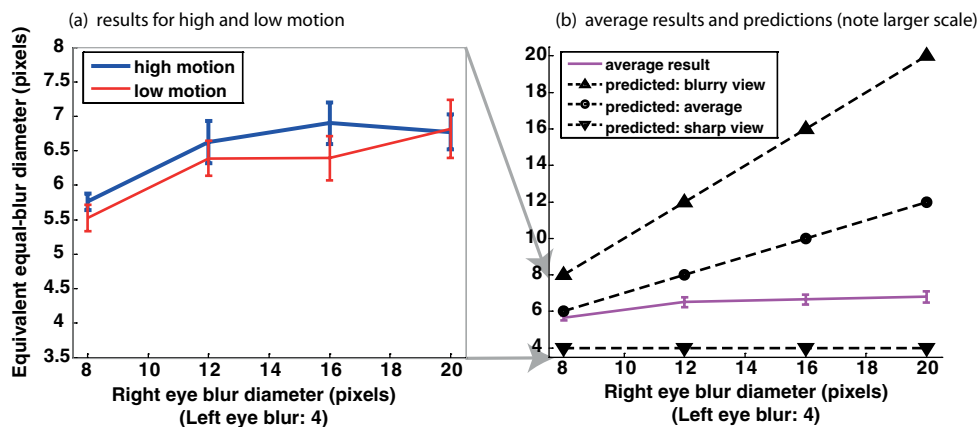

Figure 4. Results of Experiment 2, averaged across subjects. Error bars denote ±1 standard error. (b) The same results, averaged over the two motion conditions, and plotted at a smaller scale to allow space to contrast them against three predicted levels of sharpness: the sharpness if either the blurry view or sharp view alone determined the final percept, and the sharpness if the percept were the average of the sharpest and blurriest views.

showed that the loss in stereoacuity with unequal left and right eye contrasts was greater for targets of low spatial frequency, though this would predict the opposite pattern as we observed here.

From an applied perspective, this result suggests that mixed-sharpness content could be used with low-resolution displays, though the savings would be significantly smaller than for high-resolution displays. Some caution in making this generalization is due, however, since we did not simulate larger pixels directly (and the high-frequency edges between them).

4   General discussion

When the world appears different to each eye, binocular rivalry is often the outcome, with first one eye dominating and then the other. The phenomenon under study here does not exhibit this characteristic. Instead, the final percept is a weighted average of the two eyes. But what is the optimal weighting? If each eye were equally likely to convey the true nature of the external world, then an equal weighting would be preferable. Errors in focus, however, do not increase the sharpness of the image formed on the retina, so retinal input is unlikely to be sharper than the distal stimulus. On this basis, the sharpest image should perhaps receive all the weight, with the blurry image ignored. Instead, we found that the blurry view did receive a small but measurable weight. Perhaps some of the problem is due to the somewhat difficult processes of fusing the sharp and blurry views such that the high spatial frequency content is mapped onto the proper locations in 3D space. The remaining blur may in part be due to failures of this mapping. Alternatively, it could be argued that the slight loss of sharpness is functionally advantageous in normal conditions, where an eye with degraded vision retains some value (albeit reduced) as a source of information.

In terms of sharpness, it may be optimal to give full weight to the sharper eye, but for the purposes of stereopsis it is critical that binocular neurons respond to both eyes. Perhaps it is impossible to downweight the blurry eye completely and still allow the stereo system to function properly.

It is interesting to note that the previous work on stereoacuity reductions from monocular blur (e.g. Hess et al., 2003) shows a very different pattern than for apparent blur we studied here. If the stereoacuity results generalized to apparent sharpness, then the fused percept would actually appear less sharp than the blurriest view! This clearly does not occur. We can only conclude that the stereoacuity results do not reflect a general principle about combining differing views between eyes, and instead reflect the needs of the stereo system.

From a basic science perspective, the biggest question our results raise is why in our first experiment the Mixed-Blur 4 condition looked so much sharper than the other three conditions, and why those other three conditions had essentially the same level of apparent sharpness even though the blur filter diameter doubled between them. This same lack of dependence on the blur filter size was seen in the second experiment. A more parametric stimulus set (gratings or 1/f noise) may be helpful in determining the full relationship, but it is unclear that this will help determine the mechanistic cause itself.

Our results differ from those of Schmidt (1994), in that we found additional losses in sharpness with additional blur (compare Mixed-Blur 4 to the other conditions), whereas she found little change in Snellen acuity with additional blur. Perhaps even while apparent blur increases, judgments that rely on access to high spatial frequencies are unaffected?

From an applied perspective, our results provide guidance as to when a mixed-sharpness encoding would be preferable over equal-sharpness encoding. The simplest way to calculate this is to determine the bandwidth requirements of a given mixed-sharpness video, and compare that with the bandwidth requirements of a perceptually equivalent equal-sharpness video. Whichever uses less bandwidth is preferable, though if the difference is small an equal-sharpness video is probably the safer choice.

Determining the bandwidth requirements depends on the encoding technology. We will consider one approach, that of downsampling one view to a smaller size before encoding, and then upsampling it at playback to equal the spatial scale of the other view. This process would attenuate high spatial frequencies, like our blur filter did. It will also introduce some artifacts, however, which will depend on the upsampling method used. To simplify, we will ignore the artifacts and assume that the bandwidth required to transmit the mixed-resolution image is proportional to the number of pixels in the left and right views after downsampling. But how much can the image be downsampled? This will depend on the quality of the upsampling algorithm; the better the algorithm the sharper the result will be for a given amount of upsampling. Another way to think of this is that to achieve a constant level of sharpness, different upsampling methods will require more or fewer pixels as input. Since upsampling is an active research topic with many approaches, a full evaluation of the different methods is beyond the
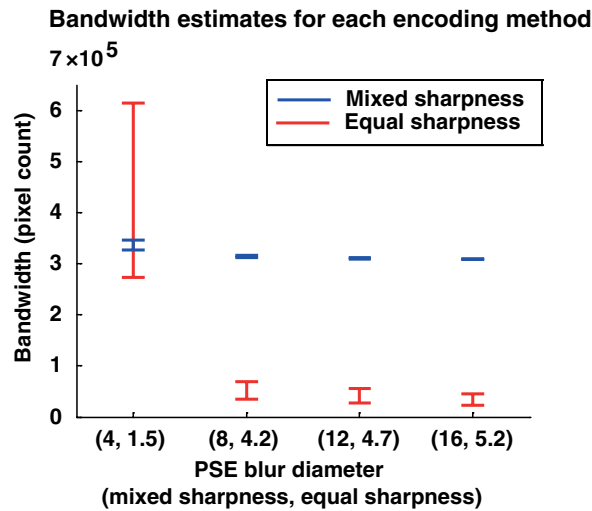
Bandwidth estimates for each encoding method

Figure 5. Bandwidth estimates for each of the mixed-blur conditions in Experiment 1 (and the perceptually equivalent amount of equal-blur). The upper and lower error bars denote the pessimistic and optimistic estimates of the number of pixels required to encode these videos. Lower values indicate a more efficient coding scheme.

scope of this paper. Instead, we will establish an upper and lower bound on how much downsampling could be applied to produce a final image as sharp as the blurred half of the mixed-resolution stimuli we showed our subjects.

One possibility is to scale down by the diameter of our blur filter, in both width and height (thus a 4-pixel diameter filter corresponds to a 1/16th scaling factor of total pixels). This is an optimistic estimate of the savings, since resampling averages over a square area, which is larger than the circular area of our blur filter. If we want to resample over a range no larger than the blur filter, then we must instead scale down by the width and height of a square that is inscribed inside the circle. This is our pessimistic estimate of how much we can downsample the image. Here we show both estimates; with a good upsampling algorithm, the actual result will be somewhere in between. Figure 5 shows the bandwidth estimates, expressed as pixel counts, for a stereo video of 1280×480 (640×480 per eye) for the four mixed-blur conditions in Experiment 1, and the empirically determined perceptually equivalent equal-resolution videos.

For the mixed-resolution videos, the optimistic and pessimistic bandwidth estimates are quite close to each other. Furthermore, increasing the blur diameter from 4 to 16 only reduces the bandwidth by about 10%. This is because the bandwidth of the mixed-sharpness content is dominated by the view that is not resized. From this perspective, Mixed-Blur 4 is probably the most useful of the conditions tested, given that it is much sharper than the other conditions, but only slightly more costly to transmit.

More interesting, however, is the bandwidth comparison between mixed resolution and the perceptually identical video produced by downsampling both left and right views equally. Only for Blur 4 is there any competition—otherwise resizing *both* views is much more efficient. This is perhaps unexpected, but stems from the fact that the bandwidth is the product of width and height (i.e. the area) and thus grows quadratically. Thus, a small reduction in both dimensions results in a significant total bandwidth reduction. Meanwhile, in the mixed-resolution encoding scheme, one eye's view is always transmitted in full resolution.

For the Blur 4 condition, the results depend on which equation (pessimistic or optimistic) is used to calculate bandwidth. With the pessimistic equation, mixed resolution uses almost half as much bandwidth, a significant savings. Using the optimistic equation, however, the equal-resolution condition actually uses less bandwidth by a small amount. Thus, the true gains in bandwidth will depend on the quality of the upsampling method.

We have conducted a similar analysis for the results of Experiment 2 and have found that for every single mixed-resolution condition, the corresponding equal-resolution encoding would be more efficient. In the interest of space, we will therefore discuss it no further. Thus, for bandwidth reduction purposes, it appears that mixed-resolution encoding is most likely to be useful with high-resolution displays.

These results highlight the importance of the upsampling process. We have recently begun work investigating if mixed-resolution content can be upsampled much better than equal-resolution content, due to the fact that one eye's view is transmitted at full resolution. We use the full-resolution view to help construct a high-resolution version of the blurry view. Initial results show that the resulting upsampled view is quite a bit sharper than using traditional techniques (Jain & Nguyen, accepted).

5 Conclusion

Across two experiments, we found that a blurry view and a sharp view combine to produce a fused percept that is much closer in appearance to the sharper view than the blurry one. There is some loss in sharpness, however, and because of the quadratic relationship between image dimensions and bandwidth, only one of the conditions we studied is likely to provide a bandwidth savings over the much simpler method of downsampling both views. We used a high- and (simulated) low-resolution display and found the best performance in the high-resolution (Experiment 1) configuration, suggesting that this is where mixed-resolution encoding has the most promise.

Our results highlight the value of measuring image degradation with psychophysical methods, instead of only collecting subjective (i.e. 1–5) quality ratings. The slight reduction in apparent sharpness we measured suggests that if we had collected subjective ratings, they would have been relatively high, like those reported by Stelmach et al. (2000). Only with our psychophysical data was it possible to determine that this "slight" reduction in sharpness would make an equal-resolution encoding more efficient for most of the conditions we studied.

Nonetheless, even in situations where it is slightly less efficient, the mixed-resolution approach might still be preferable. For instance, mixed-resolution allows the addition of a 3D video stream with very little additional bandwidth, allowing older 2D devices to show the sharpest possible content, while a 3D device can use the same stream to show full 3D content.

It is also important to consider that our results probably reflect the worst-case sharpness of mixed-resolution content. Subjects viewed the same clips repeatedly and thus had the opportunity to search for the image regions that were most diagnostic. Informally, we have noticed that these are regions with high-contrast edges that are relatively isolated from other edges that could potentially mask the spatial smearing that occurs with blur. Though common in natural images, in regular viewing these are not the only areas that people would look, and thus the loss of sharpness would be less obvious in natural viewing.

Going forward, one possible avenue for research is dynamically determining regions of the image where less blur should be applied, in order to prevent these more obvious blur signals. Such a system could be achieved by measuring the local contrast and edge strength within a window equal to the region of support of the equivalent blur filter and adjusting the resolution accordingly. Spatially varying techniques have earlier been successfully applied in video processing to compress or enhance data. Wang and Bovik (2001) implemented an image coding system exploiting the foveal frequency response, radially blurring images to achieve a lower bandwidth. A frame interpolation technique based on computing visual saliency is demonstrated in Jacobson and Nguyen (2012). Content-aware downsampling, however, will depend on a full psychophysical investigation of how spatial frequencies from the left and right eyes are weighted in the final fused percept.

We should note that there are some potential downsides to mixed-resolution content besides the reduction in apparent sharpness. First, for viewers with significantly worse vision in one eye, there is a 50% chance that the full-resolution image would be delivered to that eye. This would clearly limit the quality of the fused image, since the full-resolution image is critical to compensating for the blurry image. Second, there is good evidence that long-term exposure to unfocused images can induce changes in eye shape (Wildsoet, 1997). These changes can be prevented, however, if subjects have access to properly focused images during a substantial time period each day. Thus, mixed-resolution video is unlikely to be detrimental because only the screen itself is blurry, so every eye movement away from the screen would serve to counteract any signals that would drive long-term change. Nonetheless, for these two reasons it is worth considering an alternate method of showing mixed content than used in this paper, where the sharp and blurry views are swapped between eyes on every other frame. We have investigated this approach and found that it produces significant flicker at 30 Hz, but works quite well at 60 and 120 Hz (Jain, Robinson, & Nguyen, 2013).

Finally, there is the question of how comfortable people would find mixed-resolution content over longer periods of time, such as the duration of a TV show or movie. In Jain et al. (2013), we found

no evidence of increased fatigue after watching 10 minutes of mixed-resolution video without breaks. This is encouraging, though it would probably be worth testing over even longer viewing periods. It is interesting to note that monovision contact lens prescriptions (used to treat presbyopia) produce an experience somewhat similar to mixed resolution. In this approach, one lens is selected to give clear long-distance vision and the other clear near vision (typically, for reading). This is not identical to mixed-resolution video, since the blur amount is distance dependent, but it certainly is related. Unfortunately, the success rate of this method (the number of people who try it and elect to continue to use it) is only about 64% (Evans, 2007); however, the primary reasons for this are likely to be non-issues for our application: poor suppression of the blurry image at night, and the need for a third focal length in between the near and far ones. Nonetheless, the fact that many users are able to adapt to monovision, despite these problems, and in often with greater differences in blur between each eye is certainly encouraging.

REFERENCES

Aflaki, P., Hannuksela, M. M., Häkkinen, J., Lindroos, P., & Gabbouj, M. (2010). Impact of downsampling ratio in mixed-resolution stereoscopic video. In *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON),* Tampere, Finland, 1–4.

Brainard, D. H. (1997). The Psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. doi:10.1163/156856897X00357

Chen, M. J., Bovik, A. C., & Cormack, L. K. (2011). Study on distortion conspicuity in stereoscopically viewed 3D images. In *IEEE 10th IVMSP Workshop,* Ithaca, NY, 24–29. doi:10.1109/IVMSPW.2011.5970349

Evans, B. J. W. (2007). Monovision: A review. *Ophthalmic and Physiological Optics*, *27*, 417–439. doi:10.1111/j.1475-1313.2007.00488.x

Halpern, D. L., & Blake, R. R. (1988). How contrast affects stereoacuity. *Perception*, *17*(4), 483–495. doi:10.1068/p170483

Hess, R. F., Liu, C. H., & Wang, Y.-Z. (2003). Differential binocular input and local stereopsis. *Vision Research*, *43*(22), 2303–2313. doi:10.1016/S0042-6989(03)00406-1

Jacobson, N., & Nguyen, T. Q. (2012). Scale-aware saliency for application to frame rate upconversion. *IEEE Transactions on Image Processing, 21*, 2198–2206. doi:10.1109/TIP.2011.2179055

Jain, A. K., & Nguyen, T. Q. (accepted). (2013). Video super-resolution for mixed resolution stereo. Accepted to *2013 IEEE International Conference on Image Processing*. Melbourne, Australia, 15–18 September. doi:10.1109/TCSVT.2013.2278159

Jain, A. K., Bal C., Robinson, A., MacLeod, D., & Nguyen, T. Q. (2012). Temporal aspects of binocular suppression in 3D video. *Paper presented at the Sixth International Workshop on Video Processing and Quality Metrics for Consumer Electronics*. Scottsdale, Arizona, 19–20 January.

Jain, A. K., Robinson, A., & Nguyen, T. Q. (in press). Comparing perceived quality and fatigue for two methods of mixed resolution stereoscopic coding. *IEEE Transactions on Circuits and Systems for Video Technology.*

Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago, IL: University of Chicago Press.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3. *Perception*, *36*, 14. doi:10.1068/v070821

Kontsevich, L. L., & Tyler, C. W. (1994). Analysis of stereothresholds for stimuli below 2.5 c/deg. *Vision Research*, *34*(17), 2317–2329. doi:10.1016/0042-6989(94)90110-4

Lee, B. B., Sun, H., & Zucchini, W. (2007). The temporal properties of the response of macaque ganglion cells and central mechanisms of flicker detection. *Journal of Vision*, *7*(14), 1.1–1.16. doi:10.1167/7.14.1 ◄

Legge, G., & Gu, Y. (1989). Stereopsis and contrast. *Vision Research*, *29*(8), 989–1004. doi:10.1016/0042-6989(89)90114-4

Meegan, D. V., Stelmach, L. B., & Tam, W. J. (2001). Unequal weighting of monocular inputs in binocular combination: Implications for the compression of stereoscopic imagery. *Journal of Experimental Psychology: Applied, 7*, 143–153. doi:10.1037/1076-898X.7.2.143

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. doi:10.1163/156856897X00366

Robinson, A., & MacLeod, D. (submitted). Depth and luminance edges attract. *Journal of Vision*

Schmidt, P. P. (1994). Sensitivity of random dot stereoacuity and Snellen acuity to optical blur. *Optometry and Vision Science*, *71*(7), 466–471.

Schor, C., & Heckmann, T. (1989). Interocular differences in contrast and spatial frequency: Effects on stereopsis and fusion. *Vision Research*, *29*(7), 837–847. doi.10.1016/0042-6989(89)90095-3

Simons, K. (1984). Effects on stereopsis of monocular versus binocular degradation of image contrast. *Investigative Ophthalmology and Visual Science*, *25*(8), 987–989.

Stelmach, L., Tam, W. J., Meegan, D., & Vincent, A. (2000). Stereo image quality: Effects of mixed spatio-temporal resolution. *IEEE Transactions on Circuits and Systems for Video Technology, 10*, 188–193. doi:10.1109/76.825717

Stevenson, S. B., & Cormack, L. K. (2000). A contrast paradox in stereopsis, motion detection, and vernier acuity. *Vision Research*, *40*(21), 2881–2884. doi:10.1016/S0042-6989(00)00164-4

Wang, Z., & Bovik, A. C. (2001). Embedded foveation image coding. *IEEE Transactions on Image Processing, 10*, 1397–1410. doi:10.1109/83.951527

Westheimer, G., & McKee, S. P. (1980). Stereoscopic acuity with defocused and spatially filtered retinal images. *JOSA*, *70*(7), 772–778. doi:10.1364/JOSA.70.000772

Wildsoet, C. F. (1997). Active emmetropization: Evidence for its existence and ramifications for clinical practice. *Ophthalmic and Physiological Optics*, *17*(4), 279–290. doi:10.1046/j.1475-1313.1997.97000033.x

Wood, I. C. J. (1983). Stereopsis with spatially-degraded images. *Ophthalmic and Physiological Optics*, *3*(3), 337–340. doi:10.1111/j.1475-1313.1983.tb00622.x