PROCEEDINGS OF SPIE

Determinants of perceived image quality: ghosting vs. brightness

Wilcox, Laurie, Stewart, Jeffrey

SPIE.

Determinants of perceived image quality: ghosting vs. brightness

Laurie M. Wilcox and Jeffery A.D. Stewart
York University, Centre for Vision Research, Toronto, Canada. M3J 1P3

ABSTRACT

The physical specifications of stereoscopic eyewear are routinely documented. However, their effects on the appearance or perceived quality of 3D images is most often evaluated superficially, if at all. Here we apply psychophysical techniques to assess the influence of ghosting and perceived brightness on judgements of image quality. To determine which of these variables has the largest impact we simulated several levels of ghosting and brightness in a digital version of a 70mm 3D image sequence. We then presented these image sequences in a large-format 3D theatre and used a magnitude estimation task to assess image quality. The data were clear in showing a significant effect of ghosting on perceived quality but no effect of image brightness. From this we argue that image ghosting is a critical determinant of perceived image quality and should be a primary consideration in relevant technology decisions.

Keywords: Perceived image quality, ghosting, brightness, 3D film, stereoscopic eyewear

1. INTRODUCTION

One's choice of stereoscopic eyewear has significant effects on the quality of depth perception from a 3D scene. Although the *physical* consequences of the eyewear on the resultant images are routinely and carefully documented their *perceptual* impact is most often evaluated in a more cursory manner. When perceptual evaluations are performed, responses are often obtained from individuals within the institution or company who (a) have considerable experience viewing stereoscopic displays, and (b) are not naïve to the intent of the comparison. Further, this information tends to be obtained without rigourous experimental control. This practice has a number of important potential consequences. The most obvious of these is that the observations may not be generalizable to the experience of the average viewer. For example, it is well known that stereoacuity, or the ability to resolve stereoscopic depth differences, improves markedly with experience[1]. Therefore, it is quite likely that the experienced viewer will be more sensitive to some deficits in the 3D scene, for instance those resulting from camera misalignment or projection errors. On the other hand, there will also be greater tolerance of other attributes such as diplopic images caused by extremely large depth offsets. Further, prior knowledge of the intent of a particular comparison will create implicit biases in an individual's responses, again making it difficult to generalize their data.

In the experiments described here we apply psychophysical techniques to assess the perceptual consequences of the choice between two types of stereoscopic eyewear technology: polarized filters vs. a combination filter / shutter system. It is well established that polarized filter –based eyewear is susceptible to interocular cross-talk or ghosting.[2] The advantage of the combination technology is that this ghosting is virtually eliminated; however the added shutter system reduces the overall luminance of the display significantly more than polarized filters alone. The question we address here is which of these two factors (perceived luminance or ghosting) most influences perceived image quality.

2. METHODOLOGY

2.1 Subjects
A total of 77 paid observers participated in this study. They were recruited by advertisement at York University (Toronto, Ontario) and ranged in age from 17 to 40 years. All had normal or corrected to normal visual acuity and were able to see depth via stereopsis. 60% of the subjects had seen a 3D film prior to participating in this study.

2.2 Stimuli
The stimuli used in this study were two reels of 70mm film (one for each eye's view) depicting a series of 2sec sequences of a red car driving into a park on a sunny day. The original 2s sequence was manipulated digitally using Cineon™ software to simulate 8 levels of ghosting and 4 levels of luminance. Ghosting was simulated by adding various percentages of the left-eye image to the right-eye image. The percentages used here refer to the percentage of one eye

that was added to the other ie 0, 5, 10, 15, 20, 25, 30, and 50 percent. The luminance values (100, 92, 84, and 75 percent) are expressed in terms of the percentage reduction from the original image attained by rescaling the original range of intensities. Unfortunately, in a 3D display many variables influence the amount of light reaching an observer. These include aspects of the printing process, the screen characteristics, the observer's distance and angle relative to the screen, projector alignment as well as the type and age of bulb used. Therefore we measured the maximum luminance at a number screen locations from a variety of positions in each theatre the day prior to testing. Note that although the range of test luminances may seem narrow, pilot tests revealed that perceptually the range was broad, and in fact at luminances lower than 75% the images were reportedly too dark to make accurate ratings. After rendering, the images were printed to film. The individual conditions were then spliced in random order to create the final 70mm test reels.

2.3 Apparatus

Testing took place in two commercial large-format 3D theatres using the test reels described above. The screen size was approximately 64 x 82 ft in both locations. One theatre (A) used glasses with a combination liquid-crystal shutter and polarization technique to achieve the stereoscopic effect; the other (B) used polarized glasses. Extensive luminance measurements were taken using a Minolta LS100 spot photometer the day prior to testing in both theatres. Luminances were recorded at 9 positions on the screen in a 3x3 grid and from 9 locations in the theatre, again in a 3x3 grid. The measurements were taken through the glasses for each theatre, and therefore represent the luminance reaching the observer's eye. Under these conditions the average maximum luminance in Theatre A was 0.78 ftl (se = 0.10) while the average maximum for Theatre B was 1.01 ftl (se = 0.13). We also recorded the pre-existing or baseline amount of ghosting by presenting a bright spot on a black screen projected to one eye only. Again measurements were made at 9 screen and seat locations and the average signal to noise (S/N) ratios showed that there was virtually no crosstalk (0.001ftl) in Theatre A producing an average S/N ratio of 780, whereas Theatre B had an average S/N ratio of 25.4 (se = 3.45). Prior to entering the theatre, subjects were given miniature flashlight pens and a clipboard with the response sheets. The first page was a consent form and subjects were asked to read and sign this before entering the theatre. All subjects were seated towards the middle of the theatre to ensure a similar range of depth percepts and perceived luminance. Only participants in this study were admitted to the theatre during testing.

2.4 Procedure

A fully factorial randomized design was used with 4 (luminance) x 8 (ghosting) conditions. A classic magnitude estimation task introduced by Stevens[3] was used to evaluate perceived image quality as a function of perceived luminance (herein referred to as brightness) and ghosting. Data were collected in two sessions (one for each theatre) separated by one week using exactly the same protocol. Prior to testing, observers were shown an image of 'average' quality (intermediate brightness and ghosting) and told to assign this image a value of 100. They were then instructed to rate subsequent images relative to this reference; importantly they were free to use any numbers they wished.[1] This reference image was shown repeatedly throughout the test period (every 10th image) to remind observers of their basis for comparison. Before beginning testing the procedure was explained in detail to the subjects, they were shown how to wear the glasses, and how to turn on their pens. Subjects were asked if they had any questions, and told that once started the test session could not be paused for any reason. After all questions had been addressed a series of 10 practice trials were initiated. On each trial the experimenter called out the trial number and a white dot appeared for 1sec, followed by the 2sec film clip and a 6sec response interval. Observers were told when the practice session was completed and testing started. Following one series of trials (127 observations), the subjects were given a 15 min rest period while the film was rewound. The second session was identical to the first and included practice trials. However, subjects were not aware that the same film was used in the two test periods. In total, the subjects made 256 observations (8 per condition) and testing took approximately 50 min. At the end of the session observers were asked to answer some open-ended questions on the last page of their response booklet. These questions included whether they had prior experience with 3D films, and what image features were most important in determining their response.

---

[1] One of the main drawbacks of simple rating scales is that the resulting data are ordinal in nature that is, a rating of 4 is greater than a rating of 2 but may not be perceptually *twice* the intensity or quality. Steven's estimation procedure retains the interval nature of the ratings therefore it can be argued that the data are more informative about the relative intensity or quality of the stimuli[4].

3. RESULTS

The magnitude estimation data were averaged across subjects for each of the 32 test conditions, separately for the two theatres. Because the ratings were very consistent both within and across subjects, no normalization was necessary. The data obtained in the two test sessions are shown in Figure 1. Ratings of perceived quality dropped significantly as ghosting increased, particularly over the range of 0 to 20%. However, there was no discernible effect of brightness, except possibly a slight reduction in perceived quality in the 75% condition, *but only when no ghosting was present*. This pattern of results was the same for both theatres but note that, when no ghosting was present, quality ratings were slightly higher in Theatre A.
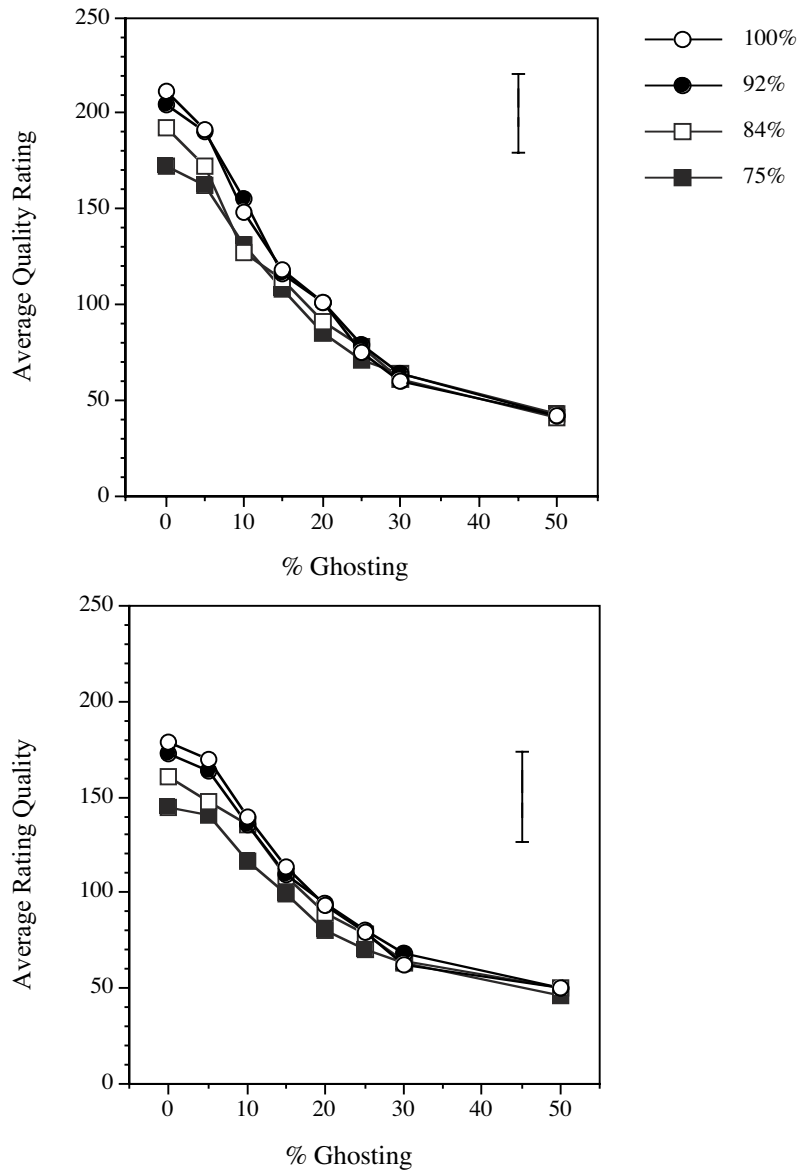
Figure 1. Rating scores are presented here for each of the four brightness levels as a function of ghosting. Results are averaged across subjects and are shown separately for the two theatres A (combination eye wear) and B (polarized eye wear). The solid line to the right indicates the average standard error of the mean across the data set.

A histogram depicting subjects' responses to the question: "What image quality (attribute) was most important in making your quality ratings" is shown in Figure 2. By far the most common attribute seen to influence image quality was ghosting.
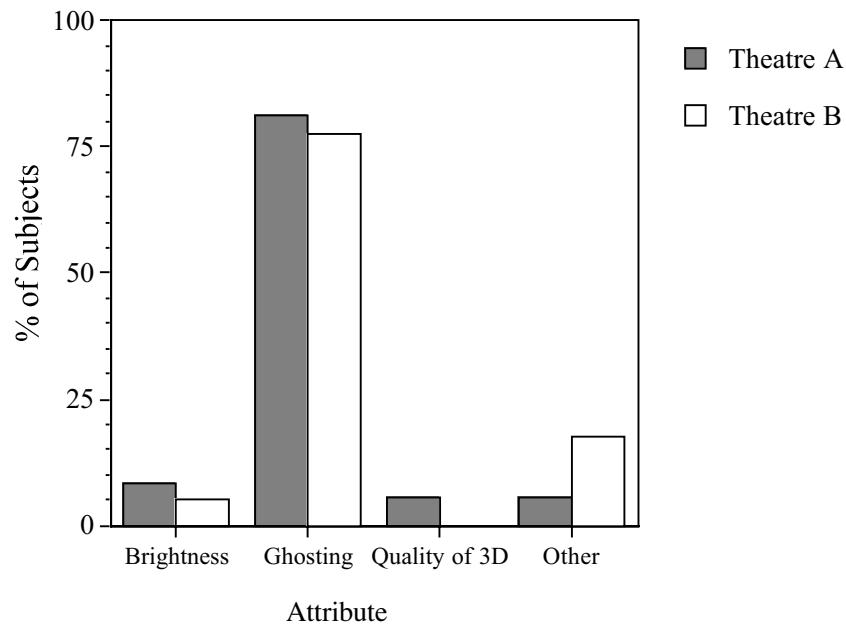
Figure 2. Shown here is a histogram summarizing the responses subjects made to the question "what image quality (attribute) was most important in making your quality ratings". The grey and white bars represent results obtained in Theatres A and B respectively.

4. CONCLUSIONS

The results of this experiment provide strong evidence that image ghosting is more critical than brightness in determining perceived quality. This pattern of results was observed when observers wore either the combination eyewear or the polarized eyewear and was supported by open-ended questionnaire responses. That is, over 75% of the observers said that ghosting was the most important attribute in determining their rating. Only 5-8% of the subjects said the same of image brightness. In fact the post-test survey shows that in our study ghosting was *the* most important factor in determining perceived image quality, even when no pedestal ghosting was present (ie Theatre A).

The observed resilience to reduced brightness is not difficult to explain; the human visual system is designed to adapt to the prevailing lighting. Thus an observer can and will adapt to the level of luminance while viewing 3D (and 2D) image sequences. However it is virtually impossible to 'adapt' to ghosting as it varies across the image according to the depth and contrast of objects. Further, in standard film sequences, the image content will vary from scene to scene, again altering the pattern of ghosting. Given the repetitive nature of our stimuli it is arguable that, if anything, the effect of ghosting is underestimated by our study. It is also worth noting that in this study the participants did not make *simultaneous* comparisons between image sequences of different luminances. It is probable that under such conditions one would find a substantial effect of image brightness on perceived quality. However, we would argue that such a comparison, though valuable for some purposes, is not representative of natural viewing conditions.

The experiment was conducted in two theatres primarily to allow for the fact that the glasses worn during testing would influence the amount of ghosting and the brightness of the stimuli, possibly adding a pedestal amount of each variable to the test sessions. Comparison of the average maximum screen luminances measured prior to testing showed that in fact the maximum luminance (assessed through the eyewear) was not markedly different for the two venues. However the use of polarized glasses in Theatre B did reduce the S/N ratio substantially. It is likely that this ghosting is responsible

for the only appreciable difference between the data obtained from the two theatres. That is ratings were lower in Theatre B (170) than in Theatre A (220) when no simulated ghosting was present.

In addition to showing that ghosting is an important determinant of image quality, our data show that for the average observer there is a reduction in perceived quality with the introduction of ghosting at levels as low as 10%, and a dramatic degradation at 20% ghosting. Therefore, efforts to reduce ghosting at almost any level and even below 10% will enhance the perceived 3D image quality. This conclusion is relevant not only to the presentation of 3D film, but also for the design of autostereoscopic displays[5] and other 3D display technologies.[6,7]

ACKNOWLEDGMENTS

REFERENCES

1.	M. Fendick and G. Westheimer. Effects of practice and the separation of test targets on foveal and peripheral stereoacuity. *Vision Research*, **23**, p. 145-150 (1983).
2.	P. J. Bos, *Stereo computer graphics and other true 3D technologies*, ed. D.F. McAllister, Princeton: Princeton University Press, (1993).
3.	S. S. Stevens. On the psychophysical law. *Psychological-Review. 1957; 64: 153-181*, **64**, p. 153-181 (1957).
4.	G. A. Gescheider. Psychophysical Scaling. *Annu. Rev. Psychol.*, **39**, p. 169-200 (1988).
5.	G. J. Woodgate, et al. Autostereoscopic 3D display systems with observer tracking. *Signal Processing: Image Communication*, **14**, p. 131-145 (1998).
6.	C. H. Chen and A. R. L. Travis. Strategies for 3D video with wide fields-of-view. *IEE Proceedings - Optoelectronics*, **148**, p. 85-90 (2001).
7.	T. Kawai. 3D Displays and applications. *Displays*, **23**, p. 49-56 (2002).