Journal of Imaging Science and Technology® 52(2): 020504–020504-5, 2008.

Capturing the Added Value of Three-Dimensional Television: Viewing Experience and Naturalness of Stereoscopic Images

Pieter J. H. Seuntiëns and Ingrid E. J. Heynderickx
Philips Research Laboratories, Prof. Holstlaan 4, 5656 AA, Eindhoven, The Netherlands
E-mail: pieter.seuntiens@philips.com

Wijnand A. IJsselsteijn
Eindhoven University of Technology, Den Dolech 2, 5600 MB, Eindhoven, The Netherlands

Abstract. *The term "image quality" is often used to describe the performance of an imaging system. Recent research showed however that image quality may not be the most appropriate term to capture the evaluative processes associated with experiencing three-dimensional (3D) images. The added value of depth in 3D images is clearly recognized when viewers judge image quality of unimpaired 3D images against their two-dimensional (2D) counterparts. However, when viewers are asked to rate image quality of impaired 2D and 3D images, the image quality results for both 2D and 3D images are mainly determined by the introduced artifacts, and the addition of depth in the 3D images is hardly accounted for. In this article we describe an experiment where we applied and tested two alternative evaluative concepts: naturalness and viewing experience. It was hypothesized that these concepts would be more sensitive to the added value of depth in 3D images. Four scenes were used, varying in dimension (2D and 3D) and noise level (six levels of white Gaussian noise). Results showed that both viewing experience and naturalness were rated higher in 3D than in 2D when the same noise level was applied. Thus, the added value of depth is clearly demonstrated when the concepts of viewing experience and naturalness are being evaluated in contrast to earlier results found using image quality. The added value of 3D over 2D, expressed in noise level, was* 2 dB *for viewing experience and* 4 dB *for naturalness, indicating that naturalness appears the more sensitive evaluative concept for demonstrating the psychological impact of 3D displays.* © *2008 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.(2008)52:2(020504)]

INTRODUCTION

Since the introduction of the television, much has been done to improve the overall experience of viewers. Improvements in picture quality, sound quality, and increasing involvement based on larger screen sizes have contributed to a better overall viewing experience. A logical next step is the introduction of three-dimensional (3D) content. Proponents of 3D-TV have argued that it will bring the viewer a whole new experience, a fundamental change in the character of the image, not just an enhancement of quality.[1,2]

Stereoscopy is the major depth cue and occurs because the human eyes are horizontally displaced in the head, resulting in a slightly different image for each eye. The brain fuses these different images to one image and extracts depth information from the difference between the two.[3] Different technologies exist for rendering the binocular depth cue from two-dimensional (2D) images.[4] Most common is the system where lenticular lenses are used to direct the images to the appropriate eye. These systems can be viewed without the use of an optical device (unaided viewing) and are called auto-stereoscopic systems. Another technique displays left and right images on the screen in a different color or a different polarization and red-green color filters or polarized filters in glasses are used to determine which image should be received by which eye. These systems, where the viewer needs to wear an optical device, are known as aided viewing.

Comparisons between television sets are done quite regularly on perceptual and/or technical aspects to determine where to put future investments. The performance of a 3D television system is often evaluated using 2D image quality models.[5] Perceived image quality is considered to be a multidimensional attribute. Earlier research in this area defined some dominant perceptual factors affecting 2D image quality, for instance, blur, brightness, color, blockiness, or noise. Psychophysical scaling experiments are used to quantify the strengths of these artifacts. People use perceptual rules to combine the measured strengths into a prediction of the overall image quality. Perception research in the area of 3D-TV has shown that the 2D Image Quality Circle model as proposed by Engeldrum is not adequate to measure the added value of depth since the depth reproduction is not incorporated in the perceived image quality.[6,7]

Therefore, we need a higher level concept that takes into account the image quality as well as the added value of depth. In this study, we evaluated the concepts' viewing experience and naturalness in order to check whether these concepts take both the image quality of the image as well as the added value of depth into account. The concept naturalness was originally introduced to determine the perceived quality of color reproduction. The concept is often defined as perceptual realism which is believed to have a depth component as well as a quality component. The concept of view-

Received Dec. 10, 2005; accepted for publication Nov. 23, 2007; published online Apr. 2, 2008.
1062-3701/2008/52(2)/020504/5/$20.00.

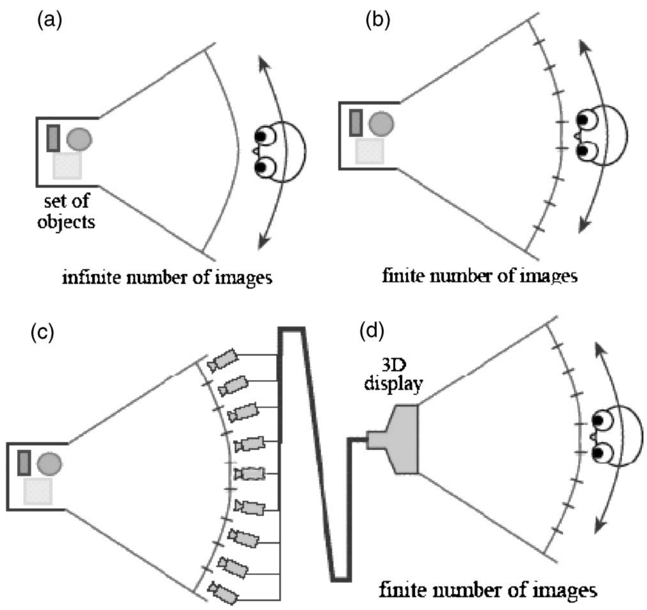

Figure 1. Panel (a) shows an observer watching a set of objects. The viewing window is divided in nine different perspective views in panel (b). In this experiment the nine different views were generated by using nine different cameras as shown in panel (c). The screen displays the nine different views in a viewing zone in panel (d).

ing experience is defined as the users' perceptual and cognitive experience of the entire application. In addition, our experiments served to calibrate the sensitivity of these concepts in relation to each other in terms of their response pattern to increasing levels of noise introduced in the 2D and 3D images. This method also allows us to quantify the potential stereoscopic advantage in terms of dB noise level.

METHOD

Design

The experiment had a mixed design with image (four images), dimension (2D versus 3D), and noise (six levels) as within subject factors, and the two different evaluation concepts (naturalness and viewing experience) tested between subjects.

Observers

Thirty observers working in a research environment (including some graduate students) participated in the experiment. Twenty observers participated in the experiment evaluating viewing experience and ten observers participated in the experiment evaluating naturalness. All participants had a visual acuity of >1 (as tested with the Landolt-C test) and good stereo vision, <30 s of arc (as tested with the Randot stereo test). The viewing distance was 1.5 m.

Materials

Equipment

A 20″ Philips nine-view autostereoscopic display was used in this experiment. The advantage of this display, besides 3D viewing without glasses, is the support of motion parallax (laterally) enabling viewers to look around objects by moving their head. Figure 1 explains the basic principle. Figure 1(a) shows an observer watching a set of objects. The left and right eye both receive a different view of the scene. By moving the head, the observer receives different views of the scene enabling him to see nine different views. Figure 1(b) shows the same viewing window, but this time for practical reasons divided into a finite set of horizontal frames. When properly positioned, each eye receives a view from a different frame, thereby preserving disparity. When moving the head, both eyes will shift to a different view, resulting in perceived motion parallax, but with a reduced amount of views. In a practical setup, nine different views were generated using nine cameras [Fig. 1(c)] and these nine views were integrated in a multi-view autostereoscopic display [Fig. 1(d)]. A set of nine successive views is called a viewing zone and repetition of this viewing zone enables multiple viewers to watch 3D. Figure 2 shows three zones consisting of nine views each. The resolution of the display was 1600×1200 pixels and the optics were optimized for a viewing distance of 1.5 m. Custom built software was used to display the image material on the Philips multi-view autostereoscopic display.

Figure 2. Three viewing zones consisting of nine different perspective views each. The repetition of viewing zones enables multiple viewing.

Stimuli

The image material used in this experiment consisted of four still images, *Minibeamer, Puzzle, Rose, and Shaver*, recorded with a nine camera setup. The advantage of recording all the views with nine cameras instead of converting a 2D image into nine views, is that all required information is available and no distortions due to limited depth information are introduced in the 3D material. Displaying the nine views on the multi-view autostereoscopic display resulted in 3D perception of the image because each eye receives a different view with a different perspective. The 2D situation was simulated by implementing the middle view (view five) into all nine views. In this case, the observer always perceives the same image on both eyes, resulting in 2D perception. The middle view (camera five) of each image is shown in Figure 3.

Since our main goal was to quantify the added value of depth through the concepts viewing experience and naturalness in terms of the affordable loss in image quality, an

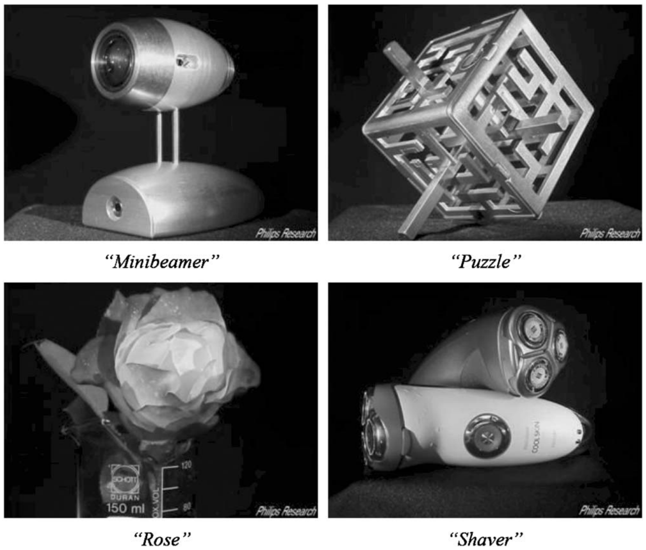

Figure 3. The four panels show the original scenes *Minibeamer, Puzzle, Rose*, and *Shaver*.

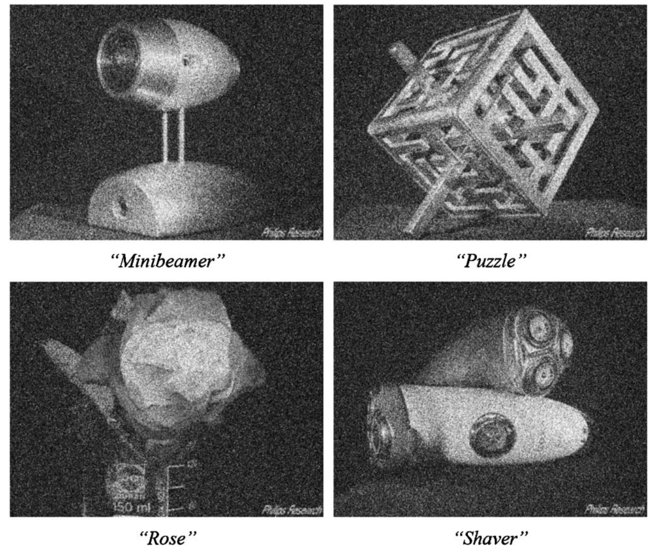

Figure 4. Noise impaired scenes *Minibeamer, Puzzle, Rose*, and *Shaver*.

appropriate image distortion had to be chosen. Artifacts like, for instance, blurring, blocking, and ringing appear in different forms on different TV systems and their visibility depends on image content. Additive noise, however, seems to manifest itself in the same way over many different systems, and in principle, is image independent. It can be modeled as the image $\mathbf{f}(i,j)$ being the sum of the true image $\mathbf{s}(i,j)$ and the noise $\mathbf{n}(i,j)$. The model is shown in Eq. (1)

$$\mathbf{f}(i,j) = \mathbf{s}(i,j) + \mathbf{n}(i,j). \tag{1}$$

The noise is modeled with an independent, additive model, where the noise $\mathbf{n}(i,j)$ has a zero mean ($x=0$) Gaussian distribution described by its standard deviation (σ), or variance (σ^2). This means that each pixel in the noisy image is the sum of the true pixel value and a random, Gaussian distributed noise value. The additive noise is evenly distributed over the frequency domain (i.e., white noise). The white Gaussian noise impairment was implemented using the MATLAB image noise filter with five levels of noise ($x=0$, $\sigma^2=0.00125$, 0.0025, 0.005, 0.01, 0.02). An increasing σ^2 parameter produced more noise in the images. Figure 4 shows the four scenes with additive noise ($x=0$ and $\sigma^2=0.02$).

Procedure

The experiment consisted of two sessions: one for measuring viewing experience and one for measuring naturalness. In both sessions exactly the same setup was used. The observers were given a brief instruction about the experiment on paper as well as a definition of the concepts' naturalness and viewing experience (see Introduction). Any remaining questions were answered and subsequently a short training session was conducted. The training session allowed the participants to get used to the setting as well as the tasks. In the training, six still images were presented with different noise levels, including the extremes used in the actual experiment. The observers were asked to rate viewing experience and naturalness on a scale labeled with the adjective terms [bad]-[poor]-[fair]-[good]-[excellent] according to the ITU[8] recommendation for subjective quality assessment. Participants were free to mark their assessment anywhere on the vertical rating scale. The order in which the images appeared was randomized throughout the experiment and each image was evaluated twice. The images were displayed for 10 s followed by a gray field for 3 s. In total, 20 participants had to indicate their viewing experience 96 times [4 images $\times 6$ distortion levels (original +5 noise impairment levels) $\times 2$ conditions (2D and 3D) $\times 2$ (repetition)]. Exactly the same setup was used for the naturalness ratings, only this session was done by 10 different participants. The lighting conditions of the room were constant for all participants and the level of light in the room was 25 lux, measured perpendicular to the display in the direction of the viewer.

RESULTS

Figure 5 shows the mean ratings for viewing experience averaged over the four images. On the horizontal axis the different noise levels are presented (increasing noise along the hotrizontal axis). The vertical axis represents the averaged values for viewing experience, from bad to excellent. The two lines in the figure represent the dimensions 2D and 3D. Error bars reflect the standard error of the mean.

A multivariate analysis of variance (ANOVA) (with Noise, Image, and Dimension as factors) was carried out on the raw subjective ratings to test the main effects and interactions for statistical significance. The results revealed significant main effects of Image ($F(3,17)=6.413$, $p<0.01$), Dimension ($F(1,19)=5.251$, $p<0.05$), and Noise ($F(5,15)=46.521$, $p<0.001$) on the viewing experience ratings. No significant interactions between Image, Dimension and Noise were found. Figure 5 clearly shows the main effect of a decreasing viewing experience with increased noise level for both 2D and 3D images. The viewing experience of 3D

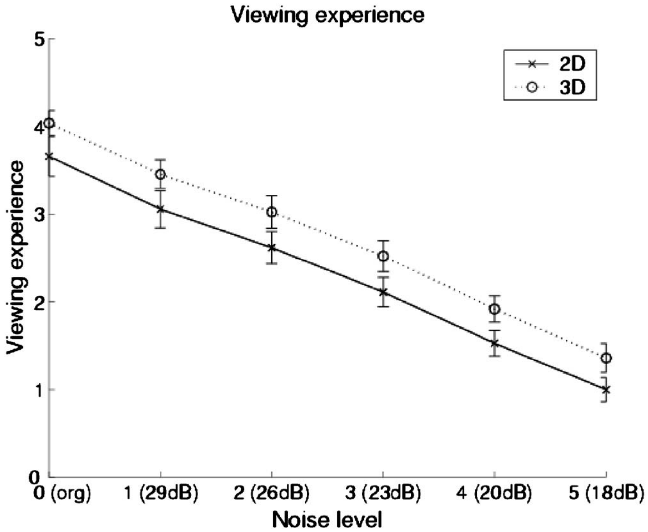

Figure 5. Mean viewing experience ratings averaged over all scenes. The horizontal axis represents the original image (org) and five noise impaired images (PSNR) and the vertical axis represents the subjective ratings for viewing experience. The lines in the figure represent the dimensionality (2D and 3D).

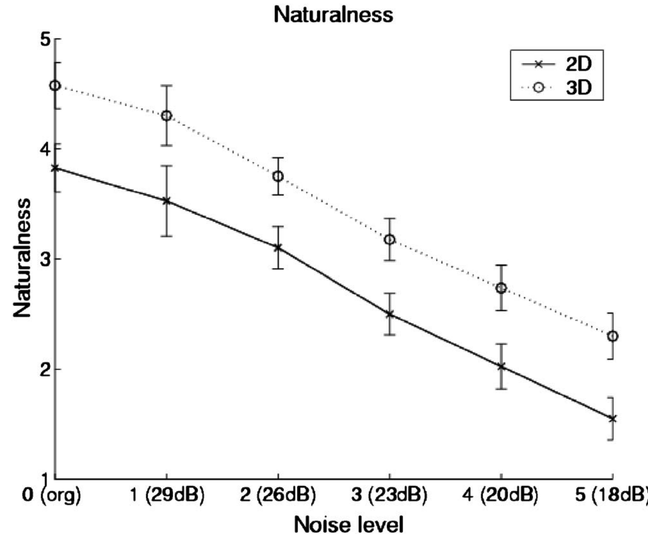

Figure 6. Mean naturalness ratings averaged over all scenes. The horizontal axis represents the original image (org) and five noise impaired images (PSNR) and the vertical axis represents the subjective ratings for naturalness. The lines in the figure represent the Dimension (2D and 3D).

images is rated systematically higher than for 2D images for all noise levels explaining the main effect of Dimension. The main effect of Image was mainly caused by different parallel shifts in the four images, but the main effects of Noise and Dimension were clearly visible in all images. The difference in viewing experience between 2D and 3D for a given noise level is equivalent to a difference in noise level of around 2 dB. Thus, 3D images with 2 dB more noise than their 2D counterparts result in an equal viewing experience. So, the evaluation term viewing experience takes into account the added value of depth, as this is the only difference between the 2D and 3D images.

Figure 6 shows the mean ratings for naturalness averaged over the four images. On the horizontal axis the different noise levels are presented (increasing noise along the horizontal axis). The vertical axis represents the averaged values for naturalness from bad to excellent. The two lines in the figure represent the dimensions 2D and 3D. Error bars reflect the standard error of the mean.

A multivariate ANOVA (with Noise, Image, and Dimension as factors) was carried out on the raw subjective ratings to test the main effects and interactions for statistical significance. The results revealed only significant effects of Dimension ($F(1,19)=9.448$, $p<0.013$) and Noise ($F(5,15)=16.285$, $p<0.004$) on naturalness ratings. No significant interactions between Image, Dimension, and Noise were found for any of the subjective ratings. Figure 6 clearly shows the main effect of a decrease in naturalness with increasing noise level for both 2D and 3D images. The naturalness of 3D images is rated higher than for 2D images for all noise levels explaining the main effect of Dimension. The difference in naturalness between 2D and 3D is around 4 dB, when expressed in an equivalent difference in noise level.

Figures 5 and 6 both show that noise considerably decreases viewing experience and naturalness ratings for both the 2D and 3D cases. Furthermore, both figures show a higher score for the 3D mode than for the 2D mode, which implies that both viewing experience and naturalness take into account the added value of depth. The difference between 2D and 3D is larger for naturalness than for viewing experience, which implies that the added value of depth is taken more into account in naturalness than in viewing experience. The fact that the difference between 2D and 3D ratings remains constant over all the noise levels suggests that the perceived depth is independent of the noise level.

DISCUSSION

Our results show that both viewing experience and naturalness are sensitive image evaluation concepts when it comes to measuring the added value of stereoscopic depth as well as the image quality (in this case noise). Earlier studies demonstrated that when participants are asked to rate image quality in impaired stereoscopic images, the added value of depth is hardly taken into account, if at all. However, when asking observers to assess viewing experience or naturalness, they do not only assess the level of impairment (in our case, the induced noise level), but also other aspects in the image, such as depth, which is illustrated by the fact that there are two distinctive lines for the assessment of 2D and 3D images. So, the added value of depth is taken into account when observers are assessing viewing experience, and even more so when they are assessing naturalness (see Figs. 5 and 6).

The results of the multivariate ANOVA tests show that both Noise and Dimension significantly affect viewing experience and naturalness. For viewing experience Image also had a significant influence (vertical shift of the 2D and 3D line), but the added value of depth as measured by viewing experience was clearly recognized in all four images.

The method applied to quantify the added value of depth expressed in noise level yields an appropriate and useful measure. The potential stereoscopic advantage can thus be quantified in terms of dB noise level. The difference in viewing experience and naturalness between 2D and 3D images expressed in noise level is, respectively, 2 dB and 4 dB. In other words, more noise is allowed in 3D images (respectively 2 dB and 4dB) for an equal viewing experience and naturalness of 2D and 3D images.

The results in Figs. 5 and 6 show a remarkably linear and thus predictable behavior, while being quite stable (low error) within the chosen stimulus set. Apparently observers are well capable of assessing the image impairment and added value of depth in the range used in this experiment.

Thus, quantifying naturalness or viewing experience by means of introducing a controlled impairment, such as noise, and expressing the results in units of this impairment yields a sensitive and reliable metric. Although the first results are very encouraging, more insight into the behavior of viewing experience and naturalness in combination with different types of 2D and 3D artifacts as well as moving 3D material will be needed. Future research could determine to what extent the underlying aspects of viewing experience and naturalness are accountable for the difference in assessment of 2D and 3D images.

ACKNOWLEDGMENT

The authors would like to thank Philips Research Laboratories Eindhoven for the laboratory space and the technical support that made this project possible.

REFERENCES

[1] W. IJsselsteijn, *Presence in Depth*, Ph.D. Thesis, Eindhoven University of Technology, The Netherlands, 2004.

[2] C. Smith and A. Dumbreck, "3-D TV: The practical requirements", *Television: Journal of the Royal Television Society*, 9–15 (January/February 1988).

[3] I. Howard and B. Rogers, *Binocular Vision and Stereopsis* (Oxford University Press, Oxford, 1995).

[4] S. Pastoor and M. Wöpking, "3-d displays: A review of current technologies", Displays **17**, 100–110 (1997).

[5] P. Engeldrum, "A theory of image quality: The image quality circle", J. Imaging Sci. Technol. **48**, 447–457 (2004).

[6] W. Tam, L. Stelmach, and P. Corriveau, "Psychovisual aspects of viewing stereoscopic video sequences", Proc. SPIE **3295**, 226–235 (1998).

[7] P. Seuntiens, L. Meesters, and W. IJsselsteijn, "Perceptual evaluation of stereoscopic images", Proc. SPIE **5006**, 215–226 (2003).

[8] ITU, "Methodology for the subjective assessment of the quality of television pictures", *Recommendation BT.500–10*, 2000.