Performance evaluation of 2D and 3D-TV systems

I. Heynderickx; Philips Research Laboratories; Eindhoven, The Netherlands

Abstract

The Image Quality Circle, introduced by P. Engeldrum [1,2], is a useful framework for modeling image quality of TV-systems. Line-up experiments with high-end 2D TVs show that for naïve viewers the most important attributes in the assessment of overall image quality are brightness, contrast, color rendering and sharpness. When evaluating 3D TV-systems, however, recent research showed that the added value of displaying stereoscopic depth is hardly accounted for in the assessment of image quality. Hence, the Image Quality Circle model needs expansion to cover the full visual experience of 3D-TV. Alternative concepts, such as naturalness and viewing experience, are evaluated on their ability to include both image quality and perceived depth.

Introduction

The Image Quality Circle [1,2], first introduced by P. Engeldrum to conceptually describe image quality for printed image material, is meanwhile a commonly accepted framework to model the relation between technological variables of an imaging system and its resulting image quality. As such, it is also applicable to the description of quality in display systems, including quality related to the display characteristics as well as to the characteristics of the incoming signal.

The Image Quality Circle tries to bypass the problem of determining the relation between the technological variables of a display system and the resulting image quality rating as an endless workflow of changing the variables, making samples and collecting judgments; even more, since knowing the optimal settings of the variables for one imaging system has (almost) no predictive value for the optimization of another imaging system. The model approaches the prediction of image quality for a given set of technological variables as a three-steps process: (1) the overall image quality rating is an often unconsciously weighted sum of image quality attributes, (2) these attributes are determined by the physical characteristics of the light output, and (3) the light output is related to the technological variables of the imaging system (for those who are not familiar with the model, see the first part of figure 6, which is copied from [1,2]). Although this model is conceptually very clear and simple, it is far from trivial in its application. Indeed, the issue of predicting image quality is now translated into two issues: (1) defining the image quality attributes, and how they are weighted in the overall assessment, and (2) determining how each attribute is linked to the physical light output coming from the display screen. In next chapter on 2D quality some first results on the relative importance of attributes are presented.

Recent developments in TV-systems try to go beyond optimizing the reproduction quality of the video material. They rather attempt to enrich the total visual experience. Ambilight TV and 3D-TV are two examples in this direction. In an Ambilight TV (see figure 1), LEDs are mounted around the TV panel, illuminating the wall surrounding the TV. The induced extension of the field of view enhances the feeling of being involved in the displayed video material and creates an atmosphere in the room. In a 3D-TV (see figure 2), the total visual experience is enriched by displaying stereoscopic depth, such that part of the video content lies behind or in front of the display screen. First experiments have indicated that the total visual experience with these enriched TV systems is not fully captured with the concept of image quality. Hence, to be able to predict the performance of these systems in terms of their technological variables, the Image Quality Circle model needs expansion. For a 3D-TV, this observation is further discussed in the chapter on 3D quality.

Figure 1: Picture illustrating the concept of an Ambilight TV

Figure 2: Picture illustrating the concept of a 3D-TV

2D Quality

For the Image Quality Circle, the relation between the overall quality rating and the underlying attributes is needed. This relation is far from trivial, since it is expected to depend on image content, application context, ambient environment, etc. The relative importance of brightness and color saturation, for example, might be different for a computer monitor display (mainly showing data and text) than for a television display (mainly showing video).

This balance might also be different for a television display, viewed in a relatively dark environment, than for a mobile display, viewed in a sun-lit environment.

In a first attempt to model the relation between image quality and the underlying attributes, we focus on the assessment of high-end TVs, which implies that we neglect the occurrence of artifacts, and hence also their effect on image quality. Line-up experiments are used to find out what the most important quality attributes are for relatively naïve viewers [3]. In these experiments four television sets are put on a line (as shown in figure 3) and all display the same image material. The four television sets are hidden behind black boards, such that only the screen area is visible, while the design of the set is covered. The video content displayed on the sets is changed over experimental session. In total, 12 SD and 5 HD video sequences are used. Each sequence is assessed on the four displays by on average 50 participants. They are requested to rank the four displays in order of overall preference, and to explain in their own words why they prefer one set above the other. Assuming that the preference ranking reflects image quality, the comments given by the participants indicate the most relevant image quality attributes for naïve viewers assessing high-end television sets.

Figure 3: Picture of the experimental set-up of a television line-up

For one of the line-up experiments (experiment 1, described in more detail in [3]) the comments of the participants are categorized along attributes by the test leaders. The results per video sequence are given in table 1 for the most important attributes. It should be noted that some participants motivate their choice with more than one comment, belonging to different attributes. That explains why the sum of the cells of a row in table 1 differs from 100%. The main conclusion from table 1 is not surprising: for naïve viewers assessing high-end television sets color rendering in general is the most important attribute, closely followed by brightness, contrast and sharpness. Table 1 also clearly shows that the relative importance of the attributes depends on the image content. For the scene "Ninjas", which is a dark grayish night scene, the attributes brightness and contrast are more important than color rendering. For the scenes "Windmill" and "Eagle", which are highly detailed with mainly natural (hence rather unsaturated) colors, sharpness is the most important attribute.

Of course, these line-up experiments only provide some first qualitative indications of the relative importance of the attributes. To model the relation more quantitatively, the attributes should be independently varied in a controlled way. Then, the first issue is that these four most important attributes are not independent. E.g., any change in the gamma characteristic of the display affects perceived brightness, contrast and color. Hence, a new set of "attributes" has to be defined that can be independently varied. The second issue is related to varying the attributes in a controlled way. To make changes in the attributes comparable in extent, just-noticeable differences (jnds) can be used. But, these jnds are not known yet for natural images viewed under common viewing conditions. First attempts to tackle these issues are described in [4].

Table 1: Number of times (in % of participants) per sequence that a given attribute was decisive in the image quality assessment.

sequence	brightness	contrast	color	sharpness
Horse riding	29	14	***77***	37
Ferriswheel	***45***	13	36	36
Shrek	30	19	***70***	16
Ninjas	***67***	***59***	30	33
Windmill	45	33	31	***55***
Eagle	33	42	44	***60***
Gladiator	33	33	***63***	45
Newsreader	22	30	***67***	45
Beach	39	33	***84***	43
Stairs	35	9	***48***	26
Soccer	23	14	***52***	***45***
Bathing	***59***	8	***58***	24

3D Quality

3D display systems contribute to an enriched visual experience by displaying objects in front of and behind the screen. To optimize this visual experience, it is necessary to know its relation with the technological variables of the 3D display. The visual experience is expected to include "2D quality" aspects (brightness, contrast, spatial and temporal resolution, etc.), "3D quality" aspects (depth resolution, inter-eye crosstalk, etc.) and aspects reflecting the "enrichment" of rendering stereoscopic depth. Understanding the balance between these aspects is particularly relevant for autostereoscopic 3D displays (i.e. systems that reproduce stereoscopic depth without the need of additional means, such as glasses), since in such systems stereoscopic depth can only be rendered at the expense of "2D quality" aspects, as e.g. the spatial and temporal resolution.

Again, the concept of the Image Quality Circle model seems appropriate for the optimization of a 3D display system in terms of its technological variables. Earlier experiments [5,6], however, have shown that image quality ratings hardly account for the added value of having stereoscopic depth in a 3D display. In one of the experiments, two natural scenes are displayed on a mirror stereoscope with different camera base distances, i.e. 0, 4 and 8 cm. A difference in camera base distance implies a difference in the amount of reproduced depth with 0 cm being a 2D reproduction of the displayed content. The images for the left and right eye are then compressed at five different levels (orig, Q30, Q20, Q15 and Q10) using JPEG coding. Ten participants are requested to score the image quality of the stimuli on a 5-points adjective scale. For one of the two images the results are depicted in figure 4 (reproduced from [6]). It clearly shows that the decrease in image quality rating as a function of increased compression level is fully independent of the camera base distance. At any compression level the image quality ratings are equal for the 2D and 3D displayed content. Apparently, the observers base their quality judgment on the visibility of compression artifacts, not accounting for the fact of having stereoscopic depth. This suggests that the term "image quality" is too limited to cover the full assessment of the visual experience of a 3D display.

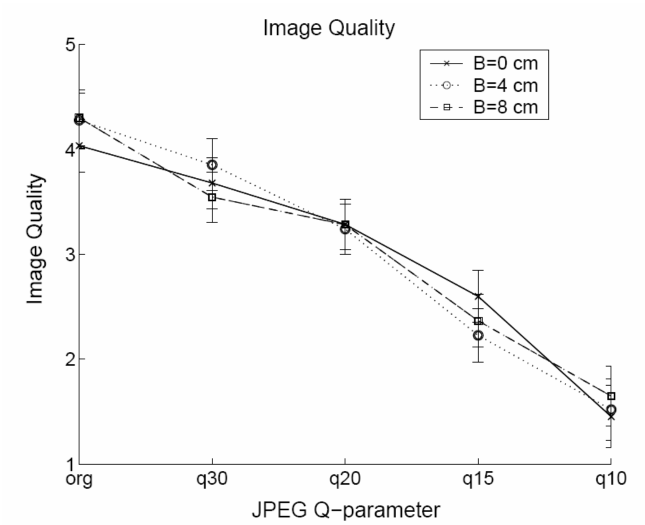

Figure 4: Image quality rating as a function of the JPEG compression level for different camera base distances B. Note that B=0cm corresponds to a 2D display.

Based on this observation, alternative concepts for the assessment of 3D displays are evaluated, such as presence, naturalness and viewing experience. Presence is extensively discussed in relation to 3D and Virtual Reality (VR) displays in [7]. It is shown to be a useful concept measuring the performance of a 3D or VR display, but mainly when using moving image content. Presence ratings are considerably lower when using still image content. Naturalness was first used in relation to the reproduction of colors in natural image content [8,9]. Later, it is also applied in research on stereoscopic displays [6,10]. These studies show a high correlation between the naturalness and image quality ratings, but with a systematic shift between the optima: images with the highest quality are not necessarily the most natural ones. To our knowledge, viewing experience is not used in performance assessments yet, but we expect it to reflect an overall experience, including more than only image quality.

In a series of experiments the terms naturalness and viewing experience in addition to the terms image quality and perceived depth are used to evaluate the performance of 3D displays. Note that presence is not included, mainly since we also use still image content. In these experiments we vary the amount of reproduced depth (by varying the camera base distance), and introduce "2D" artifacts (by adding controlled amounts of white Gaussian noise) and/or "3D" artifacts (by assigning inappropriate depth values to (parts of) objects) to the natural image content. Twenty observers are requested to judge for this image material image quality, perceived depth, naturalness and viewing experience, each on a 5-points adjective scale. One of the results is shown in figure 5, in which the naturalness and viewing experience ratings are given as a function of the level of noise for a 2D and 3D display (for more details see [11]). It clearly shows that both the naturalness and viewing experience ratings decrease as a function of increasing noise level. In addition, both ratings are systematically higher for the 3D mode than for the 2D mode. Hence, both terms apparently incorporate the added value of having stereoscopic depth; and, naturalness does that to a larger extent than viewing experience.

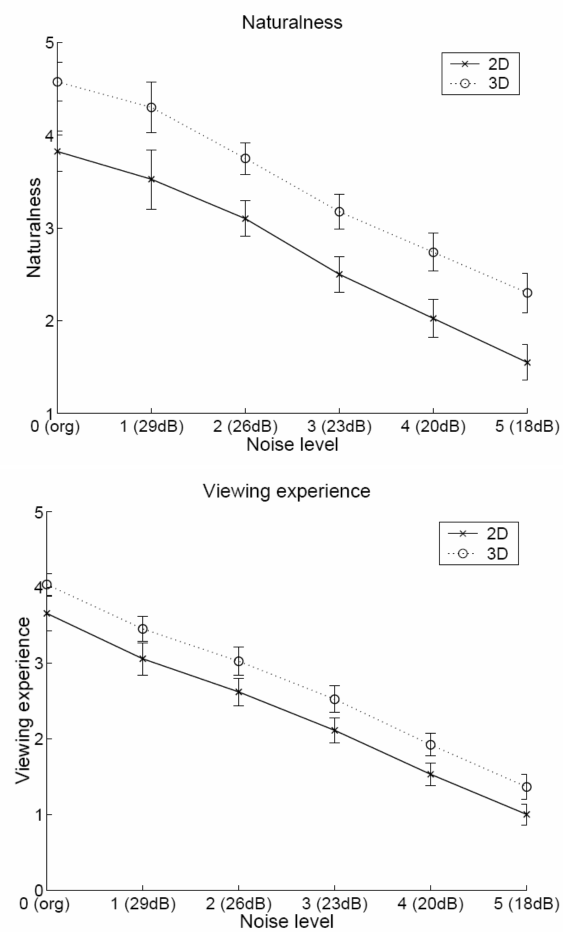

Figure 5: Naturalness and viewing experience ratings as a function of the level of noise in a 2D and 3D display.

This conclusion drawn from one experiment is representative for what we found in the other experiments. In general, our experimental results suggest that the overall visual experience of a 3D display can be modeled as a weighted combination of the assessed image quality (accounting mainly for the visibility/annoyance of artifacts) and the perceived depth. Assuming linear relations as:

Naturalness = α_{nat} IQ + β_{nat} depth,

Viewing Experience = α_{ve} IQ + β_{ve} depth,

typical results for α and β are given in table 2. They indicate that the perceived depth is more accounted for in naturalness than in viewing experience. The high value of α, however, also suggests that for both terms the correlation with image quality is high.

Table 2: Typical values for the relative importance of image quality (α) and perceived depth (β) when assessing naturalness and viewing experience of a 3D display, including the R^2-value of the linear fit.

	α	β	R^2
Naturalness	0.88	0.21	0.98
Viewing experience	0.92	0.12	0.98

Based on these observations, we propose a model that extends the Image Quality Circle as depicted in figure 6. The total visual experience of a 3D display can be assessed by naturalness ratings, which are a weighted sum of image quality ratings and perceived depth ratings. The image quality ratings can be predicted from the technological variables by using the Image Quality Circle as presented in literature [1,2]. This Image Quality Circle does not only account for "2D quality" aspects, such as e.g. noise, compression artifacts, spatial and temporal resolution, but also for "3D quality" aspects, such as e.g. inter-eye crosstalk and depth resolution. An additional block, however, is needed in the overall visual experience model to account for the added value of stereoscopic depth.

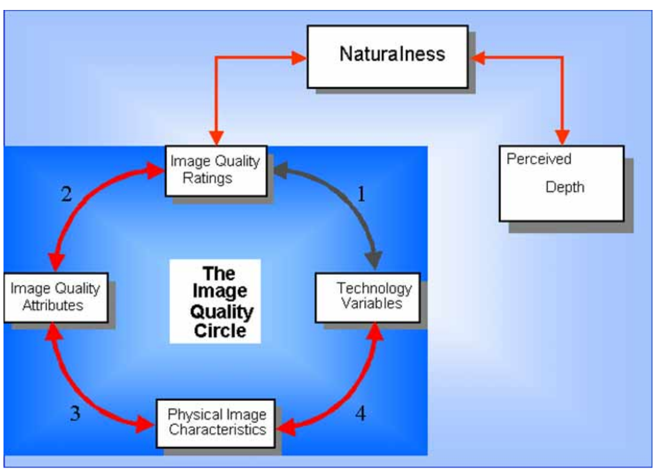

Figure 6: Model proposed for the evaluation of the visual performance of a 3D display

Conclusions

The Image Quality Circle is a useful framework for optimizing the quality of imaging systems, including displays. Knowledge on the relative importance of the image quality attributes is an essential, but not trivial step in this model. Line-up experiments have shown that for naïve viewers assessing the quality of high-end TVs, brightness, contrast, color rendering and sharpness are the most important quality aspects. For a further quantification of the relative importance more controlled experiments are needed.

For the performance evaluation of a 3D display system, the Image Quality Circle needs expansion, mainly because the added value of having stereoscopic depth is not included in the image quality evaluation. A series of experiments has shown that naturalness accounts for both the perceived image quality and the perceived depth. Hence, it is a more balanced concept for the evaluation of the visual performance of 3D displays.

Acknowledgments

The author is very grateful to Dr. P. Seuntiëns and Dr. W. IJsselsteijn of the University of Technology Eindhoven for their close and fruitful collaboration on the research on "3D quality".

References

[1] P.G. Engeldrum, Psychometric scaling: a toolkit for imaging systems development (Imcotek Press, Winchester, 2000).

[2] P.G. Engeldrum, "A Theory of Image Quality: The Image Quality Circle", J. Imaging. Sci. and Technol., 48, 447 (2004).

[3] I. Heynderickx and E. Langendijk, "Image quality comparison of PDP, LCD, CRT and LCoS", SID05 Digest, pg. 1502 (2005).

[4] J. Xia et al. "The just-noticeable difference in chromaticity, black level, white level and contour rendering in natural images", Proc. IDW/AD'05, pg. 1821 (2005).

[5] P. Seuntiëns, L. Meesters and W. IJsselsteijn, "Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation", ACM Trans. App. Perception, in press.

[6] P. Seuntiëns, Visual Experience of 3D TV, PhD. Thesis, University of Technology Eindhoven, The Netherlands, 2006.

[7] W. IJsselsteijn, Presence in Depth, PhD. Thesis, University of Technology Eindhoven, The Netherlands, 2004.

[8] H. de Ridder, F. Blommaert and E. Fedorovskaya, "Naturalness and image quality: Chroma and hue variations in color images of natural scenes", Proc. SPIE Vol. 2411, 51 (1995).

[9] H. de Ridder, "Naturalness and image quality: saturation and lightness variation in color images", J. Im. Sc. Tech. 40, 487 (1996).

[10] W. IJsselsteijn, H. de Ridder and R. Hamberg, "Perceptual factors in stereoscopic displays: the effect of stereoscopic filming parameters on perceived depth and reported eye-strain", Proc. SPIE Vol. 3299, 282 (1998).

[11] P. Seuntiëns et al., "Viewing experience and naturalness of 3D images", Proc. SPIE Vol. 6016, 43 (2005)

Author Biography

Ingrid Heynderickx received her PhD degree in Physics from the University of Antwerp in 1986. In 1987 she joined the Philips Research Laboratories in Eindhoven, where she worked on different research topics. Since 1999, she is responsible for the research on visual perception of display systems within Philips. Since 2003 she is chairman of the Applied Vision committee of the Society for Information Display. In March 2005 she is appointed Guest Professor at the Southeast University of Nanjing and in September 2005 she became part-time Professor at the Department of Media and Knowledge Engineering at the University of Delft.