Perception of detail in 3D images

Ingrid Heynderickx[a,b] and Ronald Kaptein[a]

[a]Philips Research Laboratories, High Tech Campus 34, 5656AE, Eindhoven, The Netherlands
[b]Delft University of Technology, Mekelweg 4, 2628CD, Delft, The Netherlands

ABSTRACT

A lot of current 3D displays suffer from the fact that their spatial resolution is lower compared to their 2D counterparts. One reason for this is that the multiple views needed to generate 3D are often spatially multiplexed. Besides this, imperfect separation of the left- and right-eye view leads to blurring or ghosting, and therefore to a decrease in perceived sharpness. However, people watching stereoscopic videos have reported that the 3D scene contained more details, compared to the 2D scene with identical spatial resolution. This is an interesting notion, that has never been tested in a systematic and quantitative way. To investigate this effect, we had people compare the amount of detail ("detailedness") in pairs of 2D and 3D images. A blur filter was applied to one of the two images, and the blur level was varied using an adaptive staircase procedure. In this way, the blur threshold for which the 2D and 3D image contained perceptually the same amount of detail could be found. Our results show that the 3D image needed to be blurred more than the 2D image. This confirms the earlier qualitative findings that 3D images contain perceptually more details than 2D images with the same spatial resolution.

Keywords: 3D, perception, sharpness, detail, resolution

1. INTRODUCTION

In stereoscopic displays a different view is sent to the left and right eye to create a 3D impression. To generate those views, a loss of temporal or spatial resolution is inevitable. In glasses based systems, two views are generated leading to a factor 2 decrease in resolution, which is often hardly noticeable. In modern auto-stereoscopic 3D displays however, more than 2 views are typically generated to allow for look-around and multi-viewer capabilities (see Fig. 1).[1,2] This leads to a further decrease in spatial resolution.

Another aspect that affects the perceived resolution of 3D displays is interocular crosstalk, due to the imperfect separation of the left- and right-eye image. For large disparities, crosstalk leads to ghost images, but for small disparities it is visible as blur. Especially in auto-stereoscopic systems, crosstalk can be quite high with levels up to 40%.[3] In the course of years, display resolutions of course get higher, making the resolution loss less visible. On the other hand, future auto-stereoscopic display systems may generate more and more views, thereby decreasing or canceling the effect of the native resolution increase.

In order to be able to make a sensible choice for the optimal number of views, the perceived resolution of 3D displays deserves further investigation. Is the perceived resolution of a multi-view auto-stereoscopic display indeed significantly lower compared to the perceived resolution of the underlying panel? Earlier results suggest otherwise. Hakkinen et al[4] recently showed that sometimes 3D videos perceptually contain more detail and that viewers need more time to look at everything. A similar effect can be seen by looking at an image of for example a dense forest or a field of wheat. In case of a 2D image, it is often quite difficult to discern individual wheat plants or leaves. In a 3D image however, this distinguishing is much easier, giving the impression of more detail.

In contrast, Hakkinen et al[4] also reported that some people found 3D videos more blurred than 2D videos. When thinking of blur as the reverse of sharpness, the results of Hakkinen et al[4] suggest that in the case of 3D images, the perceived amount of detail and perceived sharpness are not correlated, and hence, relate to a different perception. As far as we know, perception of detail in 3D images has never been investigated in a systematic manner, but sharpness is. However, the sharpness results are not always in agreement. It has been

Further author information: (Send correspondence to I. Heynderickx)
I. Heynderickx: E-mail: ingrid.heynderickx@philips.com, Telephone: +31 40 2747855

Image Quality and System Performance VI, edited by Susan P. Farnand, Frans Gaykema,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7242, 72420W · © 2009 SPIE-IS&T
CCC code: 0277-786X/09/$18 · doi: 10.1117/12.806186

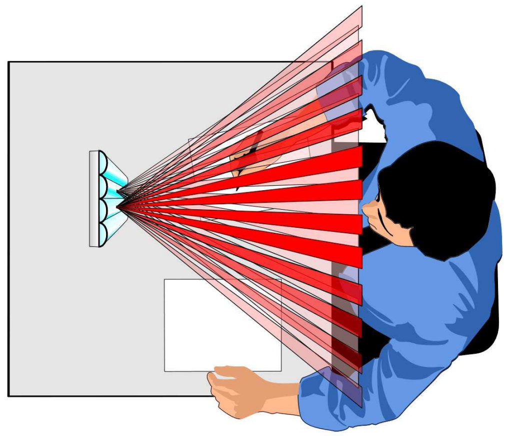

Figure 1. Generation of multiple views in a lenticular auto-stereoscopic 3D display. This allows for multiple viewers and look-around capabilities, but leads to a decrease in resolution.

suggested that stereoscopic pictures should be sharper because position uncertainty is reduced and because of the stereo hyper-acuity effect.[5] Results of Emoto & Mitsuhashi[6] suggest that 3D is indeed slightly sharper than 2D, especially at horizontal edges. They suggest that for vertical edges, de-occlusion problems may lead to a decrease of sharpness in 3D. Tam et al.[7] found no difference in perceived sharpness between 2D and 3D. However, they noted that many subjects reported eye strain, and suggest that double images (diplopia) may have given rise to a decrease in sharpness. This may also be the reason that Hakkinen[4] found that people sometimes found 3D videos more blurred.

In this context also the difference between 2D and 3D image quality is relevant. Image quality has always been a major evaluation criterion to assess the subjective performance of an imaging system.[8,9] For 3D displays however, it has been shown that people do not take the stereoscopic depth into account when judging image quality.[10–13] Something similar may be the case in the judgment of perceived sharpness. Sharpness is clearly an important and well known 2D image-quality attribute.[8] It is therefore conceivable that when people are asked to judge image sharpness, they ignore the stereoscopic depth. This would support the notion that perceived resolution or amount of detail may not simply be the same as perceived sharpness in case of 3D images.

It is therefore not straightforward to assess the issue of perceived resolution in a 3D display using a perception experiment. When asking for sharpness, people may possibly ignore the stereoscopic depth, as is happening with image quality. Since in the study of Hakkinen[4] people often spontaneously mentioned "more details" when watching 3D, we decided to use "perceived amount of detail" or "detailedness" in the assessment task.

To obtain a quantitative and accurate measure for the amount of detail, we decided to determine "detail thresholds" using an adaptive staircase method. The test image was blurred and compared to an unimpaired reference, and subjects had to choose the image which contained the most amount of detail. By varying the blur level, the threshold for which 3D images contained the same amount of detail as 2D images could be found. Results indeed show that the perceived amount of detail is higher in 3D images than in 2D images with the same spatial resolution. Looking for a possible explanation, we tested whether a sampling offset between the left and right eye image may have played a role. This turned out to be not the case. All in all, our quantitative results suggest that for a given display resolution, 3D images contain more perceptual detail than 2D images.

2. METHODS

2.1 Observers

All 15 observers were selected internally within Philips Research. All participants had a visual acuity of > 1 (as tested with the Landolt C test) and a good stereo vision < 30 seconds of arc (as tested with the Randot stereo test). Subjects took about 20 to 30 minutes to complete the experiment.

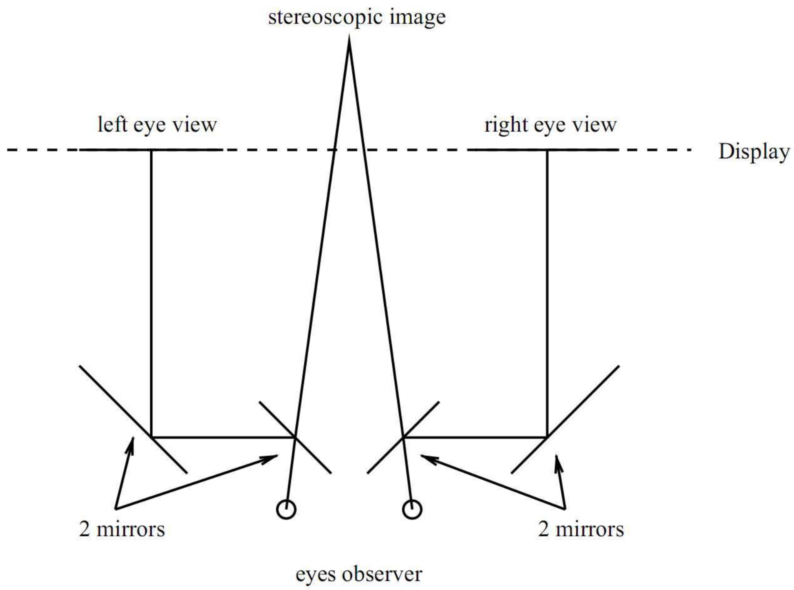

Figure 2. Schematic illustration of the stereoscope used in the experiment.

2.2 Equipment

To present 3D images without any crosstalk to the observers, a Screenscope^{TM} (mirror stereoscope) was used. The stereoscope directs the left-and right-eye image of a displayed stereo pair to the appropriate eye without any crosstalk. The principle of the stereoscope is shown in Fig. 2. In essence, it is just a simple way to increase the interocular distance in order to make it easier to show separate images to each eye. The stereoscope was attached to a 22.2 inch IBM T221 LCD, with a resolution of 3840x2400 pixels and a pixel pitch of 0.1245 mm. The viewing distance from the stereoscope to the LCD was approximately 35 cm. In order to help subjects to maintain a constant viewing position relative to the stereoscope, a chin rest was used.

2.3 Stimuli

Four different scenes were used in the experiment. They can be seen in Fig. 3. Two of the scenes ('watch' and 'phone') were used to test the 3D detail threshold. These scenes were shot from three angles: a center angle for the 2D version, and two off-center angles to create a stereo pair for the 3D version. These stereo angles were chosen in such a way that the 3D effect was strong but comfortable, staying within the 1 degree limit for comfortable viewing.[14] The images had a size of 1200x675 pixels, which equals about 15x8.4 cm on the display.

The 'house' and 'flowers' scenes were used to test the sampling-offset hypothesis mentioned earlier, and were shot with a high initial resolution. Starting from a resolution of 2400x1350 pixels, images were downscaled to 1200x675 by pixel dropping. By shifting the downscaling grid 1 pixel in the initial high-resolution image, two different images were created (see Table 1). In this way, two downscaled images without depth but with a sampling offset of 0.5 pixel were created. We choose not to use low-pass filtering in the downscaling process because then the results would depend on the exact filters used. Close inspection of the final images revealed no aliasing artifacts. Even if present, they would have been almost the same for all images.

Two images were placed side by side at the same height, such that each image was centered in the appropriate eye and the images could be fused. These two images were either two identical images for 2D, a stereo pair for 3D or two 2D images with a sampling offset. After fusing, participants saw two images per trail (one in the upper and one in the lower part of their field of view), one of which was blurred while the other remained unimpaired. Blurring was done with a Gaussian filter of filter size 9x9, with standard deviation σ.

Figure 3. Four different scenes used in the experiment. Clockwise, starting in the upper-left corner: 'watch', 'phone', 'house' and 'flowers'. The 'watch' and 'phone' scenes were available in 2D and 3D. The 'house' and 'flowers' scenes were initial high resolution and were downscaled in order to allow for a sampling offset between the two eyes.

0	1	0	1	0	1
2	0	2	0	2	0
0	1	0	1	0	1
2	0	2	0	2	0
0	1	0	1	0	1
2	0	2	0	2	0

Table 1. Pixel-dropping scheme used to downscale the 'house' and 'flowers' images from 2400x1350 px to 1200x675 px. Numbers include which of the pixels is included in the downscaled images. In this way, two images without depth but with a sampling offset of 0.5 pixel in the downscaled image could be created.

2.4 Procedure

The detail thresholds were determined as follows. Looking through the stereoscope, subjects always saw two images, one on top and one at the bottom. People were then instructed to judge 'the amount of detail' in each image, and choose the one with the most detail. One of the two images in each trial was blurred (test image), while the other one was an unimpaired reference. In each trial, subjects had to indicate which image was perceived as having more detail. By varying the blur level using an adaptive-staircase method,[15] the blur level for which the test image contained the same amount of perceived detail as the reference image could be found (see Fig. 4 for an example staircase). If subjects correctly indicated the unimpaired reference, the blur level of the test image was decreased in the next trial. If they indicated the blurred image as having more detail, the blur level was increased in the next trial. Through this procedure, the 75% threshold level was determined, i.e. the blur level which is perceived as having less detail than no blur in 75% of the comparisons. This level is also refered to as the detail threshold. All staircases began with an easy stimulus (strong blur) and the initial step size was $\sigma = 0.6$px. After the first reversal, step sizes were kept constant at 0.1px (decrease of σ) and 0.3px (increase of σ). All staircases were randomly interleaved to prevent subjects from detecting structure in the stimulus order. Catch trials, i.e. easy trials which did not belong to any staircase, were included at random for the same reason. It was randomly determined which images appeared at the top and which at the bottom.

For all 4 images the threshold for 2D detail was determined, by showing the same two 2D images at the top and at the bottom (2D condition). These thresholds served as a reference. For the 'watch' and the 'phone' scene, also 2D and 3D were compared (3D condition). This led to the blur level for which 3D contained the same

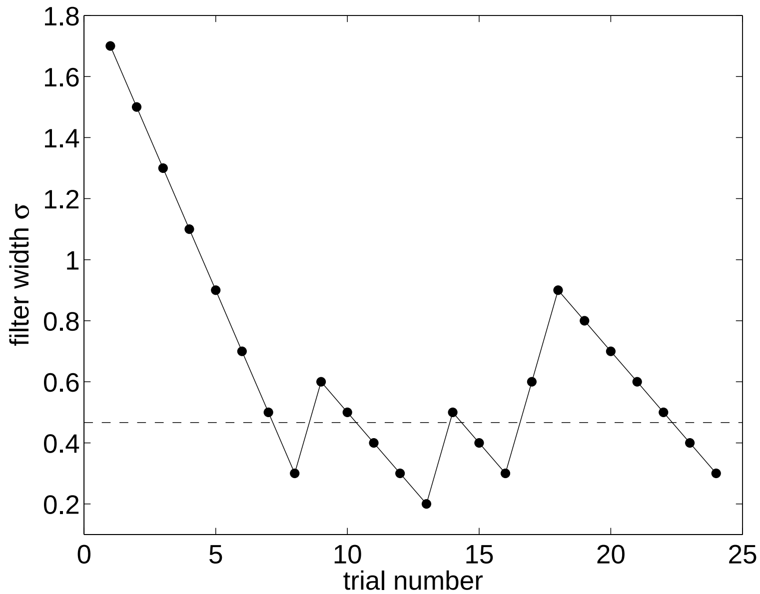

Figure 4. Typical result of a staircase. The dashed line indicates the average of the last 6 reversals. The first reversal is discarded because the step size is halved at that point, after which the step size remains constant.

amount of perceived detail as 2D. For the 'house' and the 'flowers' images, normal 2D (i.e. the same images in both eyes) was compared to 2D with a sampling offset between the two eyes (2D-offset condition). We will refer to these two images as 2D and 2D-offset respectively. So for each scene, two staircases were performed, which lead to 8 detail thresholds per participant.

In the beginning of the experiment, subjects were given instructions and the position of the chin rest was adjusted to the individual. They were shown at least 4 practice images to be able to get used to the stimuli and ask questions to the experimenter if necessary. People sat in a completely darkened room to prevent unwanted reflections from the screen and in the stereoscope, which can be very disturbing.

3. RESULTS

In total, 120 staircases were completed, 8 for each of the 15 subjects. In Fig. 4 the result of a typical staircase is shown. Typically, subjects reported that it was not always easy to assess the amount of detail. In case one of the two images was very blurred, which typically happened in the beginning of a staircase, "amount of detail" was equivalent to sharpness. However, when the sharpness difference was absent or not easy to discern, most people reported that they switched strategy and used a more higher-level concept for detail. Most people found it difficult to say what this strategy exactly was, and did not always report to have a clear preference of 2D over 3D or vice versa.

3.1 Effect of depth

The main goal was to check whether the perceived amount of detail in 3D images differs from that in 2D images. In one staircase, blurred 2D images were compared to unaffected 2D images to find the detail threshold (reference condition). In another staircase, blurred 3D images were compared to unaffected 2D images (test condition). This was done for two scenes. An ANOVA revealed no effect of scene ($p = 0.8$), meaning that we can pool the data from the two scenes. A clear effect of depth (2D versus 3D) was found ($p < 0.01$). Fig. 5 shows the individual thresholds for each condition and also the overall means. The mean for the test condition (0.70) is clearly higher than the mean of the reference condition (0.48). Although no significant effect of scene was found, the results suggest that the effect is stronger for the "watch" scene ($\sigma = 0.77$ for 3D versus 0.43 for 2D) than

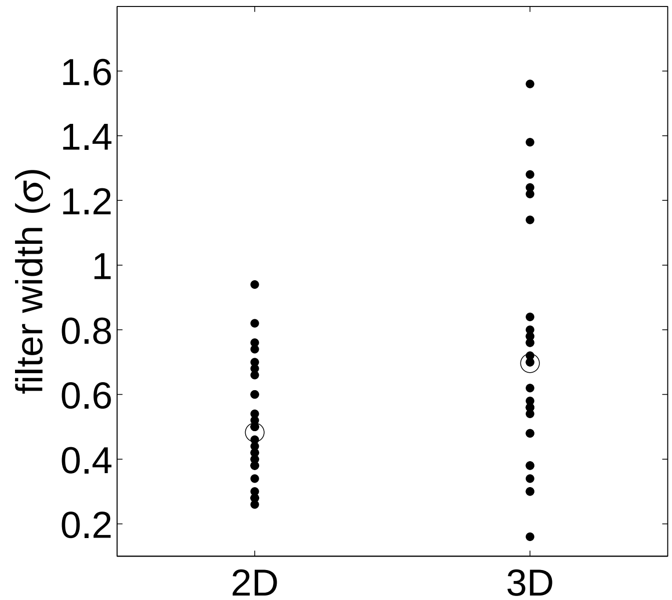

Figure 5. Results showing the 2D (reference) and 3D (test) blur thresholds. Results are pooled over scenes. Small filled circles show results of individual subjects. Large open circle shows the average results over subjects. An ANOVA revealed no significant effect of scene ($p = 0.8$), but a significant effect of depth ($p < 0.01$).

for the "phone" scene ($\sigma = 0.62$ for 3D versus 0.54 for 2D). Since the "watch" scene contained more high spatial frequencies, this is not surprising.

These results suggest that the detail threshold is higher for 3D images, compared to 2D images. This means that a 3D image can be blurred to a higher degree and still contains the same amount of perceptual detail.

3.2 Effect of sampling offset

A possible explanation for the finding that the 3D images contain perceptually more detail than the 2D images is that for the 3D images the left- and right-eye image were sampled differently. To test this, 4 extra staircases were included. Again, in one staircase a blurred 2D image was compared to an unaffected 2D image to find the blur threshold (reference condition). In this condition, both eyes were presented with exactly the same image. In the test condition, one of the images was normal 2D, while the other 2D image was sampled differently for the left and right eye (see METHODS). This was again done for two scenes. An ANOVA revealed an effect of scene ($p < 0.05$), but not of sampling offset, i.e. 2D versus 2D-offset ($p = 0.4$). Fig. 6 shows the results for the 2 scenes separately. It can be seen that the threshold in the test condition is indeed hardly different from the reference condition. For the "house" scene, the means are 0.38 and 0.43 for the reference and test condition respectively, and for the "flowers" scene the numbers are 0.48 and 0.49. Since the "house" scene contains more high spatial frequencies, the slightly lower blur threshold was expected.

4. DISCUSSION

Our results suggest that 3D images contain perceptually more detail than 2D images with the same spatial resolution. Although done with static images, this confirms earlier research that people perceive more detail in 3D videos.[4] Perceived detail can often be considered as equivalent to perceived sharpness, but there are no clear indications that the perceived sharpness of 3D images is higher than for 2D images.[5–7] This suggests that sharpness and detailedness are indeed not the same.

How to translate the blur thresholds in terms of σ to a resolution measure is not straightforward. One possibility is to look to the modulation transfer function (MTF). Resolution can then be defined as the frequency where the MTF drops below a certain limit, for example 10%. Although this may be a correct objective measure

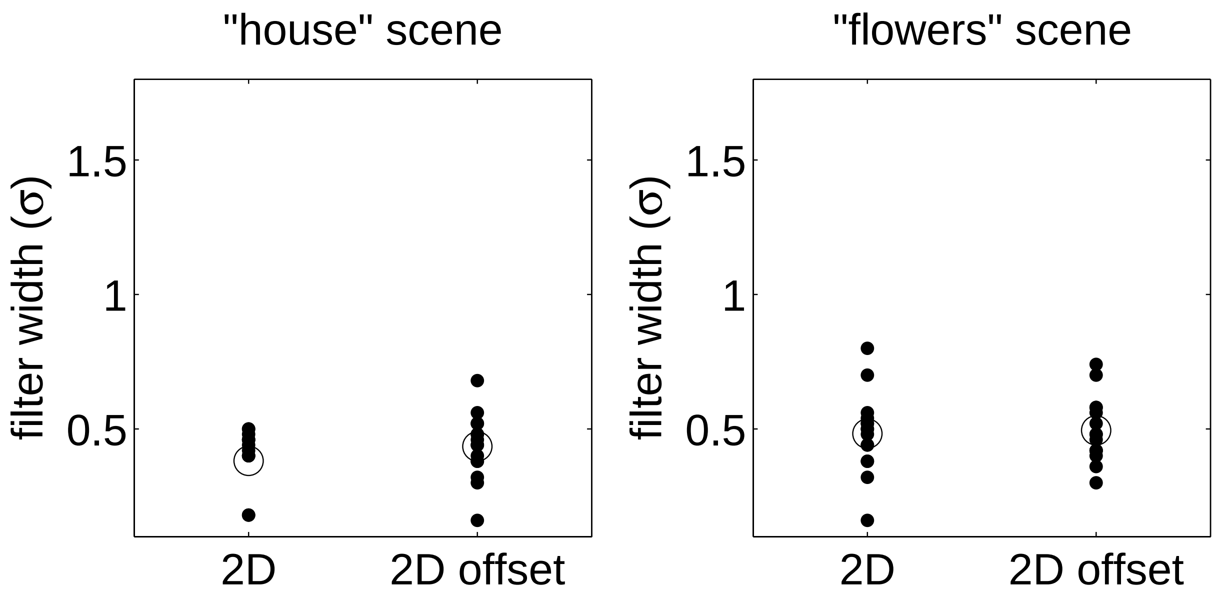

Figure 6. Results of comparison of normal 2D (reference) with 2D where there is a sampling offset between the left- and right-eye image. The two scenes are shown separately because they are significantly different ($p < 0.05$). As in Fig. 5, the small filled circles show results of individual subjects and the two open circle shows the average results over subjects. An ANOVA revealed a significant effect of scene ($p < 0.05$), but not of sampling method ($p = 0.4$).

for resolution, it is not very useful for perception.[16] Images will appear sharper when the MTF has higher amplitude at high frequencies, even if the frequency at 10% amplitude is the same. In other words, images that have the same MTF at for example 10% but a different MTF at other frequencies may lead to a very different perception of sharpness.

An alternative method to find a resolution measure would have been to really downscale large 2D and 3D images to different resolutions, and compare those. However, the exact downscaling method used (e.g. the anti-aliasing filter), has probably a strong effect on the outcome, which would make the results difficult to interpret.

If it is indeed true that 3D images and videos contain more perceptual detail, an interesting question is how that affects people's appreciation of 3D. As said earlier, it has been shown that perceived image quality for 2D and 3D images is usually the same.[10–13] This means that the image-quality circle[8,9] cannot easily be applied to 3D displays. Broader concepts such as presence, naturalness and viewing experience have been suggested[11,12,17] and used to extend the image quality circle (see Fig. 7).[13,18] Our finding that perceived amount of detail is probably higher in 3D than in 2D suggests that detailedness equals sharpness for 2D, but not for 3D.

Since naturalness is most likely positively correlated with detailedness, it would be interesting to investigate what the relation is and how it can be incorporated in Fig. 7. However, because of the strong correlation of detailedness with sharpness and to a lesser extent with depth, it will be difficult if not impossible to vary detailedness independent from these last two attributes. An obvious follow-up study is to more thoroughly investigate the relation between sharpness and detailedness in both 2D as well as 3D images and videos. Sharpness and detailedness appear to be different attributes, but a comparative study using the same setup and under the same conditions should be performed to find a definite answer.

REFERENCES

[1] van Berkel, C. and Clarke, J., "Characterization and optimization of 3D-LCD module design," in [*Proceedings of the SPIE*], **3012**, 179, SPIE (1997).

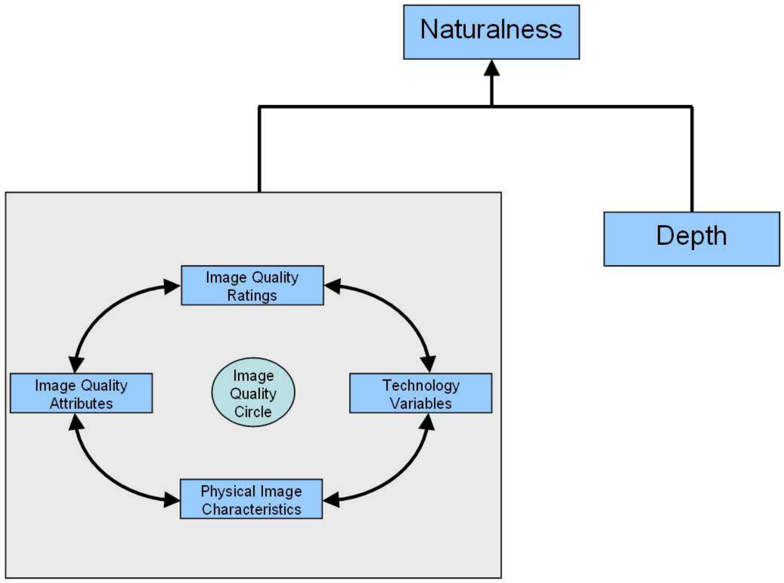

Figure 7. Conceptual extension[13] of the image-quality circle.[8,9]

[2] Willemsen, O. H., Zwart, S. T. D., and Hiddink, M. G. H., "2-D/3-D switchable displays," *Journal of the Society for Information Display* **14**, 715–722 (2006).

[3] Kaptein, R. and Heynderickx, I., "Effect of crosstalk in multi-view autostereoscopic 3D displays on perceived image quality," in [*SID Symposium Digest*], **38**, 1220–1223, SID (2007).

[4] Hakkinen, J., Kawai, T., Takatalo, J., Leisti, T., Radun, J., Hirsaho, A., and Nyman, G., "Measuring stereoscopic image quality experience with interpretation based quality methodology," in [*Proceedings of SPIE*], **6808**, 68081B, SPIE (2008).

[5] Stelmach, L., Tam, W., and Meegan, D., "Perceptual basis of stereoscopic video," in [*Proceedings of SPIE*], **3639**, 260, SPIE (1999).

[6] Emoto, M. and Mitsuhashi, T., "Perception of edge sharpness in three-dimensional images," in [*Proceedings of SPIE*], **2411**, 250, SPIE (1995).

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

[8] Engeldrum, P., "Image quality modeling: Where are we?," in [*PICS Conference Proceedings*], (1999).

[9] Engeldrum, P., "A theory of image quality: The image quality circle," *Journal of Imaging Science and Technology* **48**(5), 447–457 (2004).

[10] Seuntiëns, P., Meesters, L., and IJsselsteijn, W., "Perceptual evaluation of JPEG-coded stereoscopic images," in [*Proceedings of SPIE*], **5006**, 215, SPIE (2003).

[11] Seuntiëns, P., Heynderickx, I., and IJsselsteijn, W., "Viewing experience and naturalness of 3D images," in [*Proceedings of the SPIE*], (2005).

[12] Seuntiëns, P., Vogels, I., and van Keersop, A., "Visual experience of 3D-TV with pixelated ambilight," in [*Proceedings of PRESENCE 2007*], (2007).

[13] Kaptein, R., Kuijsters, A., Lambooij, M., IJsselsteijn, W., and Heynderickx, I., "Performance evaluation of 3d-tv systems," in [*Image Quality and System Performance V., Proc. of SPIE-IS&T Electronic Imaging*], Farnand, S. P. and Gaykema, F., eds., **6808**, 680819 (2008).

[14] Lambooij, M. T., IJsselsteijn, W. A., and Heynderickx, I., "Visual discomfort in stereoscopic displays: A review," in [*Proceedings of the SPIE*], (2007).

[15] Kaernbach, C., "Simple adaptive testing with the weighted up-down method," *Perception & Psychophysics* **49**, 227–229 (1991).

[16] Klompenhouwer, M., *Flat panel display signal processing*, PhD thesis, Technical University of Eindhoven (2006).

[17] IJsselsteijn, W., de Ridder, H., Freeman, J., Avons, S., and Bouwhuis, D., "Effects of stereoscopic presentation, image motion, and screen size on subjective and objective corroborative measures of presence," *Presence: Teleoperators and Virtual Environments* **10**(3), 298–311 (2001).

[18] Heynderickx, I., "Performance evaluation of 2D and 3D-TV systems," in [*Int. Congress of Imaging Science*], (2006).