Geometric and Subjective Analysis of Stereoscopic I3A Cluster Images

Mikko Kytö*, Jussi Hakala
Aalto University School of Science and Technology, Department of Media Technology,
Otaniementie 17, 02150 Espoo, Finland

ABSTRACT

This paper presents a geometric and subjective analysis of stereoscopic versions of close range I3A clusters (Subject-Camera distances below 5 m). The geometry of the stereoscopic pipeline from the scene to the viewer's eyes is a very relevant issue in stereoscopic media. One important factor is the camera separation, because it can be used to control the perceived depth of stereoscopic images. The computational camera separations were compared to subjectively preferred camera separations.

Participants evaluated the strength and naturalness of depth sensation and overall viewing experience from the still images with single-stimulus method. Results showed that participants were able to perceive the change of depth range even though the images were shown in random order without a reference depth scale.

A mild depth sensation was preferred over strong depth sensations. The computational camera separation differed from the subjectively preferred camera separation when the depth range of the scene was narrow. This result indicates that scenes with narrow depth should not be imaged with a long camera separation just to fill the depth budget of the display.

Keywords: Stereoscopic imaging, imaging geometry, visual experience, depth sensation, camera separation, I3A photospace

1. INTRODUCTION

It can be expected that stereoscopic photography will be incorporated in mobile phones in the foreseeable future. In International Imaging Industry Association's (I3A) Camera Phone Image Quality Initiative the typical imaging conditions are presented as six clusters in the Photospace [1]. These six clusters are calculated to cover 70 % of typical imaging conditions, so investigating the stereoscopic imaging geometry according to them is justified.

Quite much of the research on S3D imaging has focused on imaging geometry and its effect to perceived depth. However in many studies [2, 3 4, 5] the content has based on compositions of objects instead of typical imaging conditions. In this study the natural real-world content is based on three I3A clusters. In addition, one composition image with geometric objects is included to subjective and geometric analysis.

Four research questions are presented:

1. Can the participants sense the change of depth range between S3D images? When the camera separation increases the depth range increases. The aim is to follow how the participants sense this change.

2. How strong depth sensation is preferred? Depth range is varied between light and strong sensation so that the preferred depth sensation can be resolved.

3. How the computational camera separation differs from the subjectively preferred values? The geometry of scene, camera and viewing condition are known so the perceived depth can be controlled [6]. The computational camera separation is compared to subjectively preferred values.

4. Is there difference in subjective evaluations between I3A clusters? As the evaluations are made with four different contents there can be found differences for subjective evaluations.

The geometry of the stereoscopic pipeline from a scene to the viewer's eye is a very relevant issue in stereoscopic media. One important factor is the camera separation, because it can be used to control the length of the disparity range [6]. Limiting the disparity range is essential for success of stereoscopic media. Too long disparity ranges cause diplopia and too large mismatch between accommodation and convergence.

*mikko.kyto@tkk.fi; phone +358947023348; http://media.tkk.fi/en/

The length of depth range that is in viewer/display space is called depth budget (Figure 1). It is less in front of the screen than behind the screen if selected according to 1 degree rule [7]. This comfort disparity range varies between users, but for most cases it can be expected to be small enough for comfortable viewing. The appropriate depth budget depends on the viewing distance and the viewer's interpupillary distance.

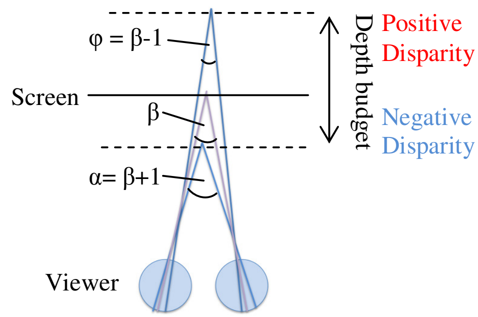

Figure 1. The depth budget according to one degree rule. Image not in scale.

Basically there are two different configurations to capture the depth with stereo cameras: parallel and toed-in configuration [8]. Parallel configuration uses built-in sensor shift or images are shifted afterwards. In toed-in method the cameras are physically rotated towards each other. Parallel configuration is preferred because the vertical disparity can be avoided.

2. I3A CLUSTER IMAGES

The typical use cases of mobile phone cameras are presented as Clusters in the I3A photospace [1]. The photospace is a statistical distribution of frequency of picture taking as function of two variables: Illumination and Subject-Camera distance. In stereoscopic imaging the interesting dimension is the Subject-Cameras distance as it affects the geometry of stereoscopic pipeline. Roughly the photospace can be split in two Subject-Camera camera distance ranges, distances between 0.5-5 m and distances over 5 m. The former distance range was selected for this study (The depth ranges of the scenes are presented in Table 1). The illumination factors were left out, because the focus was on geometric factors.

The chosen Subject-Camera distance range includes indoor Clusters (1-3) (Figure 2). The Subject-Camera distances for outdoor clusters can be much higher (over 5 m) so they might require larger camera separations than was possible to achieve with the used equipment. The first cluster is one person sitting in a front of dark background, the second cluster is one person in a living room drinking wine, the third one is persons playing a game in a living room. The fourth image is a composition of objects. The composition was modified to include more geometric objects; a bicycle tire, volleyball and a self-made depth test target with planes at different depth levels. The idea of these geometric objects was to help the user to make judgments of depth sensation and to detect geometric deformations from the S3D image.

Figure 2. Clusters 1-3 in the order following order: "Bar", "Wine" and "Game". The rightmost is the composition. On the first row there are examples of cluster images taken with mobile phones and on the second row there are images used in this study.

3. MODEL OF 3D "VISUAL EXPERIENCE"

Image quality is very difficult issue to define and research. The research related to it has been wide and there are numerous image quality models [e.g. 9]. So the investigations considering stereoscopic visual experience can be said to be an even more difficult problem to research. In this report one existing stereoscopic visual experience model (Figure 3) is used [10]. The model is interesting because it raises the "Naturalness" –attribute above "Image quality". In 2D image quality the naturalness is considered to be one attribute of image quality [11], not the other way round. The emphasis on the naturalness attribute can be expected to be important in S3D images, because in S3D images there can appear unnatural phenomena (like card board and puppet-theater) that don't appear in 2D images. On the other hand naturalness can refer to "Life-like"-experience which is commonly mentioned when describing S3D images [12].

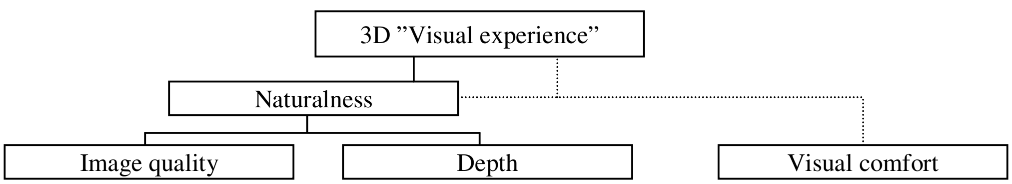

Figure 3. 3D "Visual experience" model from Seuntiens [10]

The attribute selection for the subjective test is based on this hierarchical model of 3D "Visual experience". The "low" level attribute was "Strength of depth sensation". It was selected because there was a need to follow how the user experiences the changes of depth range. Strength of depth sensation is supposed to increase as a function of camera separation, so it would work as a control attribute to follow the magnitude of perceived depth sensation.

The "Naturalness of depth sensation" attribute was selected to follow the Naturalness -attribute in the 3D "Visual experience" –model. It is a "higher" attribute than the "Strength of depth sensation" as it measures the naturalness of the depth sensation not just its amount. The "highest" attribute that was selected for the subjective tests was "Visual experience" as in Seuntiens' model. It was defined as the overall sensation of the S3D image related to how easy and comfortable the viewing experience was.

4. EXPERIMENTAL SETUP

4.1 Shooting and scene conditions

Images of the Cluster scenes (Figure 2, bottom row) were taken with a self-made stereo camera. The stereo camera consisted of two Canon Mark II digital cameras with 50 mm lenses. Table 1 summarizes the imaging and scene parameters for the selected Clusters. Lighting conditions didn't correspond to I3A specifications, because images were taken without a flash. Also because of the lighting conditions the DOFs of images were quite limited especially in Cluster 2 (Wine). The computational camera separation is computed according to [6] and depth budget selected according to the one degree rule.

Table 1. The imaging and scene parameters.

	Cluster 1 (Bar)	Cluster 2 (Wine)	Cluster 3 (Game)	Composition
Varied camera separation (cm)	2, 4, 6, 8, 10	0, 2, 4, 6, 8, 10	2, 4, 6, 8, 10	2, 4, 6, 8, 10
Convergence distance using shift (cm)	120	150	200	200
Focal length (mm)	50			
Aperture (f-number)	4	1,8	4	5,6
Scene Depth range (cm)	70 – 150	100 – 300	180 – 500	200 – 300
Object distance according to I3A	≈ 100	≈ 100	>400	
Computational camera separation (cm)	4,7	5,4	10,0	23,0

4.2 Viewing conditions

The display was autostereoscopic Tridelity SL2400. Image height was approx. 33 cm and viewing distance was approx. 80 cm. Participants used a chin support to avoid artifacts from head movements. The tests were done in normal office lighting (~100 lx, ~3000K).

4.3 Participants

Twelve participants (3 female, ages between 22-34years) attended the tests. Participants had normal or corrected to normal vision. The stereo acuity of the participants was tested with TNO stereo vision test.

4.4 Test procedure

In the beginning the participants had training with 8 images. The images of contents taken with 2 cm and 10 cm baselines were shown to the participant to help her/him to adapt to the depth range. After training session each image was shown two times to the participants in random order. One 2D image was added to Cluster 2 as a "hidden reference". The total amount of unique images was 21 (= contents × camera separation variations + hidden reference).

Participants evaluated the *strength* and *naturalness of depth sensation* and overall *viewing experience* with scale 1-7, where 7 represented the highest score. Participants had also the possibility to write what part of the image they perceive natural and why. The user interface was shown below the autostereoscopic display and the participants gave the opinion scores with mouse and qualitative evaluations with keyboard. The tests lasted for 25-70 minutes per participant, mean was 46,5 minutes.

5. RESULTS

5.1 Mean opinion scores

The numeric results of the experiments are shown in Figure 4. The strength of depth sensation increases according to camera separation as expected. However this result wasn't that clear because the images were evaluated with single-stimulus method, so the viewer couldn't compare the amount of depth sensation to the other stimuli. It has been shown that humans adapt to different depth scales quite easily if a reference depth scale is available [12]. In another study [13] the participants were able to sense the increased depth but the used differences in camera separations were higher. The interesting result is that the strength of depth sensation seems to be strong even though the image is shown only behind the screen and the disparity range is short. This phenomenon can be seen by comparing Cluster 1 and Composition. The strength of depth sensation in Composition (only positive disparity) has been evaluated equally high with Clusters 1 even though its disparity range is shorter than in Cluster 1 (both positive and negative disparities). In Cluster 2 the hidden-reference, the 2D image, was detected and was perceived with the lightest depth sensation.

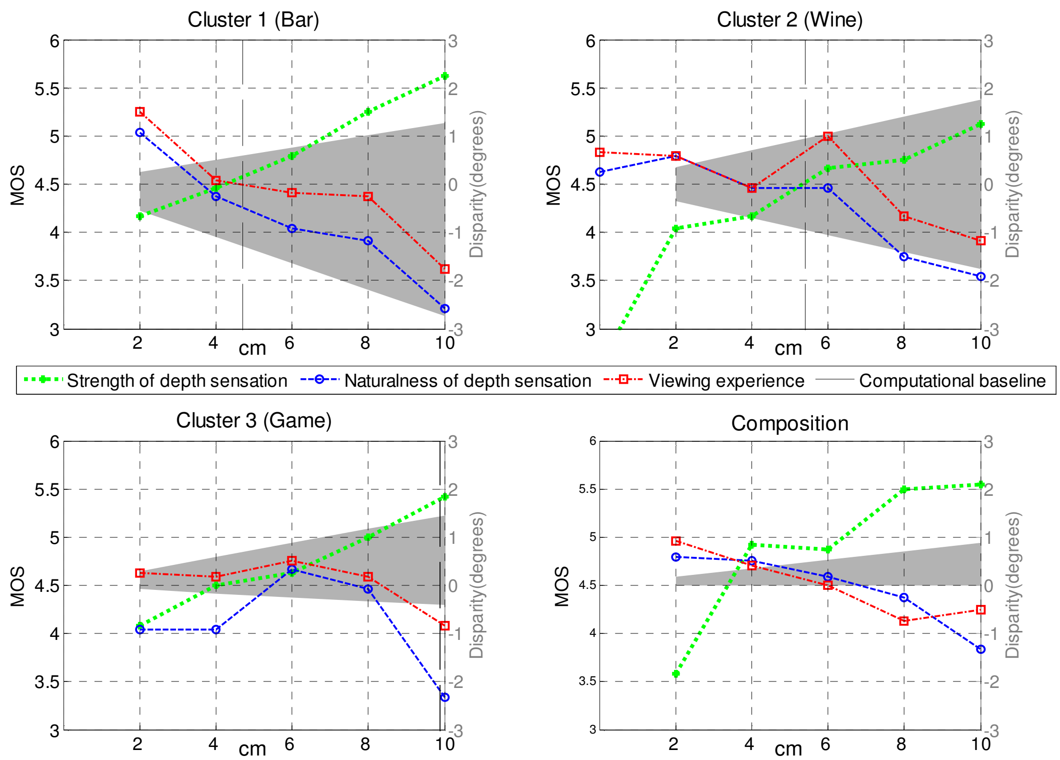

Figure 4. The MOS-values according to camera separation and disparity ranges. The disparity range (scale on right y-axis) is calculated according to [8] and shown as gray area. The mean errors for MOS-values are presented without error bars for clarity. Mean errors are close to one for every attribute.

The naturalness of depth sensation decreases after 6 cm camera separation in every content. Same kind of trend can be seen in another study [3]. In scenes with narrow depth variation (Cluster 1 and Composition) the naturalness of depth sensation decreases according to length of disparity range. In Cluster 3 there is more depth variation in the scene and the highest naturalness was achieved with a longer camera separation than in Cluster 1 and Composition.

Viewing experience changes also according to camera separation, but the impact of camera separation to viewing experience is less than to other attributes. For example in Cluster 3 viewing experience is quite constant. The viewing experience behaves same way as naturalness of depth sensation. It decreases when the disparity range increases with scenes with narrow depth variations (Cluster 1 and Composition). The limits for comfortable viewing limits are exceeded most in Cluster 1 and the negative disparity is almost three degrees with the longest camera separation. It seems that the viewing experience is high when the disparity range is between -1 to 1 degree as expected from the one degree rule.

Overall there are strong correlations between attributes. There is correlation between viewing experience and naturalness of depth sensation ($r = .87$). And both of them have negative correlation with strength of depth sensation ($r = -.60$ and $r = -.68$ respectively). All of the correlations are statistically significant ($p < .01$).

There existed a shift between quality of depth and naturalness of depth in other studies [2, 3] with higher camera separations. This similar shift can be seen with viewing experience and naturalness of depth in Clusters 1, 2 and 3. In these contents the naturalness of depth sensation is less than viewing experience with higher camera separations.

Computational camera separations were quite close to subjectively preferred values in Clusters 1 and 2. Instead computational camera separations for Clusters 3 and Composition were much higher than subjectively preferred values. The convergence distances weren't adjusted according to computations from [6] which has affected the location of disparity range. In Cluster 1 the convergence distance is farther than the computational value and in Cluster 3 and Composition the convergence is nearer than the computational value. The effect of convergence distance inside the content wasn't investigated in this study and could be added for future studies.

5.2 Effect of participants experience

Quite mild depth sensation was preferred in the test as can be seen from the MOS-values above. However the participants' experience had an effect on average evaluations (Figure 5). It seems that the more experience with stereoscopic images the more natural the depth sensation is perceived. There is a linear correlation between experience with stereoscopic images and the naturalness of depth impression($r = .58$, $p = .05$) as well as viewing experience ($r = 0.57$, $p = .05$). Also the strength of depth sensation increased according to experience with stereoscopic images, but the effect wasn't as strong as with other attributes ($r = .45$, $p = .14$). The amount of participants was low in each group (Shown as bars in Figure 5) so the results can be considered indicative.

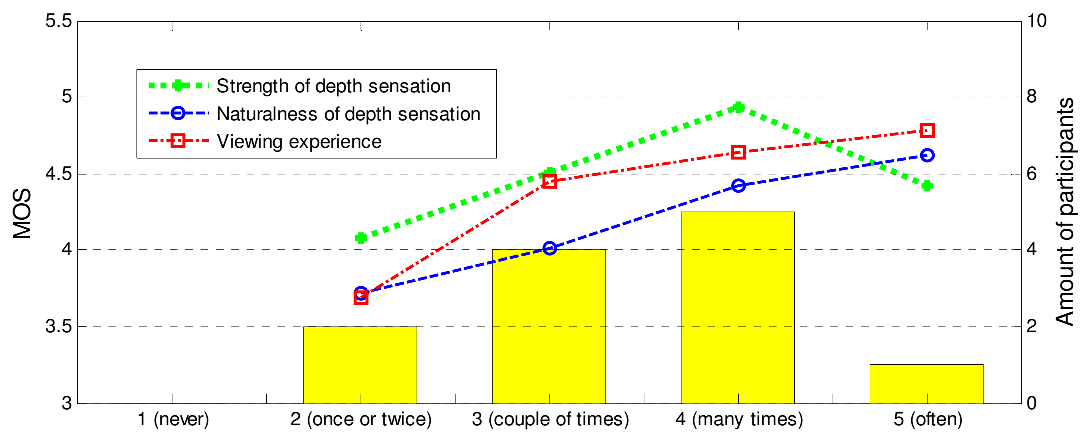

Figure 5. The effect of experience with stereoscopic images to the evaluations.

5.3 Comments from the participants

The participants had possibility to comment on what part of the image they perceived natural or unnatural. In Cluster 1 the most comments were about the face, the hands and the glasses. Mostly positive comments were about the naturalness of the person and negative comments were about the foreground glass and table with longer camera separations. In Cluster 2 the most comments considered the limited depth-of-field, which was perceived distracting. Especially the unsharp background was commonly mentioned to be annoying. The positive comments were mostly about the natural appearance of the person and wine glass. In cluster 3 the negative comments considered mostly cardboard effect on the person on right, even though the other persons were not considered to be "cut from other image". The cardboard effect might have pose sensitive characteristics as the person on the right is imaged straight from the side. The positive comments in the Cluster 3 were mostly about natural proportions between objects. From the Composition the roses were mentioned often and were said to be very natural. The geometric objects were also used in evaluations: the volley ball and tire were used to evaluate the flatness of depth. The self-made test target revealed crosstalk as it included edges with high contrast.

6. DISCUSSION

The results are actually quite positive for stereoscopic content production. The good viewing experience and naturalness of depth were gained even with short camera separation (2-4 cm). With these camera separations the visual discomfort from mismatch between accommodation and convergence is unlikely a problem. However the crosstalk might have affected the results. Crosstalk is reported to distort the image more at higher camera separations [14] which might favored shorter camera separations. Measuring and reporting crosstalk with subjective evaluations would be important to do in the future.

Another important issue in the future would be to investigate the effect of on-line changes in camera separation to get understanding how participants adapt to new depth scale during the test [12]. This memory effect might have a strong effect to the subjective results. Showing a 2D image between stimuli should be included in future studies to give participants a resting phase.

In this study the photospace defined for 2D images was used, even though the possible differences between 2D and S3D photospaces would be an interesting issue to research. What kind of S3D images mobile phone users would really take? As the S3D digital cameras enable also the range metering it would be possible to investigate what subject to camera distances occur typically. At first glance it seems that people tend to take S3D images to emphasize the depth by taking S3D images where there are objects at the different depths. So the emphasis on depth will likely emerge from scene instead of wide camera separation.

7. CONCLUSIONS

Results from the subjective tests show that the strength of depth sensation increased as a function of camera separation, which was expected based on the geometry of stereoscopic imaging. The participants were able to perceive the change of depth scale even though the images were shown in random order without a reference depth scale. The depth range behind the screen was perceived strong even when the disparity range was below one degree.

The subjective attributes correlated with each other. The strength of the depth sensation correlated negatively with naturalness of depth sensation and viewing experience and there was a positive correlation between naturalness of depth sensation and viewing experience. All correlations are statistically significant ($p <.01$). The viewing experience of the S3D images seems to be closely related to naturalness and especially naturalness of depth sensation as shown in previous studies.

The analysis showed differences between clusters for preferred depth sensation. Light depth sensation was preferred in Clusters where the depth is narrow. In Clusters 1 and 2 the computational camera separation corresponded to subjective optimum camera separation, where as in Clusters 3 and Composition the computational camera separations were much longer than subjectively preferred values. It seems that the depth sensation has to come from the scene itself not from long disparity range.

RERERENCES

[1] International Imaging Industry Association, "CPIQ Initiative Phase 1 White Paper", Available from http://www.i3a.org/, (2007).

[2] IJsselsteijn, W.A., et al., "Perceptual factors in stereoscopic displays: the effect of stereoscopic filming parameters on perceived quality and reported eyestrain," Proc. SPIE 3299, 282-291 (1998).

[3] IJsselsteijn, W.A., et al., "Subjective evaluation of stereoscopic images: Effects of camera parameters and display duration," IEEE Transactions on Circuits and Systems for Video Technology 10(2), 225-233 (2002).

[4] Seuntiëns, P., et al., "Perceptual evaluation of JPEG coded stereoscopic images," Proc. SPIE 5006, 215-226 (2003).

[5] Seuntiëns, P., et al., "Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation," ACM Transactions on Applied Perception 3(2), 95-109 (2006).

[6] Jones, G., et al., "Controlling perceived depth in stereoscopic images," Proc. SPIE 4297A, 42-53 (2001).

[7] Lambooij M., et al., "Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review," Journal of Imaging Science and Technology 53(3), (2009)

[8] Woods, A., et. al., "Image distortions in Stereoscopic Video Systems," Proc. SPIE 1915, 36-48 (1993).

[9] Engeldrum, P., "Image Quality Modeling: Where are we?," IS&T's PICS Conference, 251-255 (1999).

[10] Seuntiëns, P.J., "Visual experience of 3D TV," Doctoral dissertation, Technische Universiteit Eindhoven , (2006).

[11] de Ridder, H. and S. Endrikhovski, "Image Quality is FUN: Reflections on Fidelity, Usefulness and Naturalness," SID Symposium Digest of Technical Papers 33, 986-989 (2002).

[12] Häkkinen, J., et al., "Measuring stereoscopic image quality experience with interpretation based quality methodology," Proc. SPIE 6808, (2008).

[13] Milgram, P. and Krueger, M., "Adaptation effects in stereo due to on-line changes in camera configuration," Proc. SPIE 1669, 122-134 (1992).

[14] Seuntiëns, P., et al., "Perceptual attributes of crosstalk in 3D images," Displays 26(4), 177-183 (2005)