Interestingness of stereoscopic images

Jussi Hakala*, Mikko Nuutinen, Pirkko Oittinen
Aalto University, Department of Media Technology, Otaniementie 17, FI-02150 Espoo, Finland

ABSTRACT

The added value of stereoscopy is an important factor for stereoscopic product development and content production. Previous studies have shown that 'image quality' does not encompass the added value of stereoscopy, and thus the attributes naturalness and viewing experience have been used to evaluate stereoscopic content. The objective of this study was to explore what the added value of stereoscopy may consist of and what are the content properties that contribute to the magnitude of the added value. The hypothesis was that interestingness is a significant component of the added value. A subjective study was conducted where the participants evaluated three attributes of the stimuli in the consumer photography domain: viewing experience, naturalness of depth and interestingness. In addition to the no-reference direct scaling method a novel method, the recalled attention map, was introduced and used to study attention in stereoscopic images. In the second part of our study, we use eye tracking to compare the salient regions in monoscopic and stereoscopic conditions. We conclude from the subjective results that viewing experience and naturalness of depth do not cover the entire added value of stereoscopy, and that interestingness brings a new dimension into the added value research. The eye tracking data analysis revealed that the fixation maps are more consistent between participants in stereoscopic viewing than in monoscopic viewing and from this we conclude that stereoscopic imagery is more effective in directing the viewer's attention.

Keywords: interestingness, added value, stereoscopic, image quality, attention, recalled attention map, RAM, eye tracking

1. INTRODUCTION

1.1 Stereoscopic image quality and interestingness

Stereoscopic imaging technologies have reached a level of quality where their adoption to mainstream applications is possible. This makes it important to know how well different contents are suited to stereoscopic viewing, i.e. how high the added value of stereoscopy is. Previous research [1] has shown that when comparing stereoscopic and non-stereoscopic imagery the test participants prefer the stereoscopic images over the non-stereoscopic ones if the number of stereoscopic artefacts is small. It is not clearly known what this added value consists of. It is known that 'image quality' does not adequately encompass the added value of stereoscopy and hence the attributes naturalness and viewing experience have been used to evaluate stereoscopic images [2], but the question of what the dimensions of the added value are remains open.

Interestingness is the quality of being interesting, and interesting is defined in the Oxford English Dictionary [3] as the "*qualities which rouse curiosity, engage attention, or appeal to the emotions*". In this definition we find that interestingness is tightly coupled with attention and appeal. Assuming that pleasingness is a cause of appeal, the connection of visual appeal and interestingness is studied thoroughly in [4], where the authors find that interestingness and pleasingness are highly correlated and that they are influenced strongly by the complexity of the stimuli, specifically the main dimension of complexity, "information content". From the point of view of stereoscopic image quality this is an important connection, as we know that stereoscopic imagery inherently contains more information than the same content in monoscopic form. Another connection between image appeal and interestingness is made in [5], where the authors define image appeal in consumer photography as "*the interest that a picture generates when viewed by third-party observers*". They present the results of a first-party evaluation, which shows that the primary reason for consumers discarding a photograph from their photo album is that the subject was "not interesting". The results of their third-party evaluation study indicate that low level image quality measures are secondary to the high level attributes such as composition, when assessing image appeal.

*Correspondence: jussi.h.hakala@aalto.fi, tel. +358 9 470 23398

In previous research it has been suggested that image quality consists of three dimensions: fidelity, usefulness and naturalness [6]. In contrast to the reasoning in [7], where the authors simply discard the usefulness dimension in the case of consumer photographs, we here propose that the usefulness of a consumer photograph is closely related to the interestingness of the image. Amateur and professional photographers have during the past decade increasingly shared their works on a plethora of social and photo sharing sites, where, by definition, their purpose is to share their work, often with as large an audience as possible either from their family and friends, or from a larger community. This backs our proposed usefulness-interestingness equivalence, as interesting images draw more audience, thus making the images more useful for the purpose of attracting audience.

In this exploratory work we examine the relationship of interestingness with previously used attributes in stereoscopic image quality assessment. Based on the works presented above, our hypothesis 1 is that interestingness is significantly higher for stereoscopic content. The attributes are studied in a subjective evaluation experiment.

1.2 Binocular visual attention

As interestingness is closely related to attention, we will examine attention in more detail, specifically binocular visual attention. The research field of attention modelling is extensive, see [8] for a review. Attention arises from the bottom-up features of visual stimuli as well as the top-down processes [9]. Examples of the bottom-up salient features that draw attention include areas with higher luminance and texture contrast values compared with their surround. On the other hand top-down processes such as recognizing faces are a dominant attractor of attention [10]. Stereoscopic attention in natural stimuli has only been studied recently. Jansen et al. [11] revealed that contrary to intuition, areas with high disparity contrast do not receive significantly more fixations than flat areas. Furthermore, they found that the scan path correlates with the distance of the fixated objects in stereoscopic images. An explanation based on evolution is given, as the selection-for-action effect on the other hand helps us react to close objects that might be dangerous and on the other hand helps us discover objects within reach. This was found to be true also in monoscopic images, but disparity strengthened the phenomenon. A more recent study revealed an even more interesting phenomenon: disparity discontinuity in fact may repel attention [12]. According to the authors, this could be due to the visual system's tendency to avoid regions that require difficult disparity computations, for example in regions where occlusion occurs. We suggest that even though the actual disparity contrast does not attract attention, the disparity contrast plays an important role in the saliency of the object that it surrounds.

In the second part of our study we take our exploration of interestingness to the spatial domain and study the differences in spatial attention in natural monoscopic and stereoscopic photographic stimuli. Our hypothesis 2 is that the attention shift is significant, as we expect that there will be a sizeable shift towards objects that are easily discriminable from their background in depth, i.e. the objects that are surrounded by disparity discontinuities. Attention is studied in an eye tracking experiment.

2. MATERIALS AND METHODS

2.1 Stimuli

The stimuli were acquired with a stereo camera rig specifically built for this study. The rig was built so that two SLR cameras can be mounted on it perpendicularly and that the photographer can adjust the camera separation with ease from 0mm to 120mm. A high quality two-way mirror was used as a beam splitter in the rig, so that the left camera is moveable and the rays from the scene enter the sensor of the left camera through the mirror, whereas the right camera is immovable and it records the reflections of the scene from the mirror. Two synchronized Canon EOS 10D cameras equipped with a Canon 50mm f/1.8 lens were used for all scenes. The photographs were taken during Q4/2009.

Thirteen scenes (Figure 1) were selected so that the depth distributions in typical consumer photography situations and several levels of complexity would be covered and that there would be equal numbers of scenes with and without people, since people's faces, are a known dominant attractor of our attention [10]. Three camera separations were used in addition to the monoscopic stimuli. This was done in order to increase the likelihood that all of the scenes would be captured at a close-to-optimal camera separation. The three camera separations used were 30mm, 60mm and 120mm.

The selection was based on expected depth ranges of the scenes and the viewing geometry. This totals 52 stimuli, see Figure 2 for two stereoscopic samples.

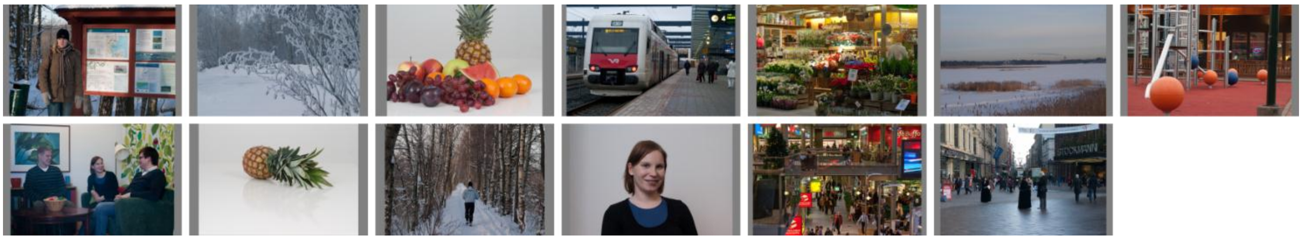


Figure 1. The left side images of the 13 stereo pairs at camera separation 30mm. Left to right, top to bottom: 1. Tourist, 2. Branch, 3. Fruit, 4. Train, 5. Florist, 6. Bay, 7. Balls, 8. Home, 9. Pineapple, 10. Runner, 11. Portrait, 12. Mall, 13. Street.

Figure 2. The stimulus 'Fruit' at 60mm camera separation and 'Street' at 120mm camera separation laid out for parallel viewing.

The lens-camera combinations used in the capturing of the images were characterized and the optical distortions from the images were removed in order to eliminate unintentional horizontal and vertical disparity. This was achieved using the Camera Calibration Toolbox for Matlab®. SIFT features [13] were used to find reliable correspondence points from the two images. These points were then used to automatically calculate the cropping and rotation that were applied to the right hand side image. The image was only rotated around its normal to avoid all distortions since the image pair would be used in subjective evaluations. Previous results [14] suggest that frame violations are highly detrimental to perceived image quality, and since in natural photography the violations would happen regularly, it was decided to remove the possibility altogether. The zero disparity level was adjusted automatically based on the correspondence points so that there was no negative (crossed) disparity in the image pair.

The stimuli were presented on a 24" Tridelity SL2400 autostereoscopic display based on parallax barrier technology with WUXGA resolution. Lighting was kept at a constant dim indirect office lighting of approximately 60lux at 3000K to minimize reflections and the background was an even monochromatic surface. The test participants rested their chin on an adjustable chinrest, which included a forehead band, to keep the viewing condition constant throughout the experiment. The chin rest was positioned so that the viewing distance was 80cm and thus the stimulus spanned 36.0° horizontally and 22.6° vertically.

2.2 Subjective evaluations

The experiment was conducted at the Department of Media Technology of the Aalto University School of Science in Q1/2010. The participants were recruited through email and an Internet newsgroup. A total of 22 people, aged 20-33 years, participated in the experiment during six days. All the participants were students or graduates of the university and had normal or corrected to normal vision, and none of them had deficiencies in colour vision. Four of the participants were female. Seven participants wore corrective eyewear. Age, gender or the use of eyewear had no significant effect on the results. Each participant received a cinema ticket as a compensation for their contribution. The participants were also interviewed about their previous experiences and attitudes towards stereoscopy in the pre-test questionnaire, and how the experiment affected their attitudes in the post-test questionnaire.

The participants read the instructions on the secondary display and filled a pre-test questionnaire. After filling the questionnaire, the participants were screened for stereo vision by displaying a test image on the screen. The test image was also used in adjusting the viewing position. It consisted of a large red circle at zero disparity surrounded by the letters ‘A’, ‘B’, ‘C’, ‘D’ and ‘E’ at different depth levels behind and in front of the circle. As the participants saw the image on the autostereoscopic display, the chin rest was adjusted so that the participants could view the image properly and that the images were not flipped. The participants were asked to recite the letters in order from the nearest to the farthest. All participants successfully completed this task.

In addition to the main attribute of focus of this study, ‘interestingness’, two more familiar attributes were also evaluated in the study. A previously used attribute in stereoscopic quality studies [2][15], ‘viewing experience’, was chosen to represent the overall quality of the viewing experience. Another attribute used in previous research, ‘naturalness’, was reformulated to ‘naturalness of depth’ as in [16] to exclude any evaluation criteria not related to depth perception, such as colour shifts.

A no-anchor single stimulus direct scaling method was selected for the evaluation tasks as it has been used in previous studies with success [2][17][18]. The scale was chosen to extend from 1 to 9 with no intermediate points instead of the five-value scale in the ITU-BT.500-11 standard [19] to find more subtle differences in the stimuli. The labels in the standard were not used either as it was likely that the labels would confuse the participants; ‘excellent’, ‘good’, ‘fair’, ‘poor’ and ‘bad’ do not describe the attribute ‘interestingness’. Instead, numbers from 1 to 9 were used and the participants were instructed that a higher number means a better score for all attributes. The resulting data was normalized as a countermeasure to the rubber-band effect present in no-anchor experiment setups, by converting the raw evaluation scores into z-scores, i.e. by subtracting the participant’s evaluations’ mean value from each of the participant’s evaluations and dividing by the standard deviation of his/her evaluations.

In order to make the evaluation task easier and thus to reduce noise, the participants were familiarized with the levels of quality by showing the actual stimuli in a random order as a practice round. They were instructed to practice the evaluation with one or two images and then browse through the photographs at their own pace. When the practice round was over, the evaluation round began with a randomized stimulus order. There was no time limit for the evaluation and it was possible to go back and forth between the stimuli. The mean total duration of the experiment was 61 minutes (σ=17min), of which the mean of the actual viewing and evaluation of the 52 stimuli was 46 minutes (σ=15min).

2.3 Recalled attention maps (RAMs)

In addition to the three attribute evaluation task, the participants were instructed to draw a transparent ellipse on top of the image in the region that caught their attention (Figure 3). The colour of the ellipse was changeable from green to red so that the participants could indicate the polarity of their selection by colouring the ellipse green if the region caught their attention positively and red if negatively. Only one ellipse was allowed per image in this study to keep the task simple. The areas are named here *recalled attention maps*, RAMs. Qualitative data was attached to the RAMs by the means of textual input by the participants. They were asked to describe what in particular caught their attention in the chosen region. This method introduces us several new possibilities, the most prominent being able to map qualitative data spatially to the stimuli, resulting in spatial qualitative data. The RAMs also enable us to localize recalled regions of interest that are important to the defined task, e.g. artefact localization, without having to resort to ambiguous stimulus segmentation and annotation.

The individual RAMs were combined to form three separate maps for each stimulus; negative, positive and total recalled attention maps. These maps were convolved with a two-dimensional Gaussian with σ of one degree visual angle in order to make the maps comparable with fixation maps from the second part of the study, and to mask the errors arising from the forced ellipse shape of the individual maps.

(a) (b) (c)

Figure 3. The experiment setup (a), a participant drawing a RAM (b) and the user interface (c).

2.4 Eye tracking

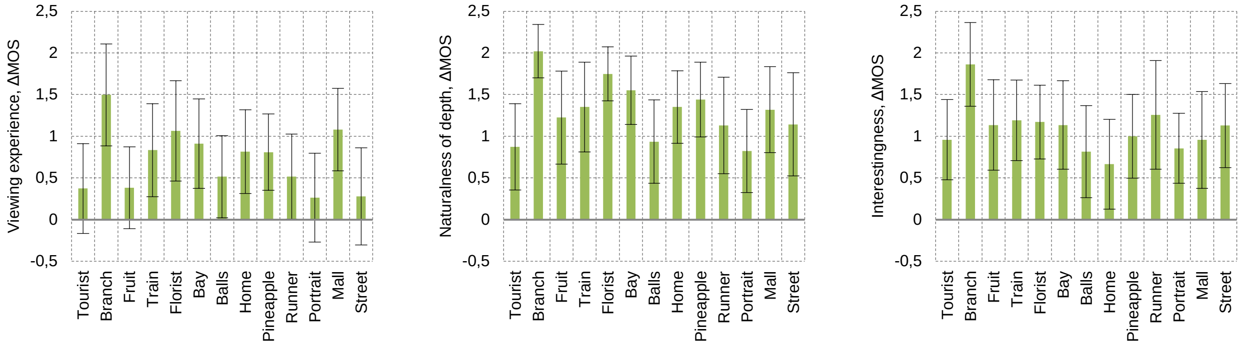
The spatial data retrieved from the RAMs is comparable to eye tracking data and the relationship of the two methods was studied in a second experiment. The experiment was conducted in Q4/2010 in the same laboratory as the first experiment, and the display, viewing conditions and recruitment methods were also same. A total of 32 people participated in the experiment, but due to poor quality data eight participants' data was discarded prior to analysis and they are also excluded from the demographics. All the participants were students, alumni or employees of the university, aged 20 to 62 years with normal or corrected to normal vision. Four of the participants wore glasses and five of the participants were female.

The experiment was a between subjects eye tracking study with monoscopic and stereoscopic viewing conditions, twelve participants in each group. The monoscopic stimuli consisted of the same monoscopic images as used in the first experiment and the stereoscopic stimuli were selected based on the first experiment's results so that for each scene the best camera separation was chosen based on the viewing experience scores. After filling the pre-test questionnaire, the participants were screened for stereo vision and the chinrest was adjusted as in the first experiment. Eye movements were recorded with the SMI RED250 eye tracker (SensoMotoric Instruments GmbH, Teltow, Germany) at 250Hz, each eye was tracked separately. After the viewing position adjustment the tracker was calibrated with a 9-point calibration procedure, and the calibration was validated. During the experiment no recalibration was done, as the total duration of the experiment was less than eight minutes. Each image was shown for 30 seconds, of which the first 20 seconds was used in the analysis and between the stimuli a fixation cross was displayed for 5 seconds. As it has been shown, instructions given to participants in eye tracking experiments change the gaze patterns [9], so in order to keep the results as generalizable as possible, the participants were only instructed to "*look at the images as you normally would*". The tracking was done binocularly and thus resulted in two separate data streams for the left and right eye. For the analysis we only used the left eye data since for most stimuli the same image was shown for the left eye in both 2D and 3D groups. We extracted fixations, saccades and blinks from the data. The fixations were mapped spatially to the stimulus coordinates and weighted with the fixation duration, after which the map was convolved with a two-dimensional Gaussian with σ of one degree visual angle in order to simulate the foveal accuracy and the eye tracker's lack of precision.

3. RESULTS AND DISCUSSION

3.1 Mean opinion scores

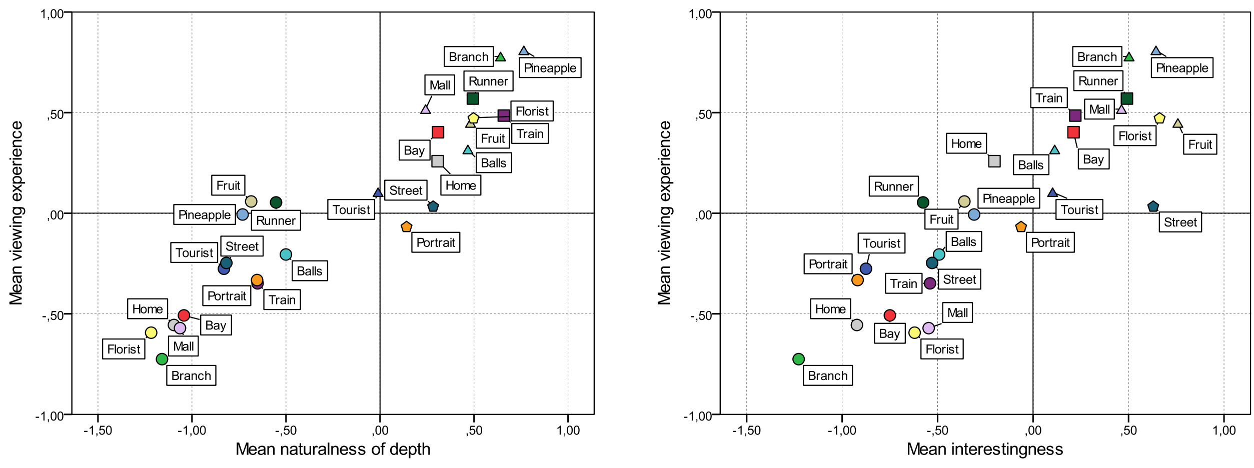
The normalized mean opinion scores were calculated for all 52 stimuli and the highest scoring stereoscopic stimulus of each scene was chosen to be compared with the monoscopic one for each attribute separately. From these MOS values the ΔMOS values were calculated by subtracting the monoscopic score from the stereoscopic score (Figure 4). The significant perceived added value of stereoscopy can be observed in all dimensions. For viewing experience four of the thirteen scenes did not reach significantly higher scores in the stereoscopic condition, but for naturalness of depth and interestingness all the scenes attained remarkably high scores in 3D. Based on these results we can retain our hypothesis 1, that stereoscopic imagery is more interesting, with confidence. The otherwise dull, but structurally complex scene Branch, which includes distinct depth differences, benefits most from stereoscopy. In this particular case the structural complexity of the scene may be the key factor affecting the difference, as well as in the scenes Florist and Mall. However, a more thorough analysis is required in order to draw conclusions.



(a) (b) (c)

Figure 4. The mean difference of normalized MOS values between the monoscopic and the highest scoring stereoscopic images for viewing experience (a), naturalness of depth (b) and interestingness (c) with 95% CI.

Based on the results, the highest scoring camera separation of each scene on the viewing experience dimension was chosen as the stereoscopic stimulus to be compared with the monoscopic one (Figure 5). It is important to note that these are not the highest scoring images in naturalness of depth or interestingness space, in seven out of thirteen scenes the interestingness was higher with a wider camera separation, and in six scenes the same separation was preferred for interestingness and viewing experience. As the following analysis is based on the chosen separations, the data is most favourable to the viewing experience attribute. The selection was made so that we get comparable results from the analysis linked to the same stimuli. Still, we can observe significantly higher ΔMOS values for naturalness of depth and interestingness compared with viewing experience. The attributes are more closely correlated in the monoscopic condition than in the stereoscopic condition (Table 1). While naturalness of depth is highly correlated with viewing experience ($\rho>0.80$, $p<0.001$), interestingness seems to possess a more independent dimensionality, with ($\rho=0.678$, $p<0.011$) in the monoscopic viewing condition and non-significant correlation in the stereoscopic condition. The same trend can be found in the correlations of the differences. While naturalness of depth ΔMOS correlates strongly with viewing experience ΔMOS ($\rho=0.889$, $p=0.000$), interestingness ΔMOS correlation is significantly weaker ($\rho=0.577$, $p=0.039$).



(a) (b)

Figure 5. Scenes' viewing experience scores plotted against naturalness of depth (a) and interestingness (b). Camera separation 0mm (monoscopic) marked with circles, 30mm triangles, 60mm squares and 120mm pentagons.

Table 1. Correlations of the attribute scores in monoscopic viewing condition (a), stereoscopic (b) and of the ΔMOS (c).

Monoscopic		Naturalness of depth	Interesting-ness
Viewing experience	ρ	.829**	.678*
	Sig.	.000	.011
	N	13	13
Naturalness of depth	ρ	1	.487
	Sig.		.091
	N	13	13

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

(a)

Stereoscopic		Naturalness of depth	Interesting-ness
Viewing experience	ρ	.814**	.525
	Sig.	.001	.065
	N	13	13
Naturalness of depth	ρ	1	.491
	Sig.		.088
	N		13

**. Correlation is significant at the 0.01 level (2-tailed).

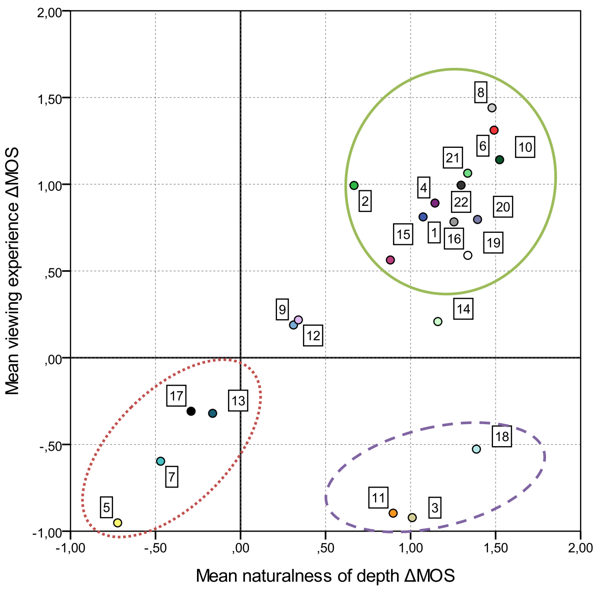
(b)

Δ		Naturalness of depth	Interestingness
Viewing experience	ρ	.889**	.577*
	Sig.	.000	.039
	N	13	13
Naturalness of depth	ρ	1	.581*
	Sig.		.037
	N		13

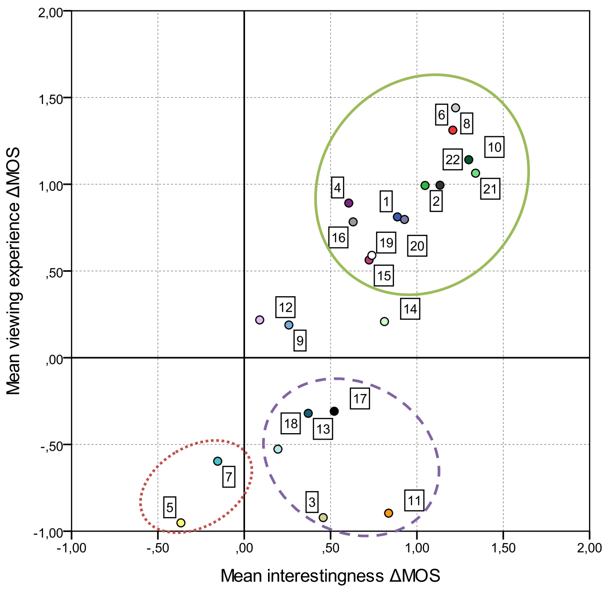
**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

(c)

As we examine the participants in the ΔMOS space (Figure 6) we find distinctive clusters. There are several possibilities for the formation of these clusters. First, it is possible that the viewing position, regardless of the care taken to ensure the best possible viewing conditions, has not been optimal. There is very little that can be done to improve the test setup in this regard, since it is up to the participant's subjective opinion when he or she is satisfied with the image on an autostereoscopic display, where there is always some crosstalk. Second, the participant's stereoscopic vision might have some deficiency, which was not detected by the stereo vision screening. In future studies, extensive vision testing is recommended. While the two first reasons address the individual's stereoscopic percept, the third reason is more interesting from the point of view of viewing experience research: a dichotomy may exist in the population regarding the appeal of stereoscopic imagery to an individual. This phenomenon is similar to the idea proposed by Datta et al. [20], who consider the personalization of aesthetics algorithms to include individual preferences. The pending qualitative data analysis is expected to shed light into this question. Also a more in-depth analysis of the dimensions of the evaluated attributes is needed.



(a)

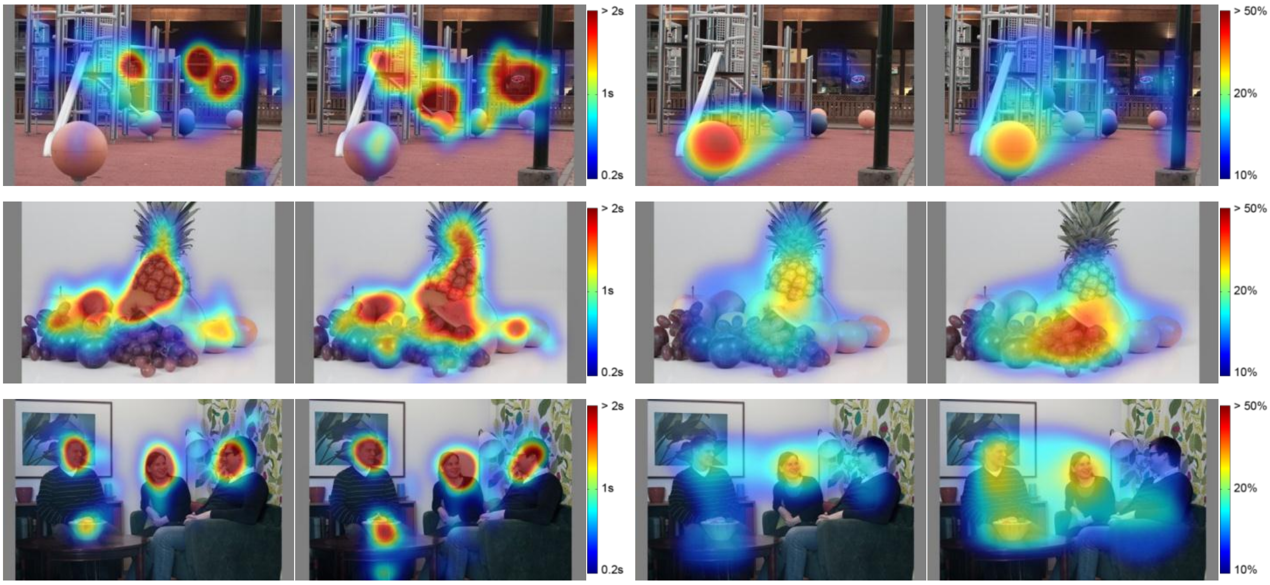


(b)

Figure 6. Participant mean difference of normalized viewing experience plotted against naturalness of depth (a) and interestingness (b). Distinct participant clusters highlighted by positive (solid green), negative (dotted red) and mixed (dashed purple) attitudes.

3.2 RAMs and fixation maps

The RAMs drawn by the participants in the first experiment and the fixation maps of the second experiment were compared. Visualizations of the maps were plotted using a mean fixation duration scale of 0.2s to 2s for the fixation maps (Figure 7a) and the percentage of participants that chose the area on a scale from 10% to 50% for RAMs (Figure 7b).

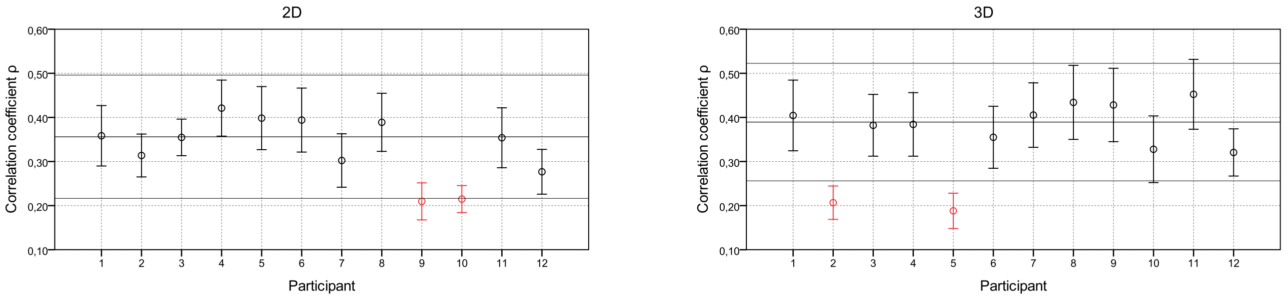


(a) (b)

Figure 7. Examples of the fixation maps (a), monoscopic condition on the left, stereoscopic on the right, scale from 0.2 to 2 seconds mean fixation time. Examples of RAMs (b), monoscopic on the left, stereoscopic no the right, scale from 10% to 50% of participants. Scenes Balls, Fruit and Home respectively from top to bottom.

We calculated the two-dimensional correlations between monoscopic and stereoscopic fixation maps and expected to see relatively weak correlation, but instead we found that the mean correlation was relatively high (ρ_μ=0.79, σ=0.075) for the thirteen scenes. Thus we reject the hypothesis 2 and conclude that the differences in fixation maps between monoscopic and stereoscopic conditions in long duration free viewing situations are generally minute and highly content dependent. The presence of bottom-up features and top-down semantics, such as faces and text, that are plentiful in natural photographic images, mask the effects of disparity on saliency.

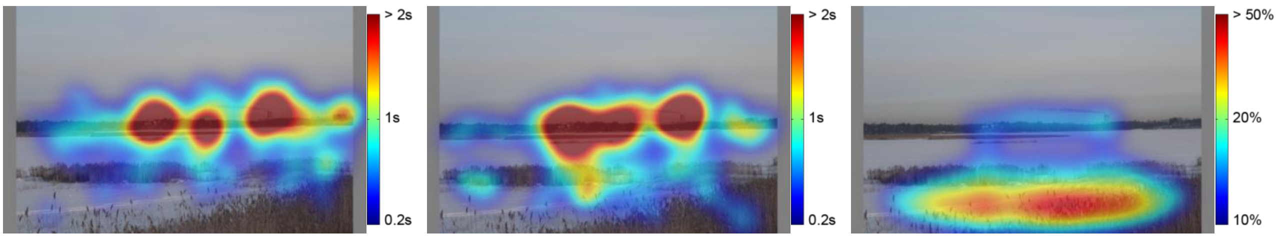
The inter-participant correlation of fixation maps is examined first by calculating the participants' mean correlations with other participants (Figure 8) as in [21]. Two outliers were removed from both groups. Second, the correlation coefficients for each fixation map between each of the subjects for each scene are calculated (Figure 10a). We can observe a trend towards higher correlation in the stereoscopic group. Also one outlier was detected, the scene Portrait, which was removed from the analysis on the basis of diverting from the mean value in monoscopic condition by almost three times the standard deviation. As we conducted a two-way ANOVA we found a significant main effect for the factor Stereoscopy ($F(1,11)=6.520$, $p=0.027$) and for Scene ($F(1,11)=7.293$, $p=0.001$). From these results we come to the conclusion that stereoscopic imagery is more effective in directing the viewer's attention.



(a) (b)

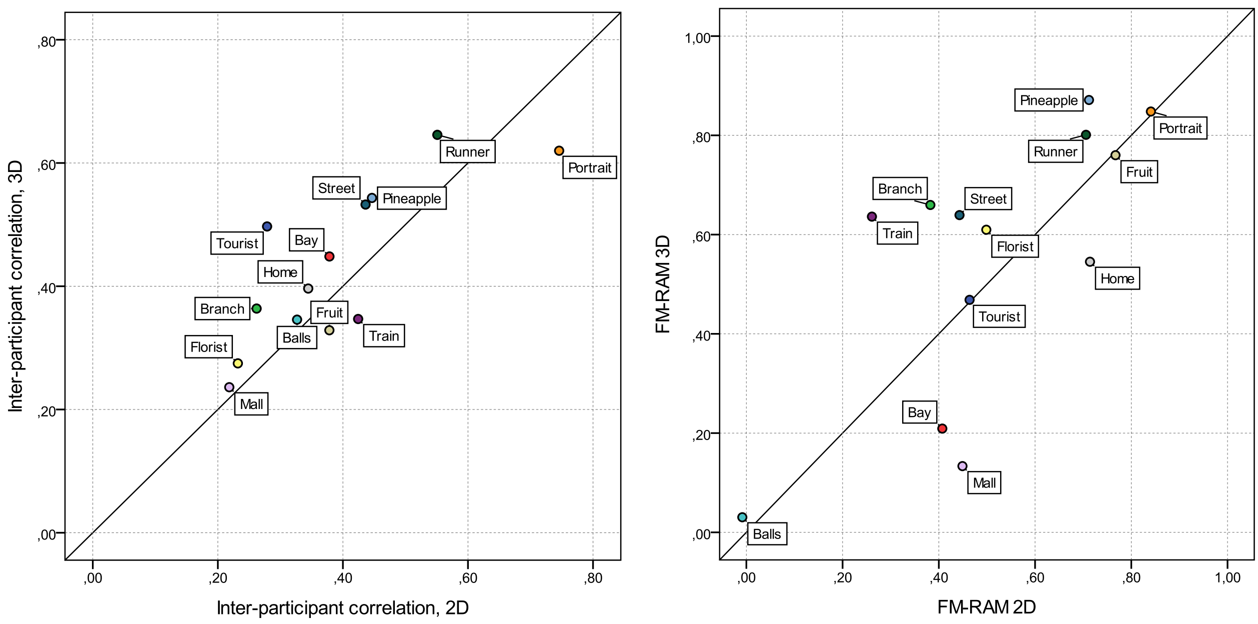
Figure 8. Mean inter-participant fixation map correlations with 95% CI for 2D (a) and 3D (b) groups. Rejected outliers highlighted in red. Post outlier rejection group mean reference line with three times std. dev. reference lines on both sides.

As we had hoped, in most cases the recalled attention maps differ from the fixation maps in the stereoscopic condition (Figure 10b). In a particular scene, Bay (Figure 9), the reeds in the foreground become clearly distinct from their background in stereoscopic viewing, and the participants of the first experiment have selected this as a region that caught their attention. This is not noticeable by examining the fixation maps. These results are encouraging for the RAM method, as it enables us to find the regions that people subjectively find important in an image with smaller interference from other saliency attractors, like the buildings in the horizon in the Bay scene. The RAMs were drawn with the specific task in mind, to find the area that draws most attention and also the underlying task of scoring the images, thus the results are not completely comparable with the task free eye tracking experiment's results. A fair amount of qualitative data was gathered and is still unanalysed: 1114 quotes were coded into 56 codes. We expect the qualitative data analysis to reveal much more about the nature of the appeal of stereoscopic imagery.



(a) (b) (c)

Figure 9. Fixation maps for the scene Bay in monoscopic (a) and stereoscopic condition (b), scale from 0.2 to 2 seconds mean fixation time. The RAM of the stereoscopic image Bay (c) highlights the interesting regions in 3D better than fixation maps, scale from 10% to 50% of participants.



(a) (b)

Figure 10. (a) Inter-participant correlation coefficient ρ by scene in monoscopic and stereoscopic conditions for the fixation maps with a reference line (c=0, m=1). (b) Scenes' two-dimensional correlation coefficients of the fixation maps (FM) and the recalled attention maps in stereoscopic and monoscopic conditions.

4. CONCLUSIONS

In this paper we presented the first results of our research of interestingness and binocular attention in natural photographic images. We introduced a novel attribute to explain the added value of stereoscopy. Our results show that interestingness brings more dimensionality into the study of the added value of stereoscopy, as the attributes used in earlier studies, viewing experience as naturalness of depth, are closely correlated. We conclude that based on our results, viewing experience and naturalness of depth do not encompass all the added value that stereoscopy contains. We recommend the use of interestingness as an addition in future studies. In the subjective experiment we observed clustering of the participants based on their preference for stereoscopy, a phenomenon that should be studied more closely.

In the eye tracking study, we found that stereoscopic content is more effective in directing the viewer's gaze, as the inter-participant correlation is higher. Significant differences in the actual fixation maps were not observed between the two viewing conditions, but the shift appears to be towards three dimensional structures. To examine the localisation of interesting regions, we presented a novel method, the recalled attention map. The RAMs gave us encouraging results, as we could easily recognize the subjectively interesting regions, and we expect to get more interesting results from the analysis of the qualitative spatial data.

REFERENCES

[1] Tam, W.J., Stelmach, L.B. and Corriveau, P.J., “Psychovisual aspects of viewing stereoscopic video sequences,” Proc. SPIE 3295, 226-235 (1998)

[2] Seuntiens, P., Heynderickx, I., IJsselsteijn, W., Van Den Avoort, P.M.J., Berentsen, J., Dalm, I.J., Lambooij, M. and Oosting, W., “Viewing experience and naturalness of 3D images,” Proc. SPIE 6016, 43–49 (2005)

[3] Oxford English Dictionary, “interesting, adj.”, http://www.oed.com/view/Entry/97743

[4] Berlyne, D.E., Ogilvie, J.C. and Parham, L.C., “The dimensionality of visual complexity, interestingness, and pleasingness,” Can J Psychol, 22(5), 376-87 (1968)

[5] Savakis, A.E., Etz, S.P. and Loui, A.C. P., “Evaluation of image appeal in consumer photography,” Proc. SPIE 3959, 111-120 (2000)

[6] S. Yendrikhovskij, [Colour Image Science: Exploiting Digital Media], John Wiley & Sons Ltd., Chichester, 394-418 (2002)

[7] de Ridder, H. and Endrikhovski, S. “Image Quality is FUN: Reflections on Fidelity, Usefulness and Naturalness,” SID Int. Symp. Digest Tech. Papers, 33(1), 986-9 (2002)

[8] Itti, L. and Koch, C., “Computational modelling of visual attention,” Nat Rev Neurosci, 2, 194-203 (2001)

[9] Yarbus, A.L., [Eye Movements and Vision], Plenum Press, New York, (1967)

[10] Kanwisher, N., McDermott, J., and Chun, M.M., “The fusiform face area: a module in human extrastriate cortex specialized for face perception,” J Neurosci, 17, 4302-11 (1997)

[11] Jansen, L., Onat, S. and König, P., “Influence of disparity on fixation and saccades in free viewing of natural scenes,” J Vis, 9(1), 1-19 (2009)

[12] Liu, Y., Cormack, L.K. and Bovik, A.C., “Dichotomy between luminance and disparity features at binocular fixations,” J Vis, 10(12), 1-17 (2010)

[13] Lowe, D.G., “Object recognition from local scale-invariant features,” Proc. ICCV, 1150-1157 (1999)

[14] McKenna, M., “Interactive viewpoint control and three-dimensional operations,” Proc. I3D 92, 53–56 (1992)

[15] Heynderickx, I. and Kaptein, R., “Perception of detail in 3D images,” Proc. SPIE 7242, 72420W-1-9 (2009)

[16] Ijsselsteijn, W., de Ridder, H., Hamberg, R., Bouwhuis, D. and Freeman, J., “Perceived depth and the feeling of presence in 3DTV,” Displays, 18(4), 207-214 (1998)

[17] Ijsselsteijn, W., de Ridder, H. and Vliegen, J., “The effect of stereoscopic filming parameters and exposure duration on quality and naturalness judgements,” Proc. SPIE 3644, 300-311 (1999)

[18] Kuijsters, A., Ijsselsteijn, W., Lambooij, M. and Heynderickx, I., “Influence of chroma variations on naturalness and image quality of stereoscopic images,” Proc. SPIE 7240, 72401E1-9 (2009)

[19] Rec. ITU-R BT.500-11, “Methodology for the subjective assessment of the quality of television pictures,” International Telecommunication Union, 1-48 (2002)

[20] Datta, R. and Wang, J.Z., “Algorithmic inferencing of aesthetics and emotion in natural images: An exposition,” Proc. ICIP, 105-108 (2008)

[21] Engelke, U., Maeder, A. and Zepernick, H.J., “Analysing Inter-Observer Saliency Variations in Task-Free Viewing of Natural Images,” Proc. ICIP, 1085-1088 (2010)