# What do people look at when they watch stereoscopic movies?

Jukka Häkkinen [a,b,c*], Takashi Kawai[d], Jari Takatalo[c], Reiko Mitsuya[d] and Göte Nyman[c]

[a] Department of Media Technology,Helsinki University of Technology, PO Box 5500 TKK, Finland
[b] Nokia Research Center, PO Box 407, 00045 Nokia Group, Finland
[c] Department of Psychology, PO Box 9, 00014 University of Helsinki, Finland
[d] Graduate School of Global Information and Telecommunication Studies,Waseda University, Japan

## ABSTRACT

We measured the eye movements of participants who watched 6-minute movie in stereoscopic and non-stereoscopic form. We analyzed four shots of the movie. The results indicate that in a 2D movie viewers tend to look at the actors, as most of the eye movements are clustered there. The significance of the actors start at the beginning of a shot, as the eyes of the viewer focus almost immediately to them. In S3D movie the eye movement patterns were more widely distributed to other targets. For example, complex stereoscopic structures and structures nearer than the actor captured the interest and eye movements of the participants. Also, the tendency to first look at the actors was diminished in the S3D shots. The results suggests that in a S3D movie there are more eye movements which are directed to wider array of objects than in a 2D movie.

**Keywords:** Stereoscopic movie, eye movements, saliency map

## 1. INTRODUCTION

The purpose of a moviemaker is to influence the viewers to pay attention to salient events of the script, so that the viewers understand the details, consequences and emotional significance of events[1]. This can be accomplished by utilizing for example shot distance, focus, angle, movement, point of view, scene composition and principles of cutting[1,2]. In a stereoscopic film the effect of these techniques might be different, as the processes of stereoscopic vision affect the way viewers pay attention and understand the scenes. Moviemakers and stereographers know how to utilize the possibilities of stereoscopy, as recent excellent stereoscopic films have shown, but there is less empirical data related to these effects. According to our recent studies the viewers most often mention experiences of reality-likeness, presence, enhanced emotions and richness of structural details when watching a stereoscopic movie[3,4]. Asking the viewers to describe their experiences is the best way to form an understanding of the underlying psychological processes[5-8], so these results tell us a lot about the experiential added value of stereoscopy. However, there are also processes that are not consciously accessible. These processes might be reflexive, automatic or too quick to enter the consciousness of the viewer. For example, eye movements and the related changes in the focus of attention are only partially guided by conscious intentions of the viewers. As the locations where the eyes stop to collect information determine what parts of the visual environment we notice, measuring the eye movements with stereoscopic film shows, which part of each shot is regarded as informative and important.

Eye movements can be divided to two main phases. Firstly, there are fixations when the eye is pointing to a single location of the scene. Secondly, there are saccades when the eyes quickly move the point of regard to another position. Information is acquired during fixations, as during the saccades the information from the eye is mostly suppressed. Experiments suggest that viewers look at the most informative areas of the scene. The definition of informative depends on the task and the contents, as it can be semantic informativeness, i.e., the meaning of the area or it can be visual informativeness, i.e., visual salience of the specific area. Visual salience means that an area is differentiated from adjacent areas by its luminance, color, texture or other feature[9,10]. It is assumed that such basic attributes draw the attention and eye movements of the person immediately when the scene viewing starts. In simplified images, like texture arrays, salient areas are formed by areas with lines that are orthogonal to their neighboring lines[11], move to different direction[11-12], have different brightness or color[11], or that are at different stereoscopic depth[12,13]. Eye movement studies have shown that salient target images can be indicated by the first saccade that occurs during the scene viewing[14], which suggests that the

IS&T/SPIE's International Symposium on Electronic Imaging: Science and Technology. Stereoscopic Displays and Applications XXI, 18.-21.1.2010, San Jose, California,USA. Proceedings of SPIE, Vol. 7524.

information defining the target is available to guide the first eye movement in a scene after a very short time period[13]. It has also been shown that salience maps based on low-level features predict eye movements in videos accurately[15].

Semantic informativeness does not affect initial fixation positions during picture viewing[9,16,17], but fixations to semantically informative areas increase as a function of viewing time[9,17]. For example, in the study of Yarbus participants looked at the faces of the people in the picture when the task was to determine their age, but looked at other things when they we instructed to understand the material circumstances of the family[9,18]. Similarly, Birmingham et al showed recently that in social scenes the social informativeness is more important determinant of eye movement patterns than low level salience maps[19].

Henderson and Hollingworth (1998) have combined the visual and semantic informativeness in their saliency map framework in which they describe a two-stage process where the eye movements are initially guided by an early parse of a scene based on low spatial frequency information that is quickly available[9]. With prolonged viewing the saliency map is modified by cognitive interest related to the scene.

There are only few studies describing the eye movement patterns in movies. In these studies it has been shown that viewers look at approximately same location when watching a movie, although there seems to be gender differences in the areas that viewers find interesting[20,21]. In our study we wanted to find out, how stereoscopic presentation affects the eye movement patterns. Based on our earlier study[3], we formed a hypothesis that in a stereoscopic film the eye movements might be more widely distributed, as in our previous study the participants reported that stereoscopic movie has much more details to see compared to 2D movie.

## 2. METHODS

### 2.1 Participants

Twenty students from University of Helsinki participated the experiment. There was a visual screening in which the stereoscopic acuity, visual acuity, horizontal near heterophoria and near point of accommodation were measured. None of the participants were excluded from the main experiment because of their vision.

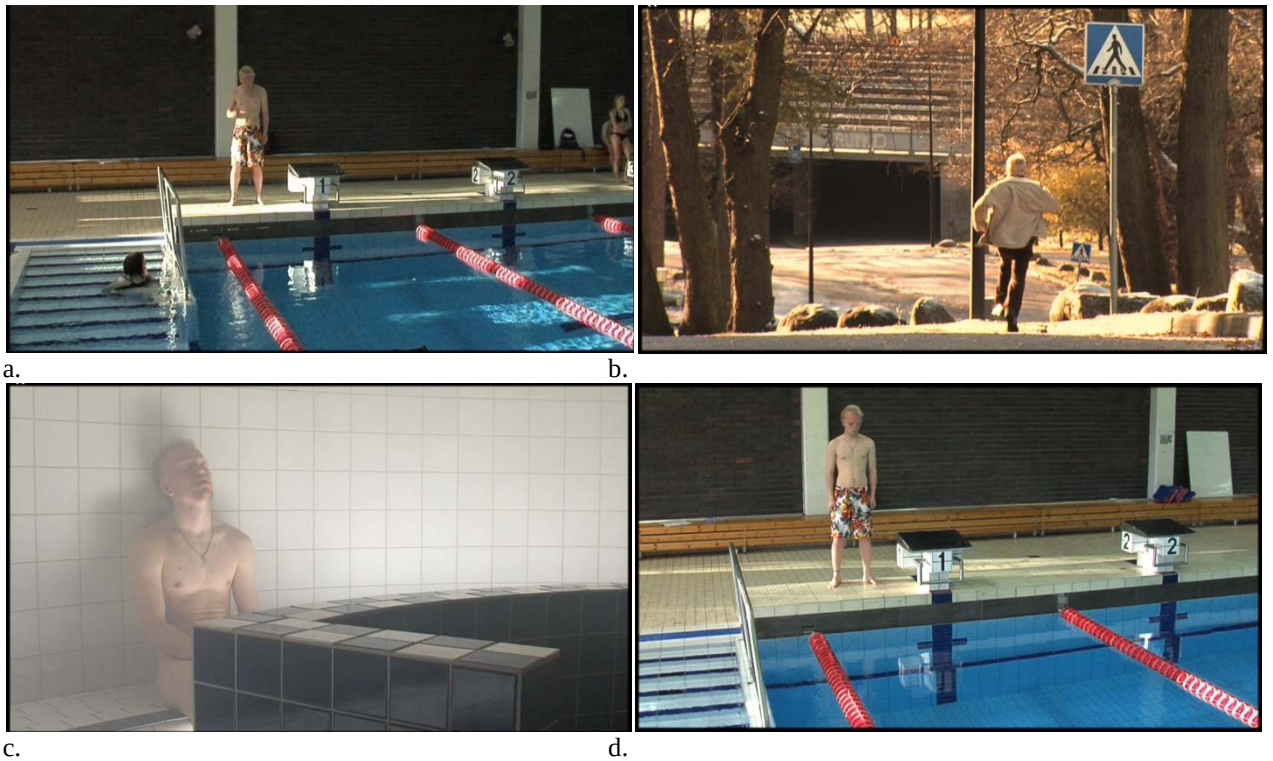
### 2.2 Contents

The short film (6 minutes 20 seconds) was produced by Stereoscape Ltd. (www.stereoscape.com) for “All different all equal” campaign by Finnish Youth Co-operation and featured a love story between a boy and a girl in a wheelchair. It consisted of 40 shots of varying length.

### 2.3 Test procedure

We used Hyundai 46-inch polarizing stereoscopic display with resolution of 1920 x 1080 pixels. The film was shown with TriDef stereoscopic player and Tobii X120 eye movement tracker was utilized to measure the eye movements (Figure 1). In the experiment the viewers watched both stereoscopic and non-stereoscopic versions of the contents from a viewing distance of 140 centimeters. In the 2D version the viewer saw two views intended for the left eye so there was no binocular disparity in the film. The viewers wore the polarizing glasses when viewing the 2D version so that the viewing conditions with 2D and S3D movies were comparable. The 2D and S3D versions were shown in random order. The participants were instructed to compare which of the versions was better.

Figure 1. The experimental setting with Hyundai stereoscopic display and Tobii X120 eye movements tracker.



a. b.

c. d.

Figure 2. Four shots from the movie that were analyzed. a) Shot 1: Dialog (22.1 seconds), b) Shot 2: Boy running (7.0 seconds), c) Shot 3: Sauna (5.9 seconds), d) Shot 4: Boy standing (5.5 seconds)

## 3. RESULTS

We selected four shots for further analysis (Fig. 2). The main criterion for selection was that they did not contain large amount of camera or object movements. Two types of eye movement visualizations were obtained from the shots. Firstly, eye movement patterns were visualized to find out the typical patterns in each scene (Fig. 4). We also visualized the eye movement patterns utilizing the heat map visualization of Tobii Studio software to indicate the clustering of the fixations (Fig. 5). The color of the heat map indicates the number of eye movements clustering to a specific area. Red color indicates higher number of eye movements, yellow and green color smaller number of eye movements. The color map has been scaled according to the eye movement distribution in each content, so the red color represents different number of eye movements in each content type. Based on our initial analysis, we selected several areas of interest (AOI; Fig. 3) in each shot and calculated the number of fixations to each of the areas of interest and the time it took to first fixate to the area of interest.

### 3.1 Shot 1: Dialog

In this 22.1 second shot boy is standing by the pool and is discussing with a girl sitting in the pool. There is also a girl standing in the right edge of the scene. The camera stays almost stationary and the only action in the scene is the dialog between the boy and the girl. The eye movements are clearly clustered around the discussion participants, and there is also a small cluster around the girl standing on the right edge of the scene (Fig.4a and 4b). Focusing to the main actors of the scene is in accordance with earlier findings. The main difference between the S3D (Fig. 4a) and 2D (Fig. 4b) versions of the scene is that in the S3D version the eye movements are more widespread, as the heat map visualizations of the scenes show (Fig.5a and 5b). In S3D version there seems to be eye movements that indicate exploration of the pool side as well as the water. This suggests that three-dimensional structures seem to be drawing the attention of the viewers away from the actors of the scene.

We divided the scene into six areas of interest (AOI) which are shown in figure 3a. When the eye movements of all participants within the AOIs are summed in Table 1, it can be seen that the there are significantly more fixations to the girl (Chi-square test, $p<0.01$) and water ($p<0.001$). Also, the total number of fixations in the 3D shot is significantly higher ($p<0.05$). Table 2 shows the time when first fixation toward an AOI was made in the shot. In the 2D version the eye movements were focused to the boy within half a second (Table 2), but in the S3D version there was no clear tendency to fixate to a specific AOI, as the longer time lags show.

Table 1. Percentage of fixations in areas of interest of shot 1. The number in parenthesis indicates the number of fixations. The fixations are a summed from the 20 experimental participants. Asterisks show significant differences (* $p<0.05$, ** $p<0.01$).

| | **Boy** | **Girl** | **Water** | **Edge** | **Background** | **Girl 2** | **Outside the display** | **Sum** |
|---|---|---|---|---|---|---|---|---|
| **2D** | 40.57% (142) | 11.71% (41) | 7.71% (27) | 15.14% (53) | 11.71% (41) | 2,00% (7) | 11.14% (30) | 100.00% (350) |
| **S3D** | 40.29% (166) | 17.72% (73)** | 12.14% (50)** | 12.62% (52) | 9.71% (40) | 1.21% (5) | 6.31% (26) | 100.00% (412)** |

Table 2. Time (seconds) before first eye movement is made toward an area of interest in shot 1. Average times of 20 participants.

| | **Boy** | **Girl** | **Water** | **Edge** | **Background** | **Girl 2** | **Outside the display** | **Length of the shot** |
|---|---|---|---|---|---|---|---|---|
| **2D** | 0.593 | 6.742 | 5.148 | 10.769 | 3.725 | 2.121 | 6.812 | 22.1 |
| **S3D** | 1.909 | 5.098 | 6.558 | 10.411 | 2.693 | 4.963 | 6.599 | 22.1 |

### 3.2 Shot 2: Boy running

In this 7.0 second shot a boy is running downhill away from camera. There are trees around him and railroad tracks in the background. During the shot a train appears from left, goes through the top of the screen and disappears to the right. Camera is almost stationary during the shot. The eye movements are clearly focused to the running boy (Fig. 4c and 4d), and the slight left-right movement of the boy is indicated by the horizontal spreading of the eye movement cluster (Fig. 5c and 5d). The eye movement patterns are quite similar in 2D and S3D versions, which might indicate that the running boy is so significant that it draws the attention of the viewer in a similar manner in both versions. This effect can be explained either by visual or semantic informativeness. From the visual informativeness point of view the moving object is a target that attracts the attention of the viewer automatically. From semantic informativeness point of view the actions of a running boy are also significant. There is also a cluster of eye movements in the railroad tracks where the train goes during the scene. The only significant difference is the higher number of fixations toward the train in the S3D version ($p<0.01$) (Table 3).

Table 3. Percentage of fixations in areas of interest of shot 2. The number in parenthesis indicates the number of fixations. The fixations are a summed from the 20 experimental participants. Asterisks show significant differences (* $p<0.05$, ** $p<0.01$).

| | Boy | Environment | Train | Outside the display | Sum |
|---|---|---|---|---|---|
| **2D** | 80.17% (93) | 15.52% (18) | 2.59% (3) | 1.72% (2) | 100.00% (116) |
| **S3D** | 68.18% (90) | 12.12% (22) | 16.67% (16)** | 3.03% (4) | 100.00% (132) |

Table 4. Time (seconds) before first eye movement is made toward an area of interest in shot 2. Average times of 20 participants.

| | Boy | Environment | Train | Outside the display | Length of the shot |
|---|---|---|---|---|---|
| **2D** | 0.16 | 0.49 | 5.03 | 1.93 | 7.00 |
| **S3D** | 0.10 | 2.38 | 4.66 | 1.87 | 7.00 |

### 3.3 Shot 3: Sauna

In this 5.9 second shot a boy is sitting still in a sauna. The camera is stationary and the boy's expression indicates that he is thinking something. The eye movements of are clearly focused to the boy's face, so the viewers are probably trying to interpret his mental state (Fig. 4f and 5f). However, in the S3D version there is also a significant clustering of eye movements to the wall-like structure in front of him ($p<0.0001$) (Fig. 4e and 5e; Table 5). It seems that structures in front of the main character capture the attention and eye movements of the participants.

Table 5. Percentage of fixations in areas of interest of shot 3. The number in parenthesis indicates the number of fixations. The fixations are a summed from the 20 experimental participants. Asterisks show significant differences (* $p<0.05$, ** $p<0.01$).

| | Boy | Front Structure | Back structure | Background | Total |
|---|---|---|---|---|---|
| **2D** | 53.72% (65) | 11.57% (14) | 14.05% (17) | 20.66% (25) | 100.00% (121) |
| **S3D** | 56.20% (68) | 44.63% (54)** | 19.83% (24) | 19.01% (23) | 100.00% (169)** |

Table 6. Time (seconds) before first eye movement is made toward an area of interest in shot 3. Average times of 20 participants.

| | Boy | Front Structure | Back structure | Background | Shot length |
|---|---|---|---|---|---|
| **2D** | 0.24 | 1.46 | 2.25 | 2.91 | 5.90 |
| **S3D** | 0.68 | 1.40 | 2.14 | 2.21 | 5.90 |

### 3.4 Shot 4: Boy standing

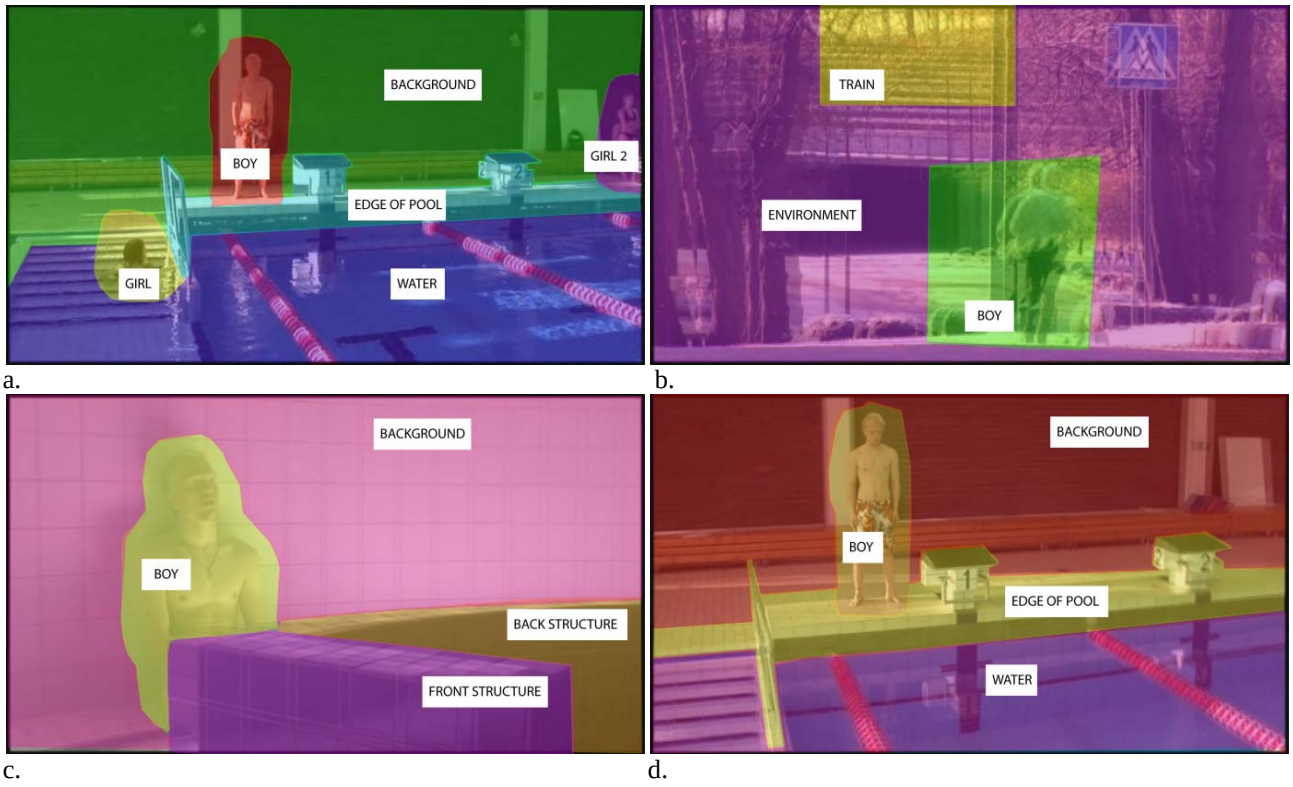
In this 5.5 second shot a boy is standing still by the pool, preparing to jump. The camera is stationary and the eye movements are focused to the boy (Fig. 4g and 4h), as the viewers are expecting him to jump to the pool. There are also eye movements toward the pool in front of the boy, which might indicate the expectation of jumping to the pool. Although the heat map (Fig. 5g and 5h) shows that in the S3D shot the eye movements are more spread to the water and to the background structures, there are no differences in the fixation counts between 2D and S3D versions (Table 7). This might be partially caused by the shortness of the shot, which reduces the total number of the fixations in the scene. The first eye movements are quickly focused to the boy in the 2D version, but in the S3D version the initial eye movements are more widely distributed.

Table 7. Percentage of fixations in areas of interest of shot 4. The number in parenthesis indicates the number of fixations. The fixations are a summed from the 20 experimental participants. Asterisks show significant differences (* $p<0.05$, ** $p<0.01$).

| | Boy | Edge | Water | Background | Outside the display | Sum |
|---|---|---|---|---|---|---|
| **2D** | 46.19% (103) | 23.77% (53) | 11.66% (26) | 13.45% (30) | 4.93% (11) | 100.00% (121) |
| **S3D** | 37.77% (88) | 22.75% (53) | 12.89% (30) | 24.46% (57) | 2.15% (5) | 100.0% (169) |

Table 8. Time (seconds) before first eye movement is made toward an area of interest in shot 4.

| | Boy | Water | Edge | Background | Outside the display | Shot length |
|---|---|---|---|---|---|---|
| **2D** | 0.21 | 3.81 | 2.47 | 2.61 | 4.13 | 5.5 |
| **S3D** | 0.68 | 2.61 | 2.91 | 2.21 | 5.17 | 5.5 |




a. b. c. d.

Figure 3. Areas of interest. a) Shot 1, b) Shot 2, c) Shot 3, d) Shot 4.

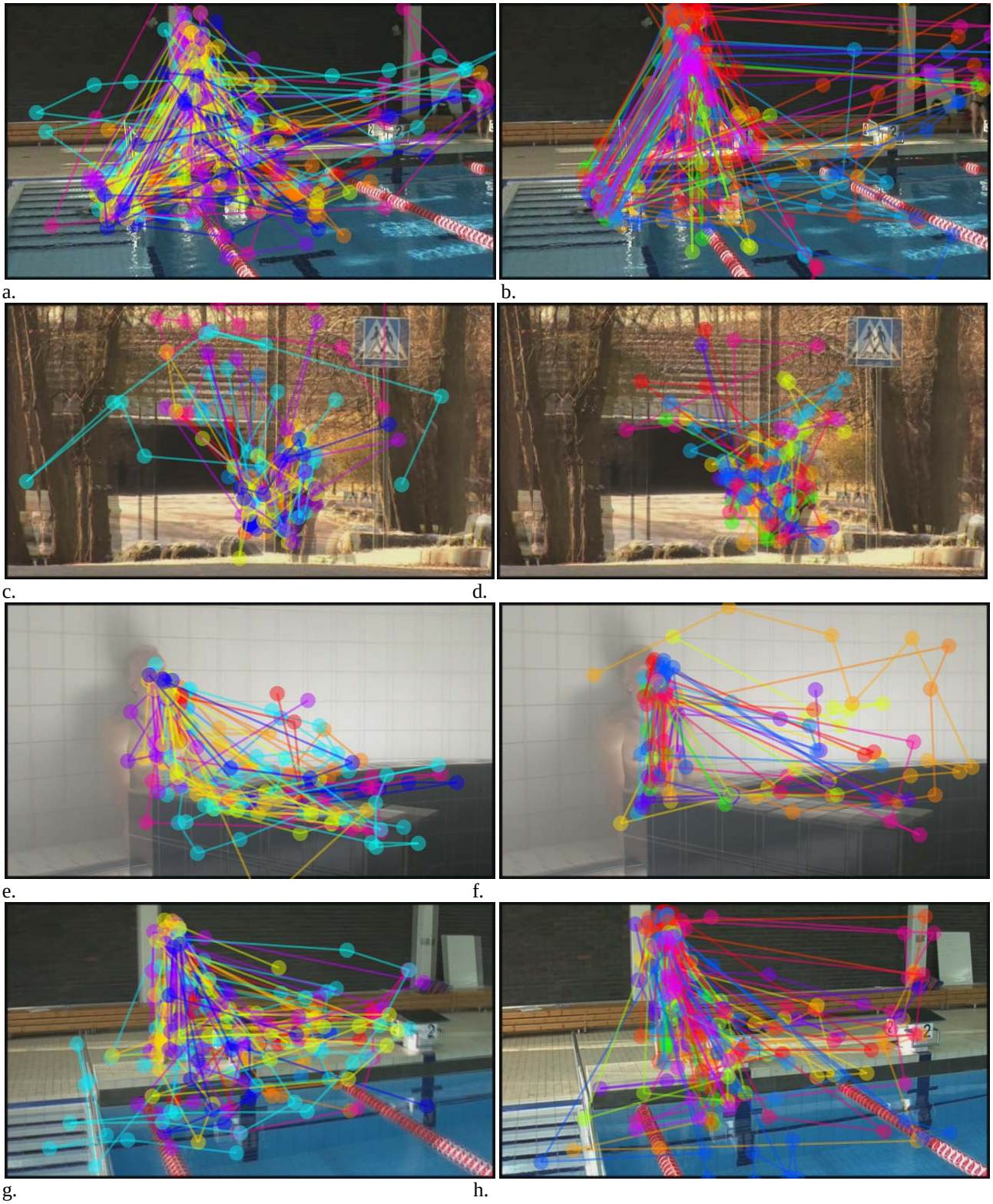


Figure 4. Gaze plots of the four selected shots. Each circle indicates a fixation and each line a saccade. Different colors indicate participants. a) Shot 1 (S3D), b) Shot 1 (2D), c) Shot 2 (S3D), d) Shot 2 (2D) e) Shot 3 (S3D), f) Shot 3 (2D), g) Shot 4 (S3D) h) Shot 4 (2D).

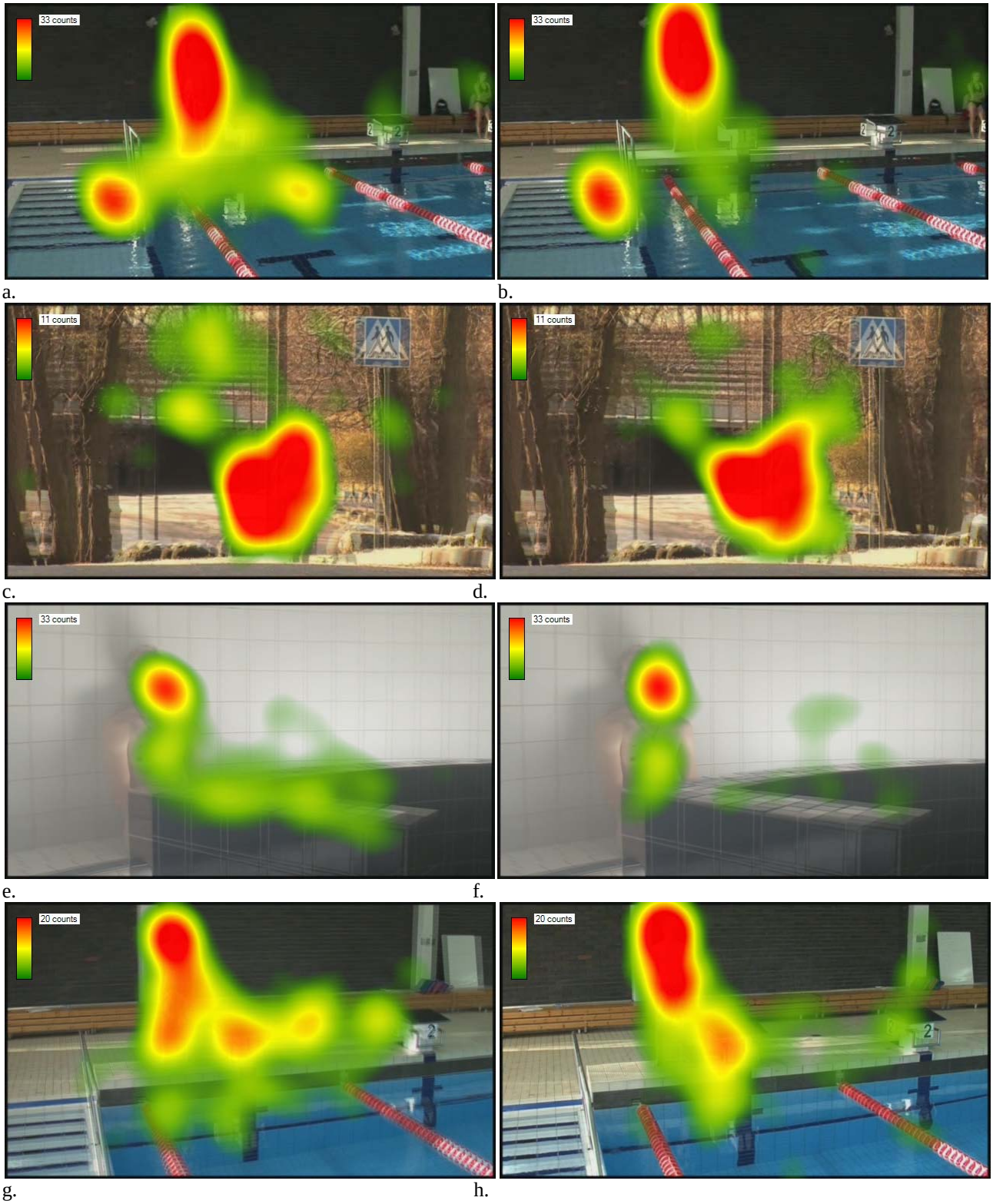



Figure 5. Heat maps of a scenes analyzed in the study. The red color indicates higher number of fixations, and yellow and green colors lower number of fixations. a) Shot 1 (S3D), b) Shot 1 (2D), c) Shot 2 (S3D), d) Shot 2 (2D), e) Shot 3 (S3D), f) Shot 3 (2D), g) Shot 4 (S3D) h) Shot 4 (2D)

## 4. DISCUSSION

When viewing movies, the eye movements of the viewers are focused to the actors and their immediate vicinity. The eye movement patterns show that the viewers are seeking socially relevant information that helps them in understanding the mental state of the actors and the relations between actors. The preference to look at the actors is present immediately at the beginning of the shot, which indicates early visual analysis of this information. The preference is probably related to the significance of social signals in our visual environment, but is also affected by the narrative structure of the movie, which leads the viewers to search for the main actors in each shot. In the S3D versions the eye movements are more widely spread. The viewers' eye movements show them exploring the details of interesting three-dimensional structures present in the shots, and the initial preference to look at actors is slightly diminished. The structures that catch the attention of the viewers are the ones coming toward the observer, like the front tiles of the shot 3, but sometimes they are also objects that are otherwise structurally fascinating, like water in shot 1.

Interesting research question for further research is whether the wider spread of eye movements in stereoscopic shots is caused by bottom-up visual processing, i.e., because the three-dimensional objects automatically capture the attention and eye movements of the viewers, or whether the effect is caused by top-down processes which are related to visible three-dimensional details and interesting structures that the participants want to explore. It is probable that both types of processes affect the final pattern of eye movements, but the exact relationship of top-down versus bottom-up processing remains to be investigated.

## ACKNOWLEDGEMENTS

This study is funded by the Academy of Finland. We thank Paul Lindroos, Eero-Matti Koivisto, Anu Welling, Anni Hirsaho and Topi Ruokolainen for their help in conducting the experiments and analyzing the data. We also thank Jaakko Sipari for technical assistance in building the experimental setup and Stereoscape Inc for providing the stereoscopic movie.

## REFERENCES

1. K. Thompson, *Storytelling in the new Hollywood,* President and Fellows of Harvard College, United States of America, 1999.
2. J. Monaco, *How to read a film,* Oxford University Press, United States of America, 1981.
3. J. Häkkinen, T. Kawai, J. Takatalo, T. Leisti, J. Radun, A. Hirsaho and G. Nyman, "Measuring stereoscopic image quality experience with interpretation based quality methodology," in Proceedings of the IS&T/SPIE's International Symposium on Electronic Imaging 2006: Imaging Quality and System Performance V, S. P. Farnand and F. Gaykema, ed., *Proc.SPIE* **6808**, pp. 68081B-68081B-12, 2008.
4. M. Pölönen, M. Salmimaa, V. Aaltonen, J. Häkkinen and J. Takatalo, "Subjective measures of presence and discomfort in viewers of color-separation-based stereoscopic cinema," *Journal of the Society for Information Display* **17,** pp. 459-466, 2009.
5. G. Nyman, J. Radun, T. Leisti, J. Oja, H. Ojanen, J. -. Olives, T. Vuori and J. Häkkinen, "What do users really perceive – probing the subjective image quality experience," in Proceedings of the IS&T/SPIE's International Symposium on Electronic Imaging 2006: Imaging Quality and System Performance III, S. P. Farnand and F. Gaykema, ed., *Proc.SPIE* **6059**, pp. 1-7, 2006.
6. J. Takatalo, J. Häkkinen, J. Komulainen, H. Särkelä and G. Nyman, "The impact of the display type and content to a game adaptation," in *Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and services, ACM International Conference Proceeding Series* **159,** pp. 17-20, 2006.
7. J. Takatalo, J. Häkkinen, J. Komulainen, H. Särkelä and G. Nyman, "Involvement and presence in digital gaming," in *Proceedings of the 4th Nordic Conference on Human-computer Interaction, ACM International Conference Proceeding Series* **189,** pp. 393-396, 2006.
8. J. Radun, T. Leisti, J. Häkkinen, G. Nyman, J. Olives, H. Ojanen and T. Vuori, "Content and quality: Interpretation-based estimation of image quality," *ACM Transactions on Applied Perception* **4**, pp. 1-15, 2008.
9. J.M. Henderson and A. Hollingworth, "Eye movements during scene viewing: An overview," in *Eye guidance in reading and scene perception,* G. Underwood, ed.,´pp. 269-293, Elsevier Science, Oxford, UK, 1998.

10. L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Reviews Neuroscience* **2,** pp. 194-203, 2001.
11. A. Treisman, "Preattentive processing in vision," *Computer Vision, Graphics and Image Processing*, pp. 156-177, 1985.
12. K. Nakayama and G. H. Silverman, "Serial and parallel processing of visual conjunctions," *Nature* **320,** pp. 264-265, 1986.
13. E. McSorley and J. M. Findlay, "Visual search in depth," *Vision Research* **41,** pp. 3487-3496, 2001.
14. J.M. Findlay, "Saccade target selection during visual search," *Vision Research* **37,** pp. 617-631, 1997.
15. O. Le Meur, P. Le Callet and D. Barba, "Predicting visual fixations on video based on low-level visual features," *Vision Research* **47,** pp. 2483-2498, 2007.
16. P. de Graef, "Prefixational object perception in scenes: Objects popping out of schemas," *in Eye guidance in reading and scene perception*, G. Underwood, ed., pp. 313-336, Elsevier Science, Oxford, UK, 1998.
17. J.M. Henderson and A. Hollingworth, "High-level scene perception," *Annual Review of Psychology* **50,** pp. 243-271, 1999.
18. A.L. Yarbus, *Eye movements and vision*, Plenum, New York, United States of America, 1967.
19. E. Birmingham, W. F. Bischof and A. Kingstone, "Saliency does not account for fixations to eyes within social scenes," *Vision Research* **49,** pp. 2992-3000, 2009.
20. V. Tosi, L. Mecacci and E. Pasquali, "Scanning eye movements made when viewing film: preliminary observations," *International Journal of Neuroscience* **92,** pp. 47-52, 2001.
21. R.B. Goldstein, R. L. Woods and E. Peli, "Where people look when watching movies: Do all viewers look at the same place?" *Computers in Biology and Medicine* **37,** pp. 957-964, 2007.