An investigation of visual selection priority of objects with texture and crossed and uncrossed disparities

Dar'ya Khaustova, Jérôme Fournier, Emmanuel Wyckens – Orange Labs, Cesson-Sévigné, France;
Olivier Le Meur – University of Rennes 1, France

ABSTRACT

The aim of this research is to understand the difference in visual attention to 2D and 3D content depending on texture and amount of depth. Two experiments were conducted using an eye-tracker and a 3DTV display. Collected fixation data were used to build saliency maps and to analyze the differences between 2D and 3D conditions. In the first experiment 51 observers participated in the test. Using scenes that contained objects with crossed disparity, it was discovered that such objects are the most salient, even if observers experience discomfort due to the high level of disparity. The goal of the second experiment is to decide whether depth is a determinative factor for visual attention. During the experiment, 28 observers watched the scenes that contained objects with crossed and uncrossed disparities. We evaluated features influencing the saliency of the objects in stereoscopic conditions by using contents with low-level visual features. With univariate tests of significance (MANOVA), it was detected that texture is more important than depth for selection of objects. Objects with crossed disparity are significantly more important for selection processes when compared to 2D. However, objects with uncrossed disparity have the same influence on visual attention as 2D objects. Analysis of eye-movements indicated that there is no difference in saccade length. Fixation durations were significantly higher in stereoscopic conditions for low-level stimuli than in 2D. We believe that these experiments can help to refine existing models of visual attention for 3D content.

Keywords: stereoscopic images, visual attention, eye-tracking, texture, binocular disparity, visual discomfort, depth, priority of selection

1. INTRODUCTION

Visual attention is the process of selecting important areas of interest out of all the abundant visual information that humans receive in everyday life [1]. Visual attention is usually studied with eye-tracking experiments. There are two ways to analyze eye-tracking data. The first one is to perform an analysis at a specific moment of time. Such analyses study the eye movements: saccades and fixations. Saccades are quick eye movements that shift from one fixation location to another. Fixations are slow eye movements that direct a small part of the visual field into the fovea in order to accurately inspect the location of the stimulus. The second type of analysis is done for the entire observation period and includes an analysis of gaze patterns and the calculation of saliency maps. Gaze patterns depend on two main mechanisms of visual attention: bottom-up (a stimulus-dependent mechanism) and top-down (an observer dependent mechanism). Bottom-up is driven by low-level features [2], [3] with eye movements that are involuntary and unconscious. Top-down attention integrates high-level cognitive processes like prior knowledge, task, and experience [4].

Models of visual attention are used as a tool to define the regions of interest (ROI) of visual stimuli. In the case of stereoscopic video selected ROI can be used to perform 3D ROI-based encoding [5], content repurposing [6] ,and for the control of the visual discomfort. Thus, there is a need to develop a 3D visual attention model. It is possible that the easiest approach is to adapt existing 2D models of visual attention to stereoscopic conditions. Therefore, it is necessary to reveal differences when observing 3D stimuli in comparison to 2D. There are a number of recent studies concerning 3D visual attention that explore how stereopsis influences our perception.

Jansen *et al.* [7] studied the influence of disparity on fixations and saccades in free viewing of 2D and 3D images of natural scenes, pink noise, and white noise. An analysis was performed using the data from the left eye. They found that

disparity information had an influence on basic eye movements, causing an increase in the number of fixations, a decrease of fixation duration over time but only for pink and white noise; and a shortening of saccade length over time. The saliency of mean luminance, luminance contrast, and texture contrast was compatible across 2D and 3D stimuli. Mean disparity had a time dependent effect for 3D stimuli. Disparity contrast was elevated at fixated regions in 3D noise images but not in 3D natural scenes. They reported that participants fixated closer locations earlier than more distant locations in the image. Previous works were supplemented by Wismeijer *et al*. [8]. They investigated whether saccades are aligned with an individual depth cue, or with a combination of depth cues. They presented an incline in depth surfaces that combined monocular perspective cues and binocular disparity cues in order to specify different plane orientations with different degrees of small and large conflicts between the two sets of cues. They discovered that the distributions of spontaneous saccade directions followed the same pattern of depth cue combination as perceived surface orientation: a weighted linear combination of cues for small conflicts, and cue dominance for large conflicts. By examining the relationship between vergence and depth cues, they reached the same conclusion as Posner [1]: that vergence is dominated only by binocular disparity. Another study [9] analyzed differences in the distribution of fixation points in 3D conditions in comparison with 2D. During the eye-tracking experiment, fixation points were collected in mono and stereo conditions from 66 images. By examining image contours, depth contours, disparity changes between fixation points, and the clustering of fixation points, only slight differences were found in the special distribution of fixation points. Huynh-Thu *et al*. [10] found that the average fixation frequency and the average fixation durations were smaller when viewing 3D stereoscopic content, while the average saccade velocity was higher when viewing 3D stereoscopic content.

Several models of 3D visual attention already exist. Zhang *et al*. in [11] proposed a bottom-up stereoscopic video attention model. It accounts for perceived depth, which is calculated from the camera's focal length, baseline distance, and disparity estimated per pixel. They take into consideration that closer objects are more important for visual attention than farther objects. As in previous model, the proposal of Wang *et al*. in [12] is based upon perceived depth, which is not calculated from camera parameters, but from the viewing distance to the display and the interocular distance of the observer. This representation of a depth map should better correlate with human perception, since perceived depth always depends on target screen size, viewing distance, and interocular distance. This model combines the generated depth saliency map with a 2D saliency map (computed by any 2D visual attention model). The resulting combination provides a saliency map for still 3D images, unlike the model of Zhang *et al*., where orientation and motion contrast are taken into account. Another model is a time-dependent computational model to predict saliency on still pictures proposed by Gautier & Le Meur [13]. Their proposed approach combines low-level visual features along with center and depth biases. The influence of disparity on saliency is investigated as well. Their results claim that visual exploration is affected by the introduction of the binocular disparity, i.e., the participants tend to first look at closer areas (in terms of depth) and then direct their gaze to more widespread locations.

In our previous work [14] , we estimated the influence of depth, comfort/discomfort, and texture complexity on visual attention. Stereoscopic contents with uncrossed disparity were used as stimuli. We did not notice any pronounced influence of uncrossed disparity on visual attention, even when stimuli were uncomfortable to watch. By calculating inter-observer visual congruency to assess the influence of texture on visual attention, we discovered that texture with high complexity attracts attention independently of the amount of disparity. Thus, the question whether objects with crossed disparity influence visual attention remains unanswered. To answer this question, we present an experiment that studies visual attention for objects with crossed disparity in comparison to 2D objects in section two. In section three, we study the role of depth and texture to determine which is more influential in the process of object selection. In section four, we provide some conclusions about the influence of depth on visual attention.

2. CROSSED DISPARITY EXPERIMENT

The goal of this experiment is to continue the research described in [14], where we studied visual attention in 3D using stimuli with only uncrossed disparity. In order to complete the previous study we designed a new experiment using 2D, 3D comfortable, and 3D uncomfortable still images with objects with crossed disparity (only in the case of 3D).

2.1. Stimulus generation

The key point of the experiment is to present stimuli with an object in front of a display with a controlled amount of depth to the observer. The depth range is controlled by changing the 3D camera baseline and the convergence distance.

In order to compute these parameters, we used the Stereo Calculator software developed by Orange Labs. The calculations can be accomplished when a display parameters (image resolution, size, viewing distance), a camera parameters (focal length, sensor size), a scene parameters (foreground, background distance of a scene, region of interest) and human visual attributes (depth of focus, inter-pupil baseline) are known. To ensure visual comfort while watching 3DTV, the software uses rules based on the optimization of a stereoscopic distortion and a comfortable viewing zone[15]. Discomfort can be avoided when the depth of focus (DoF) does not exceed ±0.2 diopters [16, 17]. The input value of DoF=0.1 diopters was used for comfortable viewing and DoF=0.3 diopters for uncomfortable. These thresholds were selected based on the former research done by Orange Labs [18].

All scenes were designed and rendered using Blender software, which allows the measurement of foreground and background distances of a scene and the accurate control of stereoscopic camera parameters. Four different scenes were selected for the experiment: “Cartoon”, “Hall”, “Pigs”, and “Table” [19]. Figure 1 illustrates examples of the generated scenes. All objects with crossed disparity were selected in a way to avoid windows violation effect. For example, in Figure 1.b, the lamp stand was behind the display plane, while the lampshade was coming out of the screen.

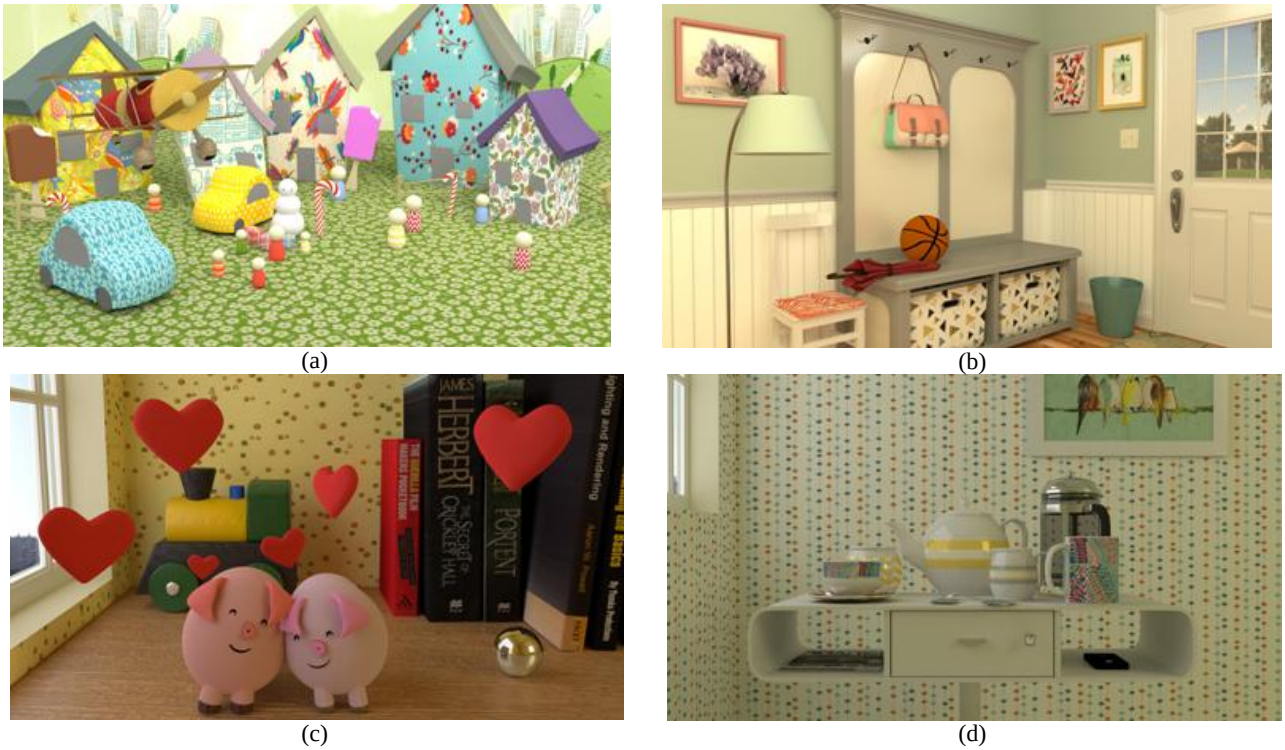


Figure 1. Stimuli used in the experiment: (a) Cartoon, (b) Hall, (c) Pigs, (d) Table.

The camera parameters calculated with Stereo Calculator are presented in Table 1.

Table 1. Scenes parameters.

Scene	DoF, diopters	Baseline, mm	Convergence distance, m	Disparity on the screen, mm	
				foreground	background
cartoon	0.1	220	8	-15	15
	0.3	660	8	-46	46
hall	0.1	342	8	-15	15
	0.3	1020	8	-46	46
pigs	0.1	53	2.2	-15	15
	0.3	162	2.2	-47	45
table	0.1	325	5.45	-16	15
	0.3	1000	5.45	-45	44

The main focus of this study is the influence of depth with crossed disparity on visual attention, so texture complexity was not taken into account. In total, 12 still images were generated (4 scenes×3 depth levels). Since it was important to prevent observers from memorizing the stimuli and hence using top-down visual mechanisms, we arranged 3 sets containing 4 images with different contents and different depth levels.

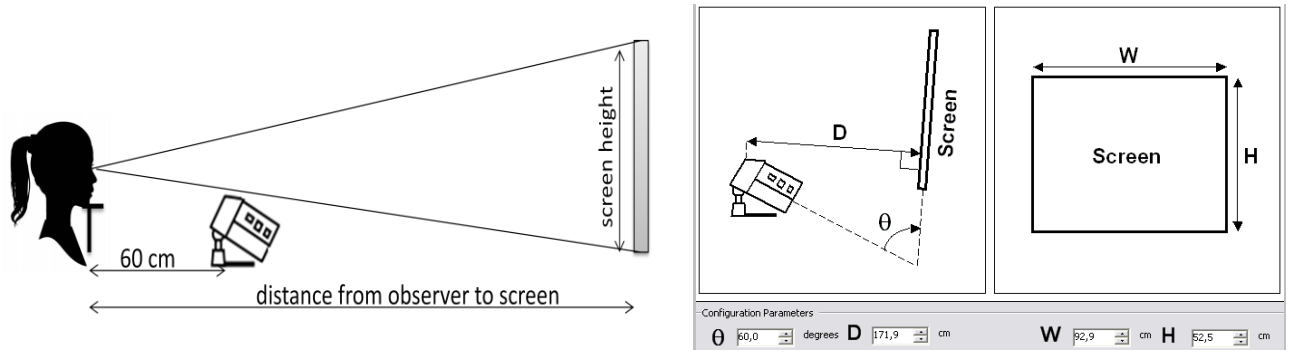
2.2. Experimental set-up

The psychophysical test set-up is schematically presented in Figure 2. It consists of:

1. 42" LG 42LW line interleaved stereoscopic display, 93×52 cm; resolution in 2D 1920×1080, in 3D 1920×540 per view;
2. 3D passive glasses;
3. Tobii x50 eye-tracker to record reflected patterns from the corneas of the eyes;
4. Chin rest to restrict head motion.

The distance from the observer to the screen was 4.5H (2.34 m). The distance from the observer to the eye tracker was 60 cm.

There are some restrictions on the placement of the Tobii eye-tracker: first, the distance from the observer to the eye tracker should be around 60 cm; second, the eye tracker should be positioned straight in front of the stimuli and at a particular angle below the user (see Figure 2.b). But once it has been configured, eye tracking is fully automatic.



(a) (b)

Figure 2. Experimental set-up: (a) Tobii x50 configuration tool, (b) General scheme of the set-up of the experiment.

To build saliency maps, it is necessary to calculate the number of pixels per one degree. Taking into account the width of the screen (SW=93 cm) and the distance from the observer to the screen (SD=234 cm), we applied trigonometry rules and calculated the visual half angle as a basis.

$$\tan\left(\frac{\alpha}{2}\right) = \frac{SW/2}{SD} \quad (1)$$

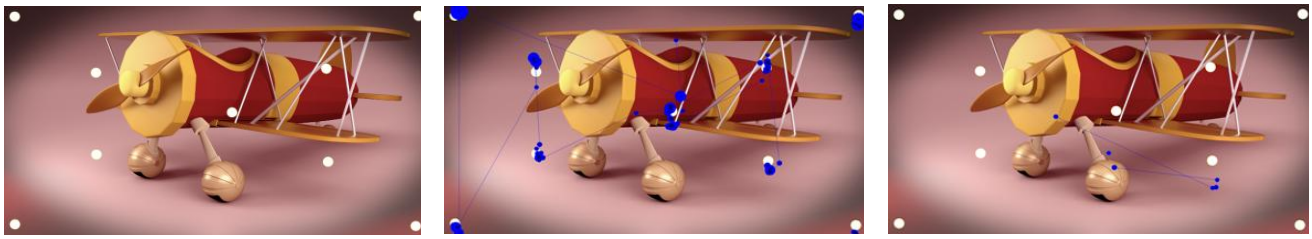
Therefore, the number of pixels per one degree is $\frac{\text{horisontal resolution}}{\alpha} = \frac{1920}{22.47^\circ} \approx 85$ pixels.

2.3. Experimental methodology

For each observer, the experiment consisted of five stages: visual test, reading of the instruction sheet, calibration, training, and, finally, the visual attention test. Monocular acuity, color vision, far vision test, fusion test, and stereoscopic acuity of the observers were checked using Essilor ERGOVISION equipment prior to the visual attention test. The instruction sheet offered some explanations on how to behave during the calibration stage, the training stage, and during the test itself. Observers were allowed to look at the images freely without any instructions.

The eye tracker requires a calibration to learn the characteristics of the eyes of each observer. The observers were asked to put on the passive polarized glasses in order to begin the calibration. During the calibration stage, an observer simply looked at a dot that appeared in different positions of the screen. The calibration procedure was fully automatic and took about 30 seconds. A five-point calibration procedure was used in our experiment. Even if the software reported that the calibration was done successfully, there were still a few special circumstances in which the system had tracking difficulties, such as for people with bi-focal glasses or people with elements (eye lids, mascara, etc.) that significantly block the eye tracker camera's view of the subject's eyes. Thus, after the automatic calibration, a specially designed chart was used in order to check whether the device was able to track the observer's gaze correctly (Figure 3a). The

calibration image contained nine white dots on the top of a picture of an airplane. Observers were instructed to focus on every white point for 3 seconds. Figure 3b presents an example of a successful calibration. If the eye-tracker had difficulties properly recording the data of an observer or the calibration process was unsuccessful, the resulting gaze plot looked similar to Figure 3c. No observers with unsuccessful gaze plots were allowed to participate in the test.



(a) (b) (c)

Figure 3. Calibration phase: (a) Calibration chart, (b) successful gaze plot, (c) unsuccessful gaze plot.

The training was done in stereoscopic mode using three images with three levels of depth: DoF=0 diopters (2D), DoF=0.1 diopters, DoF=0.3 diopters. Each image was presented for 20 seconds and separated from the subsequent one by displaying a gray screen for 5 seconds. The duration of the training was 1 minute 20 seconds. The images were different from those used in the test. The training phase was designed to familiarize observers with the test conditions.

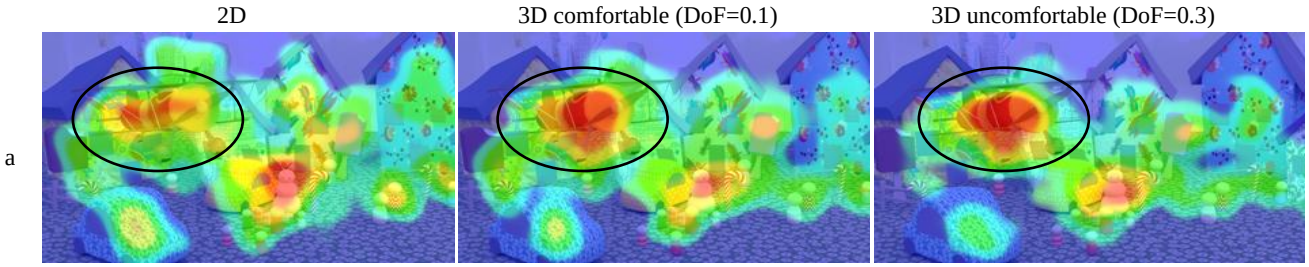
During the test, only one of the three sets of images was displayed. The duration of the test was 1 minute 40 seconds. 51 people (36 males and 15 females from 22 to 52 years old) participated in the test. Each image was tested on 17 observers.

2.4. Eye tracking data analysis

In this section we analyze eye-tracking data and study whether the introduction of crossed disparity had an effect on basic eye movement properties. The entire fixation data collected with the Tobii eye-tracker was used for analysis. All observers that could not complete the calibration process using the calibration chart were excluded before the test. In order to analyze the gaze behavior of observers, saliency and heat maps as well as fixation durations and length of saccades were computed [20, 21]. During the experiment, images were separated by a gray slide without a fixation cross in the center, which is why the first fixation of each stimulus was not discarded.

2.4.1. Qualitative analysis based on heat maps

We used heat maps representing fixated areas of a stimulus to compare the gaze patterns of all observers. This method has been used in various studies [8-10] due to its fast and convenient visualization ability of gaze behavior of an entire group of observers. The heat map consists of the stimuli as a background image (Figure 1) and a hotspot mask superimposed on top of it. A hotspot mask is a color scale, which is normalized depending on the number of fixations. Red indicates the highest number of fixations and blue – no fixations. It should be noted that the normalization process is done for each heat map independently. As a consequence, it is difficult to compare heat maps for different scenes precisely.



b

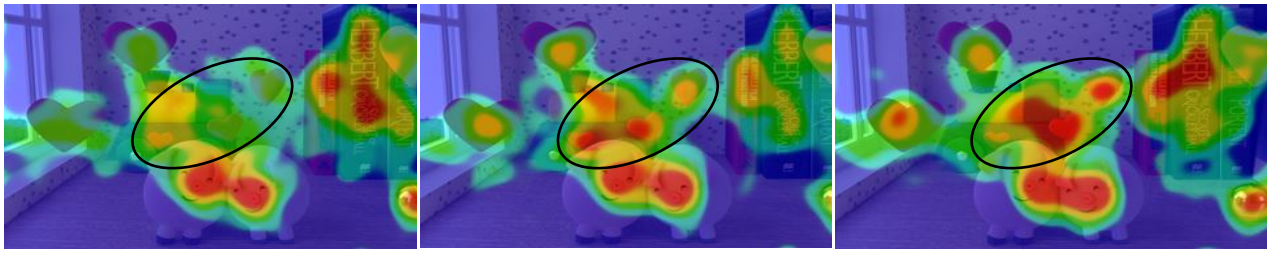


Figure 4. Heat maps for scenes (a) Room_lt, (b) Room_mt, (c) Room_ht

For example, several heat maps are shown in Figure 4 for the scenes “Cartoon” and “Pigs”; each image on this figure shows the heat map corresponding to a viewing duration of 20 seconds. In stereoscopic condition, the airplane in Figure 4.a attracts attention independent of the caused discomfort. In the 2D case, the most fixations (the large red spot) are on the snowman, while in 3D, the most fixations are on the airplane. The same situation is observed in Figure 4.b, where small hearts, which pop-out of the screen, became the main region of interest with depth and received the most fixations. We observed similar behavior for the other two scenes.

2.4.2. Quantitative analysis based on heat maps

After computing the saliency maps for each scene, which represent the density of fixations for an entire image, it was interesting to investigate their differences by calculating the correlations between pairs. As metrics, the Pearson linear correlation coefficient (CC) and Area Under Curve (AUC) were used. In the case of AUC, a higher value means a better correlation: a value of 1.0 indicates a perfect performance, while a value of 0.5 demonstrates a random performance [16]. The results for the AUC and CC metrics are presented in Table 2. The highest result for each column is marked in bold.

Table 2. AUC and CC correlation values between 2D and 3D DoF=0.1 (2D/01) saliency maps; between 2D and 3D DoF=0.3 (2D/03) saliency maps; between 3D DoF=0.1 and 3D DoF=0.3 (01/03) saliency maps.

	AUC			CC		
	2D/01	2D/03	01/03	2D/01	2D/03	01/03
cartoon	0,82	0,87	0,84	0,80	**0,94**	0,84
hall	0,84	0,85	0,83	**0,89**	0,87	0,87
pigs	0,80	0,84	0,84	0,84	0,87	0,85
table	**0,88**	**0,90**	**0,88**	0,87	0,87	**0,93**

All AUC and CC values presented in Table 2 are very high. High correlation indicates that there is no strong difference between saliency maps. This implies that visual attention is not affected by different disparities, which is in contradiction with our qualitative results. A viewing duration of 20 seconds is sufficient time for an observer to investigate every object in a still image. Consequently, saliency maps differ mainly in the density of fixations, which cannot be detected by AUC and CC metrics. Hence the difference between 2D and 3D conditions could only be discovered by comparing the number of fixations on the objects.

Huynh-Thu *et al*. performed a quantitative comparison between 2D and 3D using AUC and CC metrics. They found that the differences between saliency maps depend on the content whereas our data presented in Table 2 demonstrate a very slight difference between different scenes. One possible reason for such a contradiction is that in the Huynh-Thu *et al*. experiment, videos of different duration (from 8 to 143 seconds) were used, while we used still images with a viewing duration of 20 seconds.

2.4.3. Saccade length

Saccade length is the distance measured between locations of two fixation points in degrees. Figure 5.a presents the saccade length for every image. There is no clear tendency for saccade length, which seems to be content dependent. With a paired samples t-test we did not find any significant differences between saccade length for 2D and 3D conditions. We believe that there are not enough observations to prove a statistical significance in our case. Figure 5.b presents the average saccade length for all the scenes. It does not corroborate results from Jansen *et al.* or from Huynh-Thu *et al.*, who found that saccades in 3D were shorter and faster.

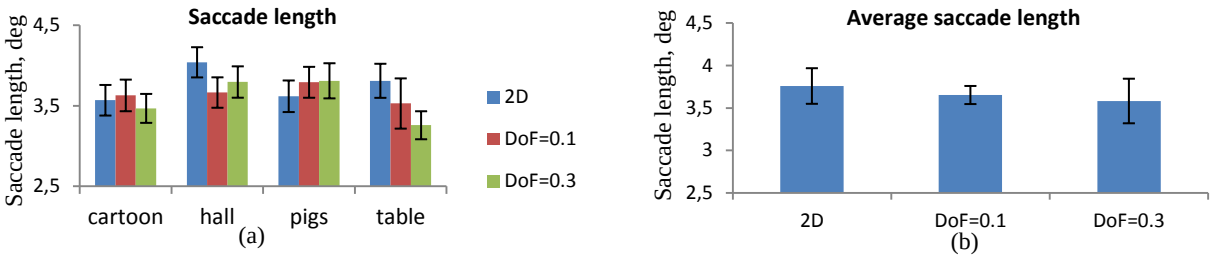


Figure 5. Influence of depth on (a) length of saccade for every image, (b) average saccade length. Error bars represent 95% of confidence interval.

2.4.4. Fixation durations

The statistical analysis of fixation durations showed that there is no relation between the fixation durations and depth levels (Figure 6.a). With a paired samples t-test, we did not find any significant influence of depth on fixation duration. Figure 6.b presents the average fixation duration for all the scenes. Our results do not corroborate the findings of Huyanh-Thu *et al*. (fixation durations were longer in 2D) nor Jansen *et al*. (fixation durations were longer in 3D). Unlike in our experiment, videos were used as stimuli in [10] and still images with uncrossed disparities in [7].

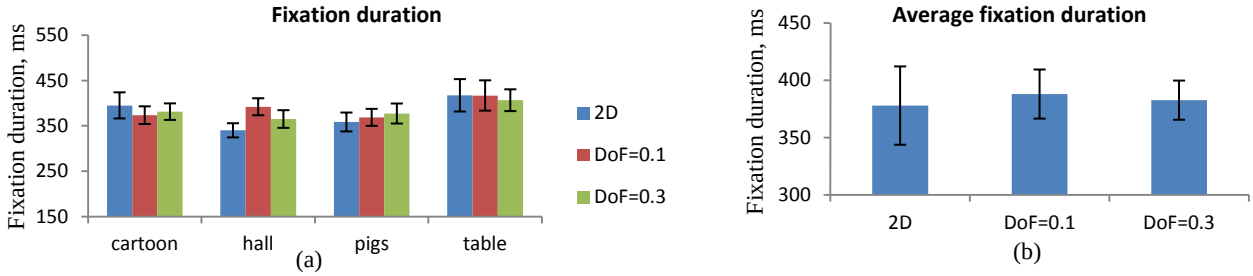


Figure 6. Influence of depth on (a) fixation duration for every image, (b) average fixation duration. Error bars represent 95% confidence interval.

3. SPHERES EXPERIMENT

The goal of the second experiment is to verify the hypothesis that texture contrast is a more influential factor in guiding our gaze than the amount of depth. By performing the eye-tracking experiment and studying gaze plots of the observers, the priority in selection of the objects in depth is investigated. Results are compared to the 2D condition.

3.1. Stimulus generation

Each stimulus contained four spheres equidistant from the center of the screen on a gray background. Any of the four spheres could be in one out of five possible locations in depth:

1) in front of a display: close to a display plane;
2) in front of a display: far from a display plane;
3) behind a display: close to a display plane;
4) behind a display: far from a display plane;
5) in the display plane.

To study the influence of texture on the selection process, two possibilities were available: a sphere could have the same gray color as the background or a checkerboard texture. By changing the camera baseline and the convergence distance, it was possible to generate three images using the same sphere set-up:

- an image with uncrossed disparity (when all of the spheres are behind the display plane);
- an image mixed disparities (when some spheres are in front of the display plane and the rest are behind);
- an image with crossed disparity (when all of the spheres are in front of the display plane).

The forth option is a 2D image – with a front view of the set-up. Figure 7 depicts one of sphere set-ups: three spheres are in front (that's why one of the spheres is invisible on the top view) and one is behind with checkerboard texture. The display plane is the bold solid line. An observer is in front of the display plane.

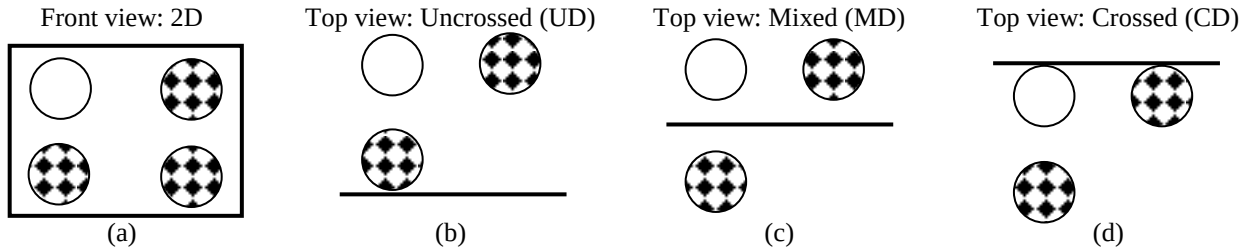


Figure 7. Four images with different disparitites and same sphere set-up (a) 2D image, (b) Image with uncrossed disparity, (c) Image with mixed disparities, (d) Image with crossed disparity. Display plane is marked with the bold solid line.

Images were generated with Blender software and depth calculations were done in the same way as described in paragraph 2.1, the amount of depth was defined as DoF=±0.1 diopters for crossed and mixed disparities and DoF=+0.15 diopters for uncrossed. We used bigger DoF for uncrossed disparities in order to increase the perceived depth because the perceived depth reaches DoF=0.2 (0.1 in front of display and 0.1 behind) in the case with mixed disparities. The calculated parameters are presented in Table 3 below.

Table 3. Calculated parameters for images with uncrossed (UD), mixed (MD) and crossed (CD) disparities.

Scene	DoF, diopters	Baseline, mm	Convergence distance, m	Disparity on the screen, mm	
				foreground	background
UD	+0.15	315	6	-	23
MD	±0.1	500	8	-15	15
CD	- 0.1	300	9.5	-15	-

In total, 56 images were generated. We used 14 different sphere set-ups (Figure 8) with 4 variations of disparities, as explained above. Each set-up in Figure 8 represents a 2D stimulus or a front view of the 3D representation. Spheres marked in bold are closer to the observer in depth, independent of disparity type. In 2D, all four spheres are on the display plane. The name of the stimulus consists of the corresponding number or sphere set-up and the designation of disparities presented in a stimulus. For example, "11_MD" means sphere set-up 11 (Figure 8) with mixed disparities (Figure 7c); the sphere in the bottom left corner comes out of the screen and three other spheres are behind the display plane.

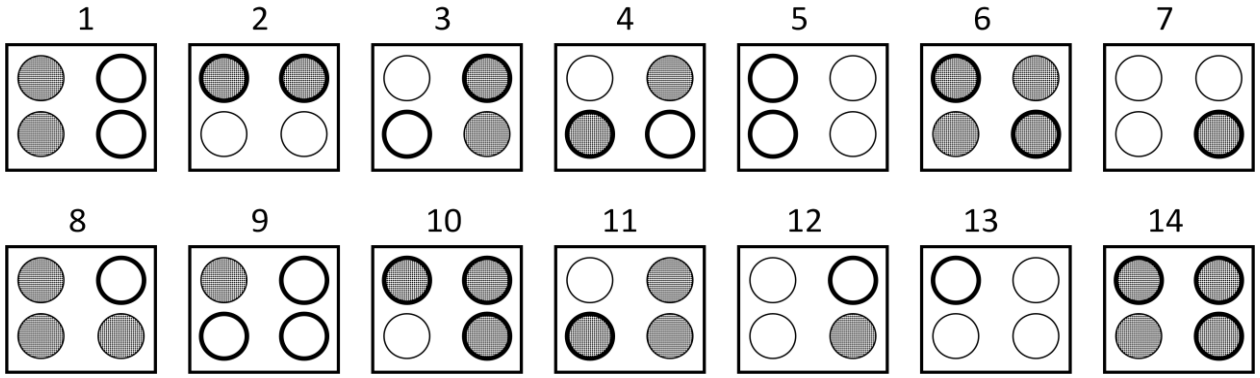


Figure 8. Scene set-ups used in experiment. Spheres marked in bold are closer to observer in depth independent of disparity type.

3.2. Experimental set-up

The same experimental set-up as in paragraph 2.2 was used in this experiment.

3.3. Experimental methodology

We used the same experimental methodology as in paragraph 2.3. The training was done in stereoscopic mode using 10 still images with various sphere set-ups and disparity combinations. Each image was presented for 5 seconds and

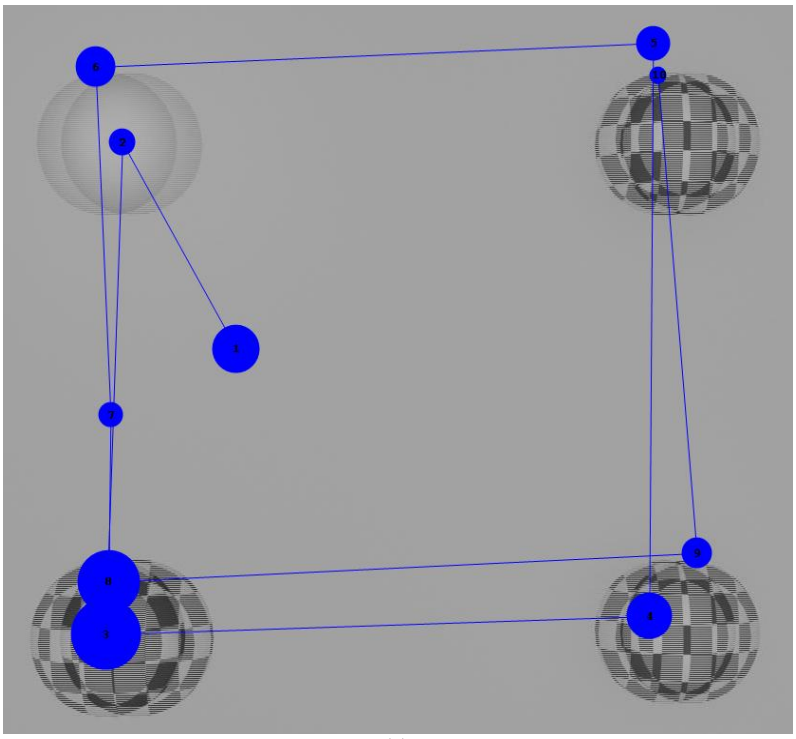
separated from the subsequent one by displaying a gray screen for 5 seconds. The training phase was designed to familiarize observers with the test conditions. The duration of the training was 1 min 40 seconds.

During the test, we randomly displayed all the prepared images to every observer. The duration of the test was 9 minutes 30 seconds. 28 people (19 males and 9 females from 20 to 52 years old) participated in the test.

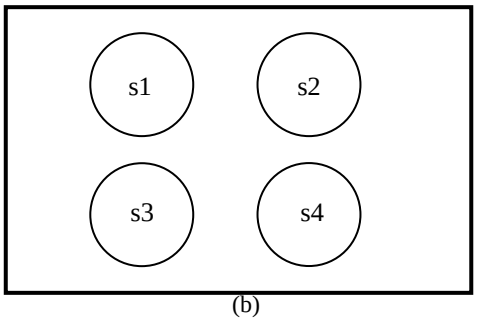
3.4. Eye tracking data analysis

Gaze plots of every observer were analyzed in order to find out if the sphere selection preference was based on texture or depth position. For every observer we created a table, which contained the sphere selection priority for each stimulus. Every sphere had a fixed position number, which is constant within all the images: the sphere in the top left corner is numbered s1, the top right corner – s2, the bottom left – s3, the bottom right – s4 (Figure 9.b). Based on the observer's gaze plot for each position number, the selection order number was collected. Figure 9.a presents the gaze plot of observer 5 for image 11_MD; position number for every sphere presented in Figure 9.b. So observer 5 selected at first the sphere with position number s1 (order is 1), then with position number s3 (order is 2), then with position number s4 (order is 3) and the last sphere with position number s2 (order 4). This data is collected for each observer for all the stimuli. See the example of such a table in Figure 9.c.

It is interesting to note that some observers looked at all of the images the same way during the experiment independent of texture or depth variation. For example, all observations were started at the top left corner and then continued clockwise; or another pattern that was observed is from top left to right and then from bottom left to right. Since the presentation of stimuli was only 5 seconds, supposedly mostly only bottom-up processes should be involved. But it seems that some observers were intentionally following the same pattern of observation during the whole test. Since the duration of the test was almost 10 minutes, we found this behavior unnatural. For 3 observers, we found that for several images all the fixation points were lost. This may have occurred due to some of them closing their eyes during the test or the signal was lost due to head movement or displacement from the chin rest. All these observers were excluded from the data analysis.



(a)



(b)

Order of selection for observer 5 for image 1

Position:	**s1**	**s2**	**s3**	**s4**
11_MD	1	4	2	3
11_2D	1	2	3	4
11_CD	2	4	1	3
11_UD	1	3	2	4

(c)

Figure 9. (a) Gaze plot of image 11_MD with mixed disparities of observer 5 (b) Fixed position number of every sphere (c) Resulting table with order of selection of spheres for image 1 with different disparities for observer 5.

After we converted all the gathered information to one data table, where texture: 0 –gray, 1– checkerboard; depth: 0 – when a sphere has zero disparity, -1 – a sphere with crossed disparity, 1 – a sphere with uncrossed disparity; order is the priority of selection of a given sphere: 1 – selected first, 2 – selected second, etc.; position is a fixed position for spheres for all the images (Figure 9.b). An example for image 11_MD for observer 5 is presented in Table 4.

Table 4. Collected data of the image 11_MD for observer 5.

Image	Observer	Position	Texture	Depth	Order
11_MD	5	s1	0	1	1
11_MD	5	s2	1	1	4
11_MD	5	s3	1	-1	2
11_MD	5	s4	1	1	3

3.4.1. Influence of depth on visual attention

MANOVA univariate tests of significance for order showed that depth significantly influences the order of selection of the spheres $F(2, 4456)=6.63$, $p<0.05$, $p=0.0013$. The analysis was performed for all the data. However, when we performed an analysis of the data separately for crossed and uncrossed disparity, we found that the influence of uncrossed disparity on the order of selection is insignificant in comparison with 2D: $F(1, 3424)=3.35$, $p<0.05$, $p=0.067$. Our previous work substantiates this result. We found that saliency maps of scenes with uncrossed disparity are very similar to 2D saliency maps. Thus, depth was not a deciding factor in directing our gaze towards the objects of interest. Analysis of crossed disparity showed that a sphere with a crossed disparity significantly influences the order of selection in comparison with 2D: $F(1, 3264)=13.14$, $p<0.05$, $p=0.0003$. These results are in accordance with the experiment described above in section 2, where we found that objects with crossed disparity attract our attention. These results are presented in Figure 10.

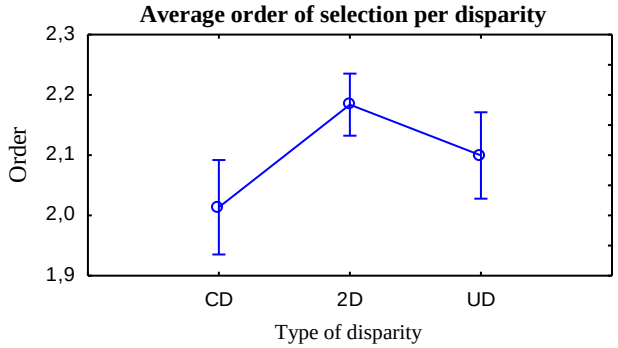


Figure 10. Average order of selection of sphere per disparity. Error bars represent 95% confidence interval.

Spheres with zero disparity, i.e., those that were situated on the display plane, received the lowest selection priority.

3.4.2. Influence of texture on visual attention

Univariate tests of significance for order showed that texture significantly influences sphere selection order $F(1, 4456)=12.31$, $p<0.05$, $p=0.0005$. Also, it is significant for both crossed $F(1, 3264)=9.95$, $p<0.05$, $p=0.0016$ and uncrossed $F(1, 3424)=8.4$, $p<0.05$, $p=0.0038$ disparities. These results are presented in Figure 11.

Independent of the type of the disparity (zero disparity, crossed, or uncrossed) spheres with the checkerboard texture were selected before spheres with no texture. Spheres coming out of the screen with the checkerboard texture had the highest selection priority.

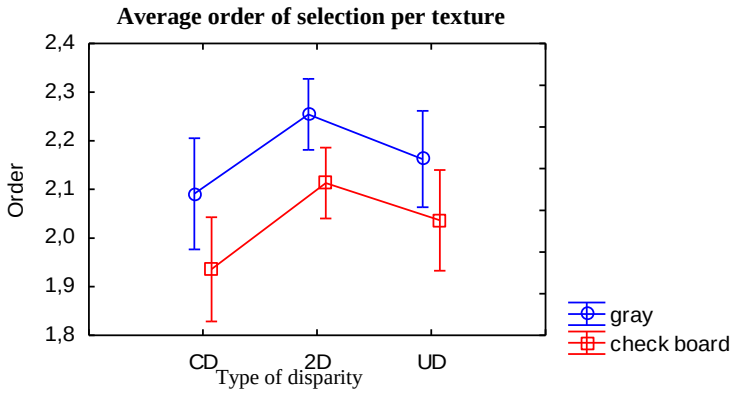


Figure 11. Average order of selection of sphere per texture. Error bars represent 95% confidence interval.

3.4.3. Influence of the position of the spheres on test results

The analysis showed that the position of the sphere significantly influences the selection priority for crossed disparity $F(3, 3264)= 31.14$, $p<0.05$, $p=0.0000001$, as well as for uncrossed disparity $F(3, 3424)=30.17$, $p<0.05$, $p=0.0000001$. There could be several explanations of such results: since the spheres were presented in two rows, observers were followed the familiar reading pattern of top left to top right, then bottom left to bottom right, as if it were a text. The higher number of sphere position resulted in the lower priority of selection. This tendency can be seen for depth (Figure 12.a), as well as for texture (Figure 12.b). Another possible explanation is that the time of presentation of one image was only 5 seconds, hence for some observers it was easier to apply a scheme of observation, for example, clockwise and follow it until the end of the experiment.

The influence of sphere position on selection order is presented in Figure 12.a. In 12.b we can see that spheres with texture are preferred (have a lower order of selection) to gray spheres. If textured spheres were situated in the top left corner of the screen, they had a significantly higher priority of selection than a non-textured sphere in the same position of the screen.

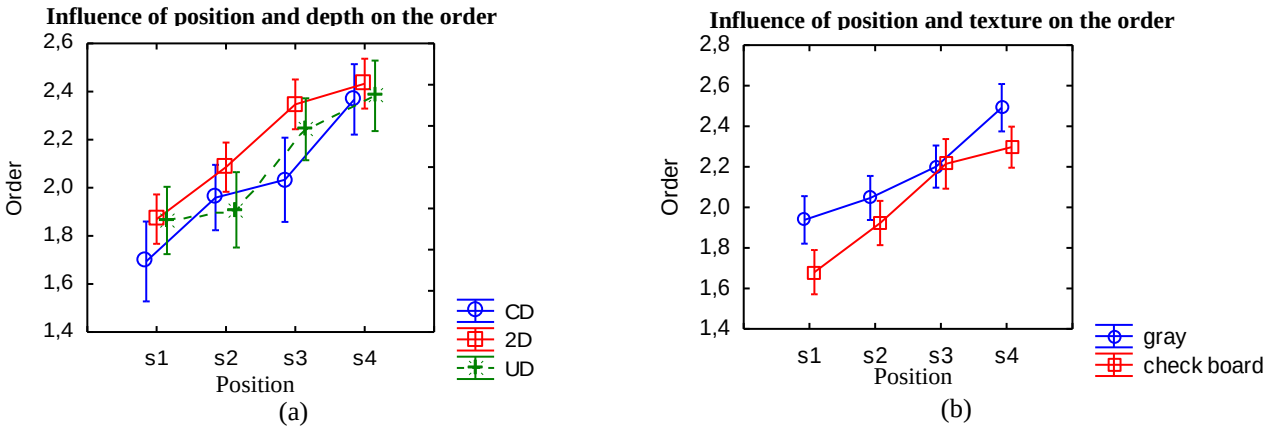


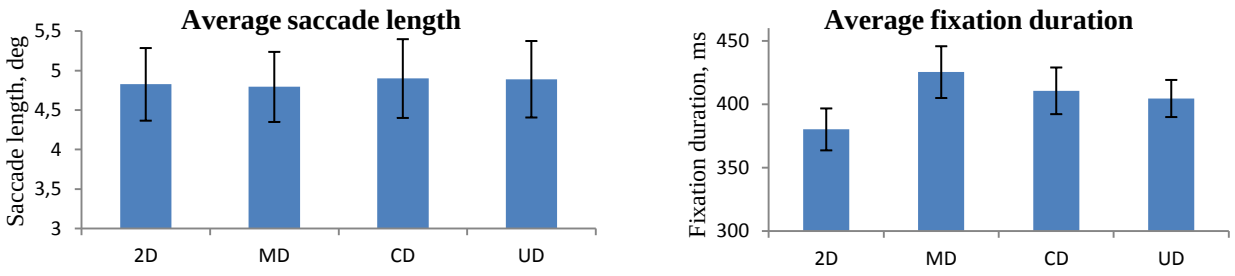
Figure 12. (a) Influence of the position of a sphere and depth on the order of selection (b) Influence of the position of a sphere and texture on the order of selection. Error bars represent 95% confidence interval.

3.4.4. Saccade length and fixation duration

During the eye tracking experiment, we displayed stimuli that were supposed to involve mostly bottom-up processes of visual attention. Hence, it was very interesting to find out if the depth component influenced saccade length and fixation duration in this experiment.

As can be seen from Figure 13.a, the depth component did not have any influence on the average saccade length. The same results were obtained with a paired samples t-test: there is no significant difference between saccade lengths for zero disparity, mixed, crossed, and uncrossed disparities.

The average fixation duration for every disparity is presented in Figure 13.b. Analyses of average fixation duration with a paired samples t-test showed that there is a significant difference for fixation duration $t(13)= -4,06$, $p<0.05$, $p=0.0013$ in scores for zero disparity and mixed disparities, as well as a significant difference between zero disparity and crossed disparity: $t(13)= -2,68$, $p<0.05$, $p=0.019$ and between zero disparity and uncrossed disparity: $t(13)= -3,98$, $p<0.05$, $p=0.0016$. A comparison of the rest of the pairs did not expose any significant results. The introduction of depth increased fixation duration independent of the type of the disparity.



(a) (b)

Figure 13. (a) Average saccade length. (b) Average fixation duration. Error bars represent 95% confidence interval.

4. CONCLUSIONS

This research was launched to generalize the studies in understanding the difference in spatial observation of 2D and 3D image content depending on texture and amount of depth. In our previous study [14], we found that the visual strategy was very similar when observing 2D images and 3D images with uncrossed disparity. To expand on the previous work, we designed an eye-tracking experiment using stimuli with crossed disparity. The difference between visual strategies when observers watch 3D images with crossed disparity (in comfortable and uncomfortable conditions) and 2D images was examined. We found that objects located in front of the display plane are more salient than objects with uncrossed disparity or 2D, even if observers experience discomfort from excessive disparity.

In the second experiment presented in this study, we evaluated features influencing the saliency of the objects in stereoscopic conditions by using content with low-level visual stimuli. We detected that textured objects are selected before non-textured independent of their depth position. Objects with crossed disparity are significantly important for selection process as well. However, there was no difference in selection preference for objects with uncrossed disparity in comparison to 2D objects. Thus, visual attention of images with uncrossed disparity is similar to 2D and, for computing a saliency map of such images, any of existing 2D visual attention models can be used. The influence of position on the screen of a sphere had a significant impact on the selection priority. This can be explained by the design of the stimuli: two objects in the top row, two objects in the bottom row, which can be read in the familiar way from top left to top right, then from bottom left to bottom right.

We analyzed eye movements in both experiments and did not find any significant difference between 2D and 3D conditions for average saccade length. Average fixation durations were higher when viewing stimuli with spheres in 3D.

REFERENCES

[1] Posner, M., "Orienting of attention," *The Quarterly Journal of Experimental Psychology*, vol. 32, pp. 3-25, (1980).

[2] Koch, C., Fau - Ullman, S., and Ullman, S., "Shifts in selective visual attention: towards the underlying neural circuitry," 19860916 DCOM- 19860916, (1985).

[3] Itti, L., Koch, C., and Niebur, E., "A model of saliency-based visual attention for rapid scene analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, pp. 1254-1259, (1998).
[4] Yarbus, A. L., "Eye movements and vision," *Plenum*, (1967).
[5] Zhang, Y., Jiang, G., Yu, M., Chen, K., and Dai, Q., "Stereoscopic Visual Attention-Based Regional Bit Allocation Optimization for Multiview Video Coding," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, p. 848713, (2010).
[6] Chamaret, C., Boisson, G., and Chevance, C., "Video retargeting for stereoscopic content under 3D viewing constraints," in *SPIE 8288, Stereoscopic Displays and Applications XXIII*, pp. 82880H-82880H-12, (2012).
[7] Jansen, L., Onat, S., and König, P., "Influence of disparity on fixation and saccades in free viewing of natural scenes," *Journal of Vision*, vol. 9, (2009).
[8] Wismeijer, D. A., Erkelens, C. J., van Ee, R., and Wexler, M., "Depth cue combination in spontaneous eye movements," *Journal of Vision*, vol. 10, (2010).
[9] Czuni, L. and Kiss, P. J., "About the fixation points in stereo images," in *Cognitive Infocommunications (CogInfoCom), 2012 IEEE 3rd International Conference on*, pp. 143-147, (2012).
[10] Huynh-Thu, Q. and Schiatti, L., "Examination of 3D visual attention in stereoscopic video content," in *SPIE 7865, Human Vision and Electronic Imaging XVI*, pp. 78650J-78650J, (2011).
[11] Zhang, Y., Jiang, G., Yu, M., and Chen, K., "Stereoscopic Visual Attention Model for 3D Video," in *Advances in Multimedia Modeling*. vol. 5916, S. Boll, Q. Tian, L. Zhang, Z. Zhang, and Y.-P. Chen, Eds., ed: Springer Berlin Heidelberg, pp. 314-324, (2010).
[12] Wang, J., Perreira Da Silva, M., Le Callet, P., and Ricordel, V., "A computational model of stereoscopic 3D visual saliency," *Image Processing, IEEE Transactions on*, vol. PP, pp. 2151,2165, (2013).
[13] Gautier, J. and Le Meur, O., "A time-dependent saliency model combining center and depth biases for 2D and 3D viewing conditions," *Cognitive Computation*, vol. 4, pp. 141-156, 2012-06-01, (2012).
[14] Khaustova, D., Fournier, J., Wyckens, E., and Le Meur, O., "How visual attention is modified by disparities and textures changes?," in *SPIE 8651, Human Vision and Electronic Imaging XVIII*, pp. 865115-865115, (2013).
[15] Chen, W., Fournier, J., Barkowsky, M., and Le Callet, P., "New stereoscopic video shooting rule based on stereoscopic distortion parameters and comfortable viewing zone," in *SPIE 7863, Stereoscopic Displays and Applications XXII*, San Francisco, California, USA, (2011).
[16] Yano, S., Emoto, M., and Mitsuhashi, T., "Two factors in visual fatigue caused by stereoscopic HDTV images," *Displays*, vol. 25, pp. 141-150, (2004).
[17] Hoffman, D. M., Girshick, A. R., Akeley, K., and Banks, M. S., "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of Vision*, vol. 8, (2008).
[18] Chen, W., Fournier, J., Barkowsky, M., and Le Callet, P., "Quality of experience model for 3DTV," in *SPIE 8288, Stereoscopic Displays and Applications XXIII*, Burlingame, California, USA, pp. 82881P-9, (2012)
[19] Monteiro, Wfg5001, Robo3dguy, and Jay-Artist, "Original design of "Cartoon", "Hall", "Pigs", "Table". ," ed: http://www.blendswap.com, (2011-2013).
[20] Le Meur, O. and Baccino, T., "Methods for comparing scanpaths and saliency maps: strengths and weaknesses," *Behavior Research Methods*, pp. 1-16, (2012).
[21] Le Meur, O., "Fixation analysis software," in *http://people.irisa.fr/Olivier.Le_Meur/publi/2012_BRM/index2.html#soft*, ed. IRISA, Rennes 1, France, (2012)