



# A Method of Evaluating User Visual Attention to Moving Objects in Head Mounted Virtual Reality

Shi Huang<sup>(✉)</sup>

Animation and Digital Arts Academy, Communication University of China,  
Dingfuzhuang East Street, Chaoyang, Beijing, China  
bit.stone@163.com

**Abstract.** Virtual reality games/films/applications bring new challenges to conventional film grammar and design principles, due to more spatial freedom available to users in 6-DOF Head-Mounted Display (HMD). This paper introduces a simple model of viewers' visual attention in environment of virtual reality while watching randomly generated moving objects. The model is based on a dataset collected from 10 users in a 50-seconds-long virtual reality experience on HTC Vive. In this paper, we considered three factors as major parameters affecting audiences' attention: the distance between object and the viewer, the speed of objects movement, and the direction of object towards. We hope the research result is useful to immersive film directors and VR game designers in the future.

**Keywords:** Virtual reality · Focus of attention · Immersive film  
VR game · VR experience

## 1 Introduction

Some business and technology specialists predict that VR (Virtual Reality) will become mainstream by 2020, and that it will eventually be led by non-games applications [1]. They believe that VR technology holds tremendous potential to reshape the future for many industries, from video game, cinema, healthcare, education, arts to engineering. Many film festivals have started VR competitive section or showcase. These events include Cannes Film Festival, Venice Film Festival, Sundance, Tribeca, Toronto, Geneva Intl. Film Festival Tous Ecrans and Dubai fest [2].

However, some film directors regard VR as an immature, somehow dangerous medium for storytelling. Because VR “gives the viewer a lot of latitude not to take direction from the storytellers but make their own choices of where to look [3].” In simulated environment, users may miss the important clue or element when exploring other trivial areas. VR designers must find a way to guide the audiences to look at the right direction, otherwise they may fail to achieve a complete storytelling. How to grab users' visual attention, how to guide the audience's fixation, becomes a very basic and essential technique in VR experience design.

Human visual attention to graphics has been studied for many years. Koch and Ullman presented a computational model of human visual attention in 1987 [4], and

Clark and Ferrier developed a practical implementation in two years later [5]. Recent research includes Judd and Durand in 2012 [6], Borji and Itti in 2013 [7], Borji et al. in 2015 [8], many different computational saliency models have been built to detect salient visual subsets. These studies normally produce saliency map from images and videos by various features, such as brightness, color, shape and other characteristics.

In VR field, visual attention for spherical images/videos has been studied in [9–11]. Bogdanova et al. propose a method to obtain saliency maps from omnidirectional images. Upenik et al. describe a model for head direction trajectories to produce continuous fixation maps. However, these studies are based on spherical VR contents, which are usually shot with VR cameras or pre-rendered on computers. These images/videos compress depth information into a spherical projection, lacking of realistic sense of space.

Current popular VR HMD such as Oculus Rift, HTC Vive or PlayStation VR has 6 DoF (degree of freedom), allowing users to walk and jump within a given space. Many VR experiences are rendered in real-time, for instance, Oculus Story Studio’s “*Lost*” (2015) and “*Henry*” (2015), Disney’s “*Coco VR*” (2017) and Magnopus’ “*Blade Runner 2049: Memory Lab*” (2017). Real-time rendering CG brings the participants more flexibility and interactivity. In these cases, a very effective trick to grab viewer’s attention is to place a moving object in the scene. For example, there is a crawling ladybug in the beginning of “*Henry*”. The director initially placed various objects with different speeds for test, and finally decided to use a ladybug to direct viewer’s attention [12]. Related studies show that human eyes are more likely to focus on moving and near objects than the opposite [13, 14].

In this paper, we propose a simple method to record users’ visual attention data in 6-DoF VR environment. We build a virtual space filled with several moving objects. To exclude potential interference of the objects’ appearance, we intentionally designed them as white spheres of the same size. Each sphere has a unique speed, generated in a random position towards a random direction in every testing.

Considering users actually observing random scenes in the experiment, we record data of objects instead of users. This is the major difference between our approach and others. When experiment repeated sufficient times, a rough model of users’ visual attention to moving objects is concluded.

## 2 Experimental Settings

In this section we describe the setting of moving objects and virtual environment, the software implemented to record data, as well as the test material and conditions considered during the experiment processing.

### 2.1 Virtual Environment

Figure 1 shows the virtual environment designed for the experiment. Users are placed in a virtual space, where is empty and dark, existing nothing but an infinite floor lightened by a directional light from the top. The floor within 10 m is illuminated and others fade into the darkness.

In front of the viewer, 20 white spheres are initially generated within a 4 m wide, 2 m high and 4 m deep cubic range. After that, spheres fly away along random directions. Each sphere is in the same color and the same size, precisely, 10 cm in diameter. We apply different velocities to each sphere, from 0.2 m/s to 4 m/s. If a sphere moves out of the given range, it will disappear at the edge and regenerated on the opposite side. As shown in Fig. 2, every sphere moves in a fixed velocity along a random direction. This process will repeat for about 1 min until the experiment finishes.

In order to observe more neatly, all directions are set to be vertical or horizontal, that is to say, there are 6 possible directions in total: top to bottom, bottom to top, left to right, right to left, forward and backward. Depending on the performance of the computer, the loading time can last for 10 to 20 s, and the actual testing time can vary from 50 to 60 s.

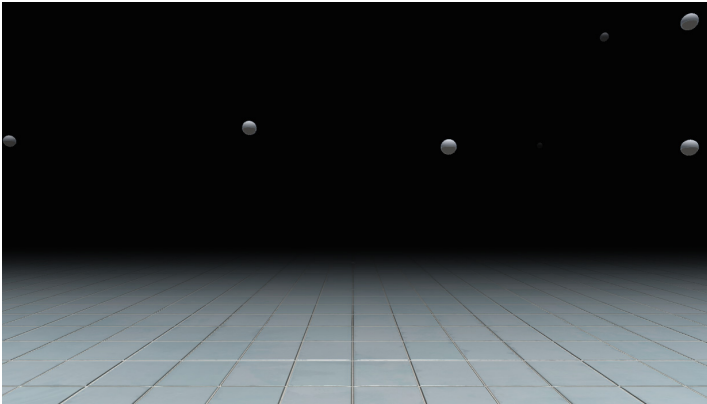


Fig. 1. Virtual environment in the experiment

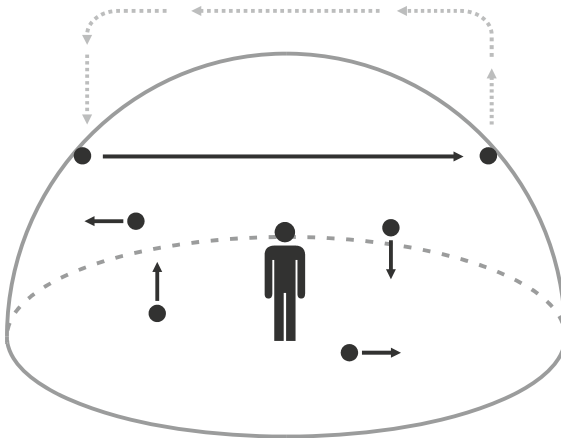
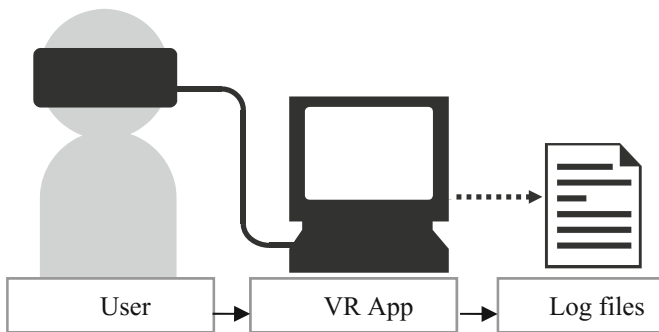


Fig. 2. Objects' movement in VR

## 2.2 Software

As shown in Fig. 3, we developed a VR application to capture and save to log files the objects' movement every 100 frames or whenever the viewer sees an object in the center of his/her visible range. The major purpose of the software is to associate each attention event to a timestamp and other related information. Once the object is observed by the user (i.e. an attention event), a corresponding log file will add a line, which includes the following data:

- The object's ID (integer)
- Timestamp of the event (float)
- Count of the object being observed (integer)
- Distance between object and the viewer (float)
- Screen position in camera vision (float)
- Whether the object is approaching or leaving (Boolean value)
- Current velocity of the object (float)
- Current direction of the object (3d vector, as described above, there are 6 directions in total).



**Fig. 3.** Workflow of the software

This VR application was developed with game engine Unity in C#, based on the SteamVR Plugin developed by Valve Corporation. This SDK allows developers to target a single interface that will work with all major virtual reality headsets; in this case, the headset was HTC Vive. We performed all tests and the data collection campaign on Windows 10.

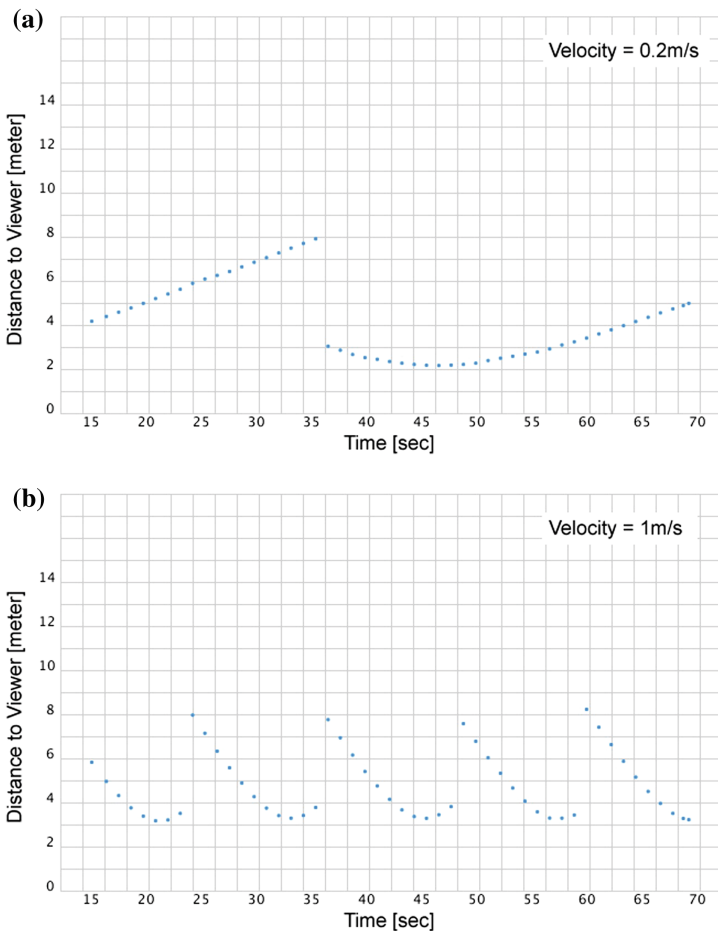
## 2.3 Testing Session

11 subjects participated in the experiment, 9 male and 2 female subjects, between 20 and 39 years old. Most subjects in the sample group are students or staffs from Communication University of China. Some are visitors from other institutes. All subjects have experiences of virtual reality. Before the beginning of the testing, oral instructions were provided to describe the main steps of the testing process. The exact flow is described as following:

1. Before the user puts the HMD on, he/she is asked if they are familiar with HMDs. The instructor records information including gender, age and level of familiarity with VR on HMDs.
2. Before testing starts, the instructor helps user to adjust HMD, such as the fine-tuning of the position and the interpupillary distance.
3. VR application starts and lasts 70 s.
4. At the end of the session, the user is informed that test is over and to remove the HMD.

### 3 Data Set

Our method respectively records the data of each moving object. Due to the velocity of a certain object is fixed every time, based on the log file name (containing velocity information), we can conveniently evaluate the data for each object.



**Fig. 4.** (a). Object distance to viewer (move at 0.2 m/s) (b). Object distance to viewer (move at 1 m/s) (Color figure online)

### 3.1 Distance Between Object and Viewer

Distance between a certain object and the viewer is a float, which is recorded every 100 frames (about 1.2 s). Object distances to the viewer are presented as blue dots, as shown in Fig. 4(a) and (b). Because objects move from one side to another repeatedly, the data in Fig. 4 shows a periodic pattern.

During the periodic movement, once the viewer captures the object in view, a data tagged by attention event will be added. These attention events are presented as red dot. And according to the object's proximity to users' visual center (in camera view), we set attention events into different sizes. Object closer to visual center, the size of related red dot is bigger. Figure 5 depicts a typical chart with attention events of an object moving at 0.4 m/s. We can put all experiments data of this object into one figure, as shown in Fig. 6.

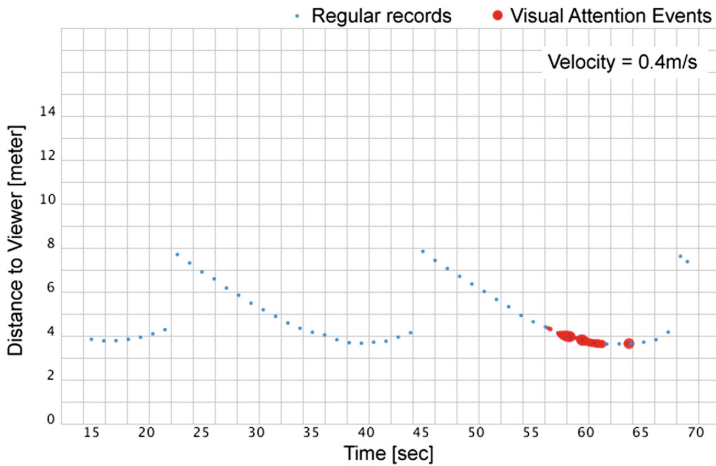


Fig. 5. Distance chart with attention event (Color figure online)

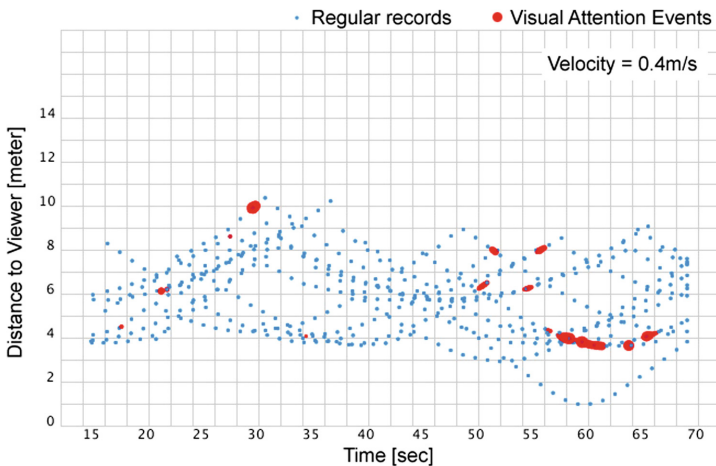


Fig. 6. Summary of all experiment data of an object (Color figure online)

### 3.2 Velocity

As stated before, each object in the virtual environment moves at a fixed speed, from 0.2 m/s, 0.4 m/s, 0.6 m/s... to 4 m/s. As shown in Fig. 7, each column contains the total attention events of an object (there are 20 objects along x-axis), and y-axis displays the distance between user and object at the moment when this event takes place. To identify potential patterns behind hundreds of dots, we set each dot as a translucent color and set its size based on its position in the camera screen. If a point is small, it means it may have been recorded accidentally.

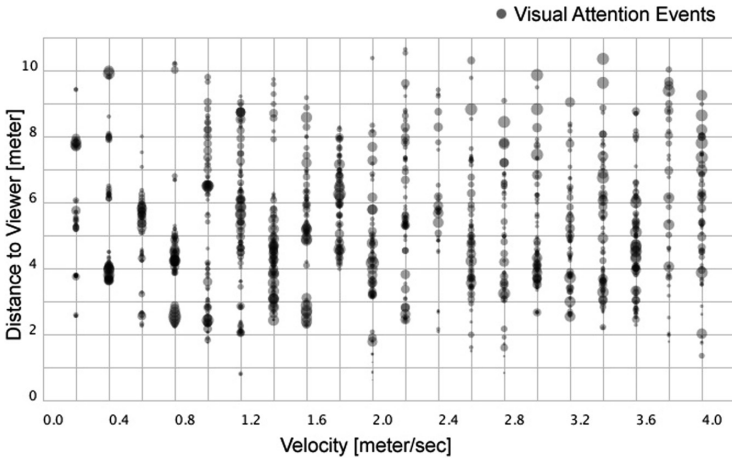


Fig. 7. Attention events distribution of velocity and distance

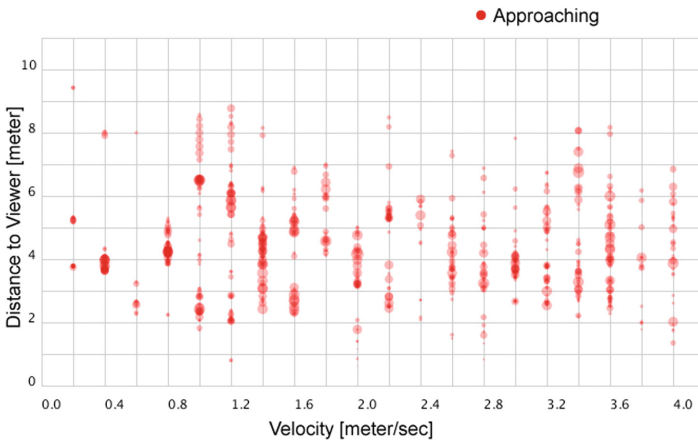


Fig. 8. Distribution of approaching events (Color figure online)

### 3.3 Approaching and Leaving States

Given that users' responding to approaching objects and leaving objects are different, it is necessary to calculate the motion states of all events to judge whether the object is approaching. Method as below: we save the distance data of target event as  $d[1]$ , and the distance data of previous second as  $d[0]$ . If  $d[1]$  greater than  $d[0]$ , it means the object is leaving the observer; Else if  $d[1]$  less than  $d[0]$ , the object is approaching. As shown in Figs. 8 and 9, we colored approaching and leaving events with different colors, the former with red and the latter with blue. Figure 10 depicts the distribution of these two states, the approaching events appear as circles, and the leaving events appear as rectangles. We find that the distances of these states have significant differences.

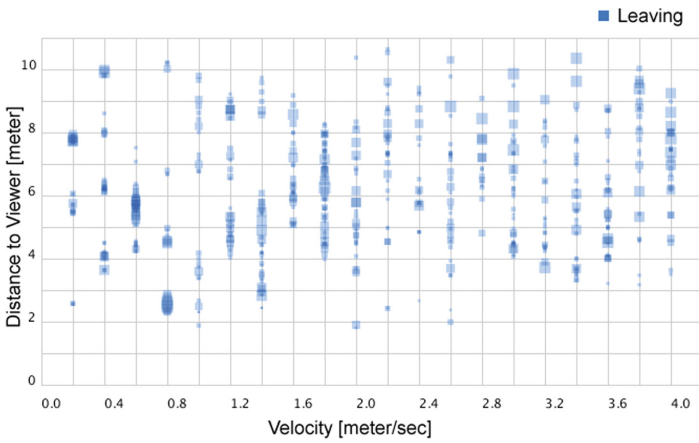


Fig. 9. Distribution of leaving events (Color figure online)

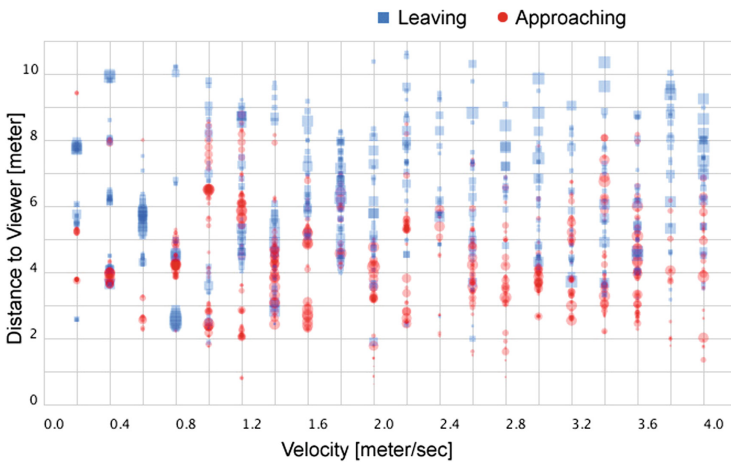


Fig. 10. Colored distribution of attention events (Color figure online)



Distances of leaving events are relatively longer. That means users often gaze at one certain moving object until it disappears far away; but they no longer look at approaching object when it passes by.

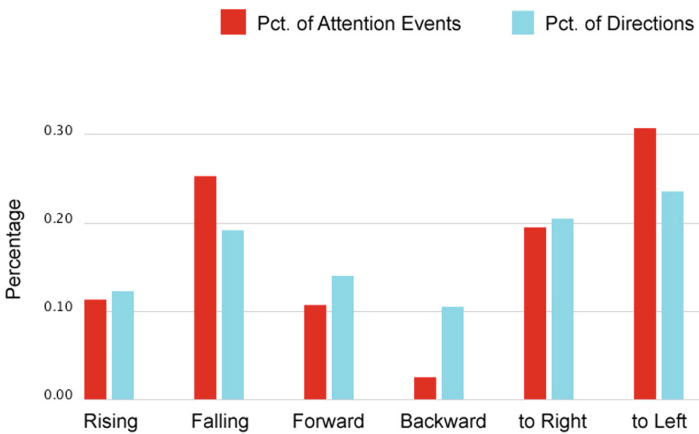
### 3.4 Moving Directions

We classify objects movement into 6 types according to their move directions. Vertical movement has 2 types: rising and falling. Horizontal movement has 2 types: from left to right and from right to left. And other two types are forward and backward. In our dataset, these states are saved as magnitudes of 3d vector. Table 1 shows the statistics of directions and related attention events. Note that the probabilities of each direction shows in experiments are not evenly distributed, which partly explains the reason why some directions get more attentions.

**Table 1.** Attention events of moving directions

No.	Direction	Magnitude	Attn. events number	Events Pct.	Direction Pct.
1	Rising	(0, 1, 0)	15718	11.30%	12.27%
2	Falling	(0, -1, 0)	35204	25.31%	19.18%
3	Forward	(0, 0, 1)	14860	10.68%	14.00%
4	Backward	(0, 0, -1)	3444	2.48%	10.49%
5	To right	(1, 0, 0)	27106	19.48%	20.49%
6	To left	(-1, 0, 0)	42786	30.76%	23.58%

As shown in Fig. 11, the probability of directions has positive correlation with attention events. And Fig. 12 shows the result of eliminating the influence of probability factor.



**Fig. 11.** Percentage of attention events and moving directions

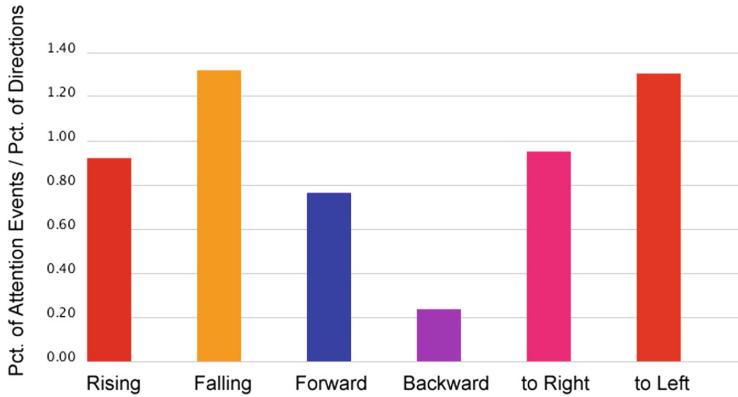


Fig. 12. Users' preference for moving direction

## 4 Conclusion

In this paper we have described a simple method to record and evaluate users' visual attention to moving objects in 6-Dof virtual reality environment. The method incorporates data analysis of randomly generated objects, which includes moving speed, moving direction, distance to viewer, attention events and timestamps. We imply that in a space filled with multiple moving objects, the speed of the objects has no significant effect on the users' visual attention. In our experiments, users spend less time to object moving along particular directions, such as forward and backward; and users normally watch leaving objects in further distance than approaching objects.

Future work will focus on improving the design of VR environment. More specifically, moving objects will be reduced to 1 or 2, while most objects will remain static. This state is more similar to the actual method of attracting users' attention in VR experiences (for instance, VR film and VR game) and may have more reference value.

## References

1. Batchelor, J.: Riccitiello: VR will be mainstream by 2020 (2016). <http://www.gamesindustry.biz/articles/2016-12-21-riccitiello-vr-will-be-mainstream-by-2020>
2. Vivarelli, N.: Virtual Reality Films Take Hold at Fall's Film Festivals (2017). <http://variety.com/2017/film/spotlight/carne-y-arena-venice-virtual-reality-cannes-1202524881/>
3. Matyszczuk, C.: Steven Spielberg says VR is a 'dangerous' movie medium (2016). <https://www.cnet.com/news/steven-spielberg-says-vr-is-a-dangerous-movie-medium/>
4. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. In: Vaina, L.M. (ed.) *Matters of Intelligence*. SYLI, vol. 188, pp. 115–141. Springer, Dordrecht (1987). [https://doi.org/10.1007/978-94-009-3833-5\\_5](https://doi.org/10.1007/978-94-009-3833-5_5)
5. Clark, J.J., Ferrier, N.J.: Modal control of an attentive vision system. In: *IEEE International Conference on Computer Vision* (1988)
6. Judd, T., Durand, F., Torralba, A.: A benchmark of computational models of saliency to predict human fixations. MIT Technical report (2012)

7. Borji, A., Itti, L.: State-of-the-art in visual attention modeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(1), 185–207 (2013)
8. Borji, A., Cheng, M.M., Jiang, H., Li, J.: Salient object detection: a benchmark. *IEEE Trans. Image Process.* **24**(12), 5706–5722 (2015)
9. Bogdanova, I., Bur, A., Hügli, H.: Visual attention on the sphere. *IEEE Trans. Image Process.* **17**(11), 2000–2014 (2008)
10. Bogdanova, I., Bur, A., Hügli, H., Farine, P.-A.: Dynamic visual attention on the sphere. *Comput. Vis. Image Underst.* **114**(1), 100–110 (2010)
11. Upenik, E., Ebrahimi, T.: A simple method to obtain visual attention data in head mounted virtual reality. In: *IEEE International Conference on Multimedia and Expo 2017, Hong-Kong* (2017)
12. Thill, S.: Oculus Creative Director Saschka Unseld: “It Feels Like We’re in Film School Again” (2015). <http://www.cartoonbrew.com/interactive/oculus-creative-director-saschka-unseld-116773.html>
13. Sheng, H., Liu, X.Y., Zhang, S.: Saliency analysis based on depth contrast increased. In: *Proceedings of 2016 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1347–1351. IEEE, Shanghai (2016). <https://doi.org/10.1109/icassp.2016.7471896>
14. Park, J., Oh, H., Lee, S., et al.: 3D visual discomfort predictor: analysis of disparity and neural activity statistics. *IEEE Trans. Image Process.* **24**(3), 1101–1114 (2015). <https://doi.org/10.1109/TIP.2014.2383327>