Under the Movement of Head:

Evaluating Visual Attention in Immersive Virtual Reality Environment

1Honglei Han

1 School of Animation and Digital Arts, Communication University of China, Beijing, China

Department of Computer Science, UNC Charlotte, Charlotte, NC 28223, USA

hanhonglei@cuc.edu.cn

2Aidong Lu 3Unique Wells

2,3 Department of Computer Science, University of North Carolina at Charlotte, Charlotte, NC 28223, USA,

aidong.lu, uwells@uncc.edu

**Abstract-**This paper presents a method to measure what and how deep the user can perceive when exploring virtual reality environments using a head mounted display. A preliminary user study was conducted to verify that user gaze behavior has specific differences in immersive virtual reality environments compared with that in conventional, non-immersive virtual reality environments, which are based on a desktop screen. Gathered from the study result for gaze behavior, the users experiencing immersive virtual reality environments are more likely to adjust their head movement to center interesting objects in their vision. Based on this finding, a quantitative method is proposed to measure the user's visual attention in such a virtual reality environment. In application part, a user personalized storyboard has been designed to capture the user's most regarded views as key frames that can depict users' exploration experience in immersive virtual reality environments.

测试是否能够修改看到。

**Keywords**: gaze analysis; visual attention; virtual reality

Introduction

Interactive virtual environments are widely used in computer games, virtual reality (VR), computer aided design, and other computer graphics-related fields. With the evolution of modern three dimensional rendering engines and Head Mounted Displays (HMDs) users can vividly experience immersion, which has been pursued in the computer graphics field for decades. Only HMD-based VR applications are discussed in this paper as they are more popular than others, such as Cave Automatic Virtual Environments (CAVEs).

测试

In immersive VR, users can explore using natural interaction methods, which are much easier to utilize excessively. If there are insufficient view-guiding methods, users may not always be able to explore such environments following the purposes of designers, as the design ideas cannot be easily expressed. Therefore, it is vital for VR film directors to be able to evaluate the user's attention. This assists the directors with understanding their users’ experiences, and then helps give ideas on how to improve them, to make sure the user focuses mostly on significant objects — not straying too much to observe background objects.

We propose a hypothesis about users having similar patterns in exploration with immersive VR environments. A preliminary user study experiment has been established to verify it. We sought to find any particular user gaze patterns in immersive VR, and used non-immersive VR environment patterns as a comparison scale.

The analysis results showed that, in immersive VR, viewers instinctively move their heads to gaze at objects they are most interested in. They adjust their own lines of vision to bring the most interesting object into their center of view. Conversely, they would rather scan using only their eye movement, redirecting the avatar's field of view when necessary in non-immersive VR.

From this finding, we derived a method uses the Logistic function to measure the visual attention degree of objects by users during exploring. It is used by applications to capture the most impressive shots when subjects are exploring the immersive VR scenes. A user may examine this so-called personality storyboard's summarization to enrich his or her own experience with using VR. This storyboard may also be used as a visualization tool to help designers evaluate their immersive VR scenes when explored by users.

# Related work

The visual attention quantitative methods can be divided into two main categories. One of them is the subjective method, which measures the reactions of users and then uses them to analyze the relationship between contents and user reactions. Another one is the objective method, which calculates features in the contents and indicates users' vision attention behaviors. These two categories always combine with each other [[1](#_ENREF_1)].

The eye tracking system being able to directly record users' eye movements is a typical subjective method for estimating visual attention. The eye movement can then be used as a strong evidence to infer what is perceived by the user [[2](#_ENREF_2)]. There are some pioneer studies that embed eye trackers into VR HMDs [[3](#_ENREF_3)]. However, raw data for eye mobility contains eye direction positions on-screen, which then has to be translated into contents within the virtual scene. Many eye movements are saccades, which are both voluntary and reflexive. There is virtually no visual information cognitively processed during a saccade [[2](#_ENREF_2), [4](#_ENREF_4)].

Other subjective methods always use software, instead of eye tracking systems, to collect user behavior to deduce user attention. Camera control is a strong hint for estimating user visual attention behavior in interactive environments. For a survey on camera control in interactive virtual environment the reader is referred to [[5](#_ENREF_5)].

Objective methods, on the other hand, can predict whether or not the user would pay attention to some objects or areas. The predication result has to be evaluated by eye tracking, or other subjective methods [[1](#_ENREF_1), [6](#_ENREF_6), [7](#_ENREF_7)]. It can be further divided into two groups: bottom-up and top-down. Bottom-up methods (stimulus-driven) tend to obtain gaze patterns from the basic features in the scene, such as color, intensity, and motion. A saliency map can be calculated through these low-level factors, and may be used as a quantitative value to indicate users' attention [[8](#_ENREF_8)]. Top-down methods (goal-directed) relate with high level processes, such as thinking, inference, and memory. They consist in simulating the cognitive processes that take place in the brain[[7](#_ENREF_7), [9](#_ENREF_9)].

Very recently, Vincent *et al.* addressed the problem of how users explore virtual environments and what constitutes saliency in immersive applications [[10](#_ENREF_10)]. After analyzing user study data, they confirmed the hypothesis that saliency in immersive VR is in good agreement with saliency in conventional displays. The same result with ours is derived: head movement can be a valuable tool to analyze the approximate regions that users attend to in a scene without the need for additional eye-tracking hardware. Contrasting with them, the immersive VR videos, which are more popular in modern VR applications, are the main study objects in our paper instead of still panoramas used in [[10](#_ENREF_10)]. A virtual object, which is the basic unit in the VR designing process, based visual attention is obtained in our paper rather than conventional region based saliency map obtained in [[10](#_ENREF_10)].

Our method in this paper can be categorized into the subjective method, since it relies on users' head motion as a strong hint of gaze and attention. The head orientation has been revealed as a main criterion of gazing in immersive VR in our user study experiment. The top-down idea in objective methods inspired us greatly in designing the user study scenes, especially concerning the importance of task oriented gaze guiding [[11](#_ENREF_11), [12](#_ENREF_12)].

Many practical applications have taken advantage of results of users' attention behavior researches, such as user experience analysis [[13](#_ENREF_13), [14](#_ENREF_14)], natural interaction design [[15](#_ENREF_15)], user interface design [[16](#_ENREF_16)], automatic recognition [[17](#_ENREF_17)], and scene alignment [[18](#_ENREF_18)].

User visual attention research has been involved in interactive virtual environment applications[[1](#_ENREF_1), [10](#_ENREF_10), [19](#_ENREF_19)]. The main purpose is to help designers improve their work and enhance user experience. In papers [[20](#_ENREF_20)] and [[21](#_ENREF_21)], the authors established the relationship between user experience and user gaze data through virtual cameras. As they believed that user visual attention is very important to user experience in interactive virtual environments, they attempted to use the attention model to ensure that the important objects are more likely to appear in the center of the screen.

In the immersive video field, Thomas *et al.* presented a visualization framework for analyzing head movements and gaze data for immersive 360°videos in the paper [[19](#_ENREF_19)]. A specialized view similarity visualization method was provided, that allows analysts to quickly identify moments of spatiotemporal agreements between the viewing directions of individual participants. It can further be used as an evaluation tool to detect whether the attention guidance of an immersive video works as expected. However, viewing directions used in that paper cannot efficiently express what particular semantic virtual objects are focused on by the participants.

We developed a so-called personalized storyboard generating system. It uses virtual object based attention values as a metric to output key frames during participants' exploring process in immersive VR environments. There have been many other user behavior abstraction systems proposed in recent years, like video summarization[[22](#_ENREF_22)]. Such systems mainly utilized the objective method to automatically extract the most important shots in contents and organize them into scripture boards[[22](#_ENREF_22)]. However, this proposed personalized storyboard system can extract the key frames that contain the most concerned objects, based on the user's own exploration behavior. It can not only improve the user's experience as a complimentary result, but it also helps designers to qualify their work in an intuitive way.

# Immersive VR exploring features

As discussed in paper[[23](#_ENREF_23)], in an immersive VR scene a participant will respond to a virtual reality as if it were real, as a function of place illusion (PI) and plausibility illusion (Psi). If you are there (PI) and what appears to be happening is really happening (Psi), then this is happening to you. Hence, you are likely to respond as if it were real[[23](#_ENREF_23)]. In non-immersive VR scenes, however, the users have to perceive information through the two-dimensional monitor and use unnatural interactive hardware to mingle with the virtual environment, such as a keyboard, mouse, and joystick. It is harder to invoke the feeling of presence, and as a result users tend to treat the virtual avatar as the third person, controlling it to explore the virtual world, instead of placing themselves in the virtual avatar's shoes.

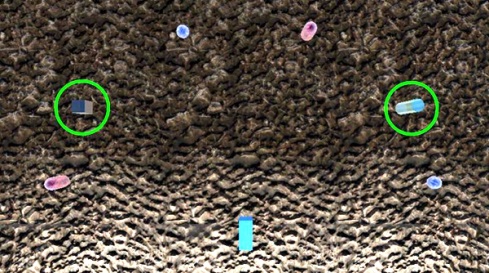
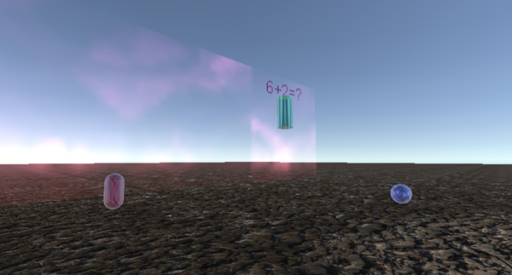
Therefore, we propose a hypothesis: when exploring in immersive VR environments, users tend to move their heads more often to perceive objects they are interested in just as they do in real life. However, in mouse and keypad controlled non-immersive VR environments, users feel it is inconvenient to move the avatars' heads, choosing to move their own eyes to gaze at objects that appear on the screen.

According to a recent state-of-art paper [[24](#_ENREF_24)], quantitative evaluation is considered the cornerstone of the scientific discovery process. The evaluation can always get verified through empirical experimentation, which is treated as evidence supporting the value of a newly introduced technique, or a contrast of the effectiveness of several established methods. These goals are pursued by performing statistical analysis, driven by a set of hypotheses aimed at answering high-level research questions.

So we decided to use behavior analysis methods to verify our hypothesis of user exploring pattern in immersive VR. It is based on recorded participants' particular interactions that better reflect the real-time exploration interests, instead of questionnaires used in many other user interaction evaluation researches, such as the recently published paper [[25](#_ENREF_25)]. As inattention blindness exists, participants most likely will not remember everything that they just observed[[26](#_ENREF_26)]. Furthermore, many errors might be introduced in using a post-process questionnaire system.

# User study scenes

There are three requirements for these user study scenes. First, they must be sophisticatedly designed to ensure that every participant can observe in a similar way, so that the visual attention pattern may be deducted. Objects in user study scenes are divided into task-related, math problems solving, and background to achieve this goal. Since it has been proved that the task-related gaze behavior can dominate over saliency[[12](#_ENREF_12)]. Second, the task-related object is supposed to be focused on or gazed at as much as possible during its perceiving stage; meanwhile, the background objects should be disregarded. We set each task-related object to rotate throughout the scenes, and used materials similar to the background to make it slightly troublesome to properly see the math problems if participants do not focus on them for a while. Finally, the scenes must avoid the object's positional configuration influence to the gazing behavior. In order to diagnose the exact relationship between task-related objects and the users' viewing direction, we let them keep translating in different trajectories through particular scenes.



(a) From user's view (b) Top view

Figure 1. Shots captured during the user study scene (a), and its top view (b). Green circled objects are task-related

There are five user study scenes developed with Unity 3D game engine. Each of them contains one or two objects with three addition problems, appearing sequentially, between two numbers less than ten for participants to solve (as seen from Figure 1(a)). They all have the same scene arrangement, as shown in Figure 1(b)}. Objects circled in green may have math problems on top of them, and a camera is located at the bottom middle (see the blue box as the representation in Figure 1(b)). Each task-related object in the scenes is rotating. The differences are the number of task-related objects and their moving trajectories. Scene 1 *static* contains only one task-related object, while scene 2 *static* contains two. In order to find out if participants' view direction is strongly correlated to task-related objects, we set the objects to be active in the last three scenes. In 1*move* scene, there is only one move task-related object moving horizontally, back and forth. The 2*moveS* scene contains two in up-and-down and left-and-right trajectories, respectively. The final scene 2*move* contains two moving task-related objects in triangular and rectangular trajectories, which are more complex compared to previous scenes. All five of these scenes contain several background objects that are motionless. They are used to compare attention differences with task-related objects. We hypothesized that participants in immersive VR scenes tend to orientate their heads towards the task-related objects, while simply scanning with their own eyeballs with a stable virtual camera in non-immersive scenes.

## User study configuration

We want to concentrate on the gaze behavior because that reflects the visual attention of participants, so then the translation degree of the virtual camera is eliminated. In immersive VR scenes, participants use a VR HMD to explore. The latest Oculus Rift HMD is used in our experiments. In non-immersive mode, the user study scenes run on a common PC, with a 24.1*''* monitor in a resolution of 1920x1200px, and participants sat around 0.45 meters away using a mouse to explore through first person view.

The level time is twenty seconds in each user study scene, followed by an answer user interface stage. Participants are required to input their answers to the math problems in the scenes. Such a time limit ensures that participants concentrate on tracing and focusing on the task-related objects. If the participants solve all the problems early, they have the option of using the “Space” key to skip to the answer stage directly. The correct rate of the answers can be used to validate the user study result.

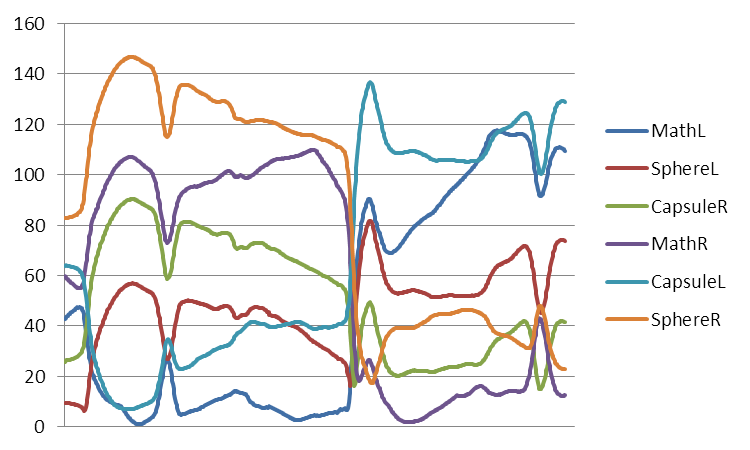
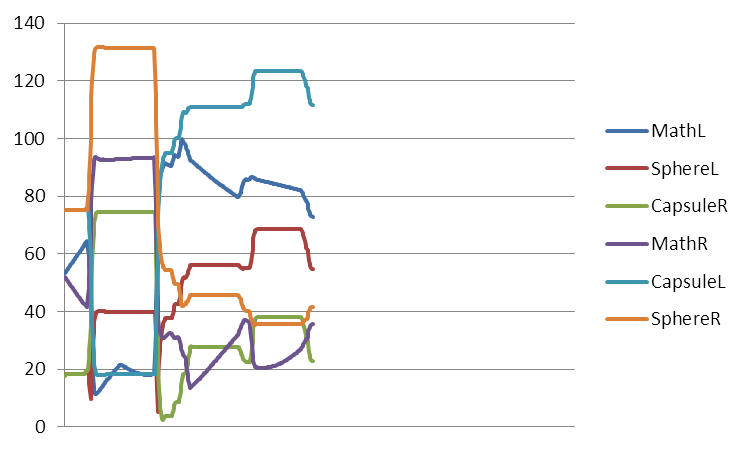
The gaze behavior we are mostly concerned with is the included angle between view direction and the vector from the camera to the object. It indicates how centered the object is in the camera. We believe the included angle is a strong hint of the objects' visual attention degree from the participant.

Every participant is required to explore all scenes using both immersive and non-immersive modes. To avoid the influence of the experience order, half of the participants will experience non-immersive mode first, while the other half will experience immersive mode.

## Result analysis

We invited twenty-five participants, who are all students and faculty members in our university, aging from 19 to 48 with fifteen males and ten females. Their backgrounds are diversified, and all are familiar with how to navigate in desktop VR applications. Only five of them have experienced HMD-based VR scenes at least once. After fully understanding the purpose and control method of the user study application, they were asked to play the scenes in both of the VR modes.

Just as the hypothesis stated in the beginning, most participants acted differently in immersive mode with what in non-immersive mode. In immersive mode, they tend to focus on the task-related objects exclusively when they were trying to solve the math problems upon the objects. In non-immersive mode, however, they tend to only keep the task-related objects inside the screen and use their own eyes to trace the moving objects instead of trying to rotate the virtual camera to trace them. A probable reason is that in immersive environments people act like they are in real world, in which they tend to keep the goal right in the center of their vision until reaching a limitation, and then use eye movement as a complement. In non-immersive mode, using a mouse to control the virtual camera would trigger a series of computer processes to refresh. It would at least somewhat distract users' attention while they are solving the math problems. People would rather leave the camera still and use their own eyes to trace the moving task-related objects.



(a) non-immersive mode (b) immersive mode

Figure 2. Included angles of a participant with different objects in scene 2*moveS* use non-immersive (a) and immersive (b) methods respectively. *X* axis is time, and *Y* axis is included angles

The gazing pattern of these two exploring methods and their differences can be viewed in Figure 2. It depicts a participant's exploration process in the user study scene 2*moveS*. We can see clearly that in non-immersive mode (Figure 2(a)), this participant first began to find a task-related object, *MathL*, and then kept the virtual camera still as the included angles of unmoving background objects stayed the same. After getting the results of the math problems on *MathL*, the participant began to find another task-related object, *MathR*. He then began to solve the math problems upon it; the virtual camera was nearly kept still until the end of the level. He only moved the camera when *MathR* moved to the edge of the screen. In immersive mode (Figure 2(b)) the curves are not so regular, as the participant tended to keep his head rotating and tried to focus on the task-related object when he was solving the math problems upon it. The included angle as he was focusing fell below 10°. It's worthy to point out that the participant earned a 100% answer correct rate in this scene using both modes, and he passed the level ahead of time in non-immersive mode.

# Object-based attention quantitative equation

We can conclude from section IV-B that objects receiving the most user attention continuously appear in the center of the screen in immersive VR mode. In the non-immersive mode, users have more freedom to use their own eyes, which is more convenient for gazing the objects on the monitor instead of using a mouse to control the virtual camera to focus them.

In order to reflect the gaze pattern in immersive VR scenes, a math model is needed to give a higher value to the objects that have included angles smaller than a certain threshold of degrees, which means they are being focused on; while giving much lower values to the other objects that have larger included angles, which means the items are being mostly disregarded. Considering the influence of objects' sizes and human habits, the threshold angle should have some tolerance. We found out that the Logistic function fairly meets our requirements, as it emphasizes the center effect. Hence, the object-based visual attention value *A* can be calculated with Equation 1, a revised Logistic function,

where *k* denotes the steepness of the curve; here we use 0.5. *deg* denotes the included angle degree between view direction and the vector pointing from the camera to the object *O*. *d* denotes the sigmoid curve's midpoint, which means that for the threshold of gazing included angles, where we use 15°as the default value. As summarized from our user study results introduced in Section IV-B, included angles below 10°can be seen as focusing in immersive VR scenes. We add 5°as a tolerance. The attention value is accumulated along with time *t* to prove the statement that the more that subjects have interest in an object, the longer they will focus on it[[18](#_ENREF_18)].

From Figure 3, the apparent correlation between saliency levels of objects and their attention degrees calculated by Equation 1 are easily seen. The data used in the chart are from the same source in Figure 2. In immersive VR scenes, the task-related objects received the highest attention values. This means that the participant always put them in the center of his or her view. While in non-immersive VR mode, the participant kept the view unmoving so much that background object *CapsuleR* was put in the center of his or her view unintentionally, resulting in its highest attention value.

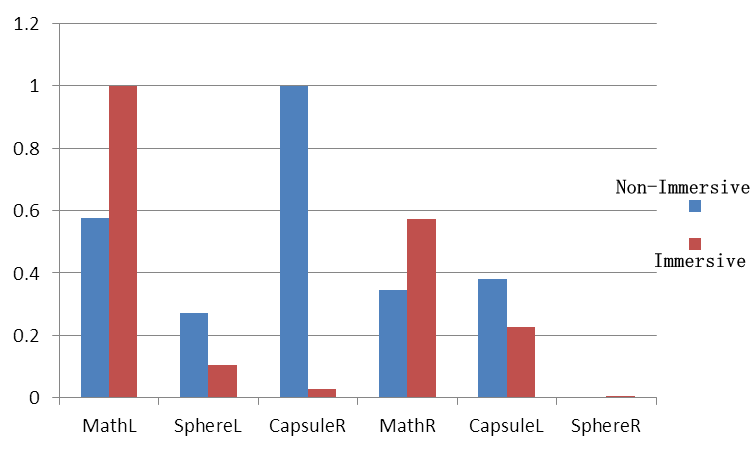


Figure 3. A user's normalized attention degree calculated by Equation1 in scene 2*moveS* using non-immersive and immersive modes respectively

# Applications

Using the object-based visual attention evaluation method proposed in Section V, designers will be able to use an intuitive technique to look into users' exploration patterns of their work. The series of quantitative charts or values using includes angle metric and object-based attention equation that can be generated through the users' exploration of the immersive VR scenes. After collecting enough user data, designers get a brief view of user experience in object attention that can help them predict whether or not their scenes are well explored, followed by their design purposes.

We developed an automatic storyboard generation system. Equation 1 is used as a criterion to decide whether or not a shot may be seen as a key frame; this relies on the fact that if an object is centered on by a person, it should be perceived as the point of interest to him or her. From the user study results introduced in Section IV-B, three seconds is a proper fixed time to decide if an object is really interested in by the user.

In order to test the efficiency, we had set up a immersive VR short film using resources from the cutting-edge demo Adam, shared by Unity. There is a realistic indoor scene that lasts for 100 seconds, where four animated characters will appear in the scene consecutively.

A script was assigned to the particular parts of each character, and some background objects, in order to record their levels of attention. There is an accumulator parameter in the script to collect the attention value calculated by Equation 1 in each fixed frame. Once the accumulator exceeds the value 3, the current screen will be captured as a key frame.

As can be seen from Figure 4, six screen shots were captured by our personalized storyboard system. These key frames roughly illustrate this participant's exploring process. In the very beginning, there was only a guard character in the scene. The participant had no choice but to stare at him for a while. The participant noticed that the character Adam came out of the door with a very strange appearance, so Adam attracted the participant's attention (shown in the second and third frames). Finally, Sebastian walked in. His design seems like a boss character, thus drawing much more attention from the participant. The personalized story board generated by our system in Figure 4 can also help us indicate some design issues. Character Lu seems missed by the participant, as he didn't get enough attention from it.

# Conclusion

A quantitative object-based attention evaluation method was proposed in this paper. Based on the results of the user study to diagnose the gaze pattern for task-related objects in VR scenes, we found that head movement is vital, especially in immersive VR environments. In accordance with this finding, we borrowed a Logistic function to evaluate the visual attention degree, with included angles between view directions and the vectors from the camera to the object.

We used the equation in an application to generate personalized storyboards which can obtain a brief exploration process for users. It has particular meanings, not only for the users, but also for designers' review purposes.

Since only head movement is taken into account to evaluate the visual attention in VR environments, the eye tracking system is not necessary. Thus, our method can be easily embedded into current immersive VR applications without adding complex calculations or equipment costs.

More results, the user study application source project introduced in Section IV, and the storyboard generation system demo used in Section VI can be found in author's website http://hanhonglei.github.io/.

# Acknowledgment

The support provided by China Scholarship Council and UNC Charlotte during a visit of Honglei Han to UNC Charlotte is acknowledged.

This work is supported by the National Key Technology R&D Program under Grant No. 2012BAH62F03 and the National Science Foundation under Grant No. 1629913.

# References

1. R. J. Peters and L. Itti. Applying computational tools to predict gaze direction in interactive visual environments. 2008
2. A. T. Duchowski. Eye tracking methodology. 2007
3. FOVE. 2017
4. R. J. Snowden, P. Thompson and T. Troscianko. Basic vision : an introduction to visual perception. 2012
5. M. Christie and P. Olivier. Camera control in computer graphics: models, techniques and applications. 2009
6. M. Bernhard, E. Stavrakis and M. Wimmer. An empirical pipeline to derive gaze prediction heuristics for 3D action games. 2010
7. S. Hillaire, A. Lécuyer, T. Regiacorte, R. Cozot, J. Royan and G. Breton. Design and application of real-time visual attention model for the exploration of 3D virtual environments. 2012
8. L. Itti. Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. 2005
9. S. Lee, G. J. Kim and S. Choi. Real-time tracking of visually attended objects in virtual environments and its application to LOD. 2009
10. V. Sitzmann, A. Serrano, A. Pavel, M. Agrawala, D. Gutierrez and G. Wetzstein. Saliency in VR: How do people explore virtual environments? 2016
11. V. Sundstedt, M. Bernhard, E. Stavrakis, E. Reinhard and M. Wimmer. Visual Attention and Gaze Behavior in Games: An Object-Based Approach. 2013
12. M. Land, N. Mennie and J. Rusted. The roles of vision and eye movements in the control of activities of daily living. 1999
13. G. N. Yannakakis, H. P. Martínez and A. Jhala. Towards affective camera control in games. 2010
14. P. Burelli. Virtual cinematography in games: Investigating the impact on player experience. 2013
15. P. Majaranta and A. Bulling. Eye Tracking and Eye-Based Human–Computer Interaction. 2014
16. T. Nakayama, H. Kato and Y. Yamane. Discovering the gap between Web site designers' expectations and users' behavior. 2000
17. K. Yun, Y. Peng, D. Samaras, G. J. Zelinsky and T. L. Berg. Studying Relationships between Human Gaze, Description, and Computer Vision. 2013
18. F. Alt, A. S. Shirazi, A. Schmidt, J. Mennen and #246. Increasing the user's attention on the web: using implicit interaction based on gaze behavior to tailor content. 2012
19. T. Löwe, M. Stengel, E. C. Förster, S. Grogorick and M. Magnor. Visualization and Analysis of Head Movement and Gaze Data for Immersive Video in Head-mounted Displays. 2015
20. A. Picardi, P. Burelli and G. N. Yannakakis. Modelling virtual camera behaviour through player gaze. 2011
21. P. Burelli and G. N. Yannakakis. Towards Adaptive Virtual Camera Control in Computer Games. 2011
22. Y. F. Ma, X. S. Hua, L. Lu and H. J. Zhan. A generic framework of user attention model and its application in video summarization. 2005
23. M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. 2009
24. C. Papadopoulos, I. Gutenko and A. E. Kaufman. VEEVVIE: Visual Explorer for Empirical Visualization, VR and Interaction Experiments. 2015
25. A. Steed, S. Friston, M. Lopez, J. Drummond, Y. Pan and D. Swapp. An 'In the Wild' Experiment on Presence and Embodiment using Consumer Virtual Reality Equipment. 2016
26. A. Mack and I. Rock. Inattentional Blindness. 1998