

Object-based Visual Attention Quantification using Head Orientation in VR Applications

Honglei Han^{a,b,*}, Aidong Lu^b, Chanchan Xu^a, and Unique Wells^b

^a*School of Animation and Digital Arts, Communication University of China, Beijing, 100024, China*

^b*Department of Computer Science, University of North Carolina at Charlotte, Charlotte, 28223, USA*

Abstract

This paper presents a method to measure what and how deep users 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 results 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. A user personalized storyboard is designed to capture the user's most regarded views as key frames that can depict the users' exploration experience in immersive virtual reality environments.

Keywords: gaze analysis; visual attention; virtual reality; interaction; eye tracking

(Submitted on October 22, 2018; Revised on November 21, 2018; Accepted on December 23, 2018)

© 2019 Totem Publisher, Inc. All rights reserved.

1. Introduction

Interactive virtual environments are widely used in computer games, virtual reality (VR), computer aided design, and other computer graphics related fields. Users operate many different kinds of methods to control avatars in virtual environments to navigate in them, and they experience them through rendered images that are actually the views from a virtual camera. With the evolution of modern three-dimensional rendering engines and Head Mounted Displays (HMDs), users can vividly experience immersion, which has been something 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).

However, because there are various interaction freedoms in virtual environments, users may not always be able to explore such environments following the purposes of designers, as the design ideas cannot be easily expressed. It is special in immersive VR that users can explore using natural interaction methods, which are much easier to utilize excessively. However, if there are insufficient view-guiding methods, users may get confused, or even lost, in virtual environments. Immersive VR users are not always experts; they usually experience the content once or twice. They typically would not know what exactly to pay attention to in a new immersive VR film or game, if specific hints are not properly provided. They may focus too much on some background objects and miss the key points of a story, leading to them missing out on the full story and ruining the experience.

Based on the interaction frequency and purpose, virtual environment applications can be divided into two main streams. One is interactive applications, such as digital games that contain massive interactions. The second stream is interactive films, such as virtual reality films and 360° panorama videos, in which interactions play a secondary role. Experienced designers can implement a virtual interactive scene that uses some design principles, such as placing particular salient objects, to guide users' views and allow them to experience the scene under the framework. It is easier to achieve this goal in interactive applications as there are specific tasks for each scene. Tasks can attract users' attention, and task-related contents

* Corresponding author.

E-mail address: hanhonglei@cuc.edu.cn

may serve as strong hints to guide users' exploration behavior. In immersive VR films, however, it is easy to get lost since there is usually a lack of specific tasks. The immersive VR short film named "Lost"¹ produced by Story Studio is a successful example of achieving the goal of user view guiding. It uses a firefly in a dark forest as a hint to guide the audience to the main stage gradually. From then on, a series of prominent objects guides the audiences' view to ensure they follow the narrative of the film.

Therefore, confirming whether particular objects that designers carefully set up in the virtual scene have received enough attention from users is eventually important, as narratives can only be expressed out through the audience's perspective. If a significant object is not noticed by the user, then the object would not have any influence on his or her experience. It is vital for VR film directors to be able to evaluate the user's attention, making sure he or she focuses mostly on significant objects and does not stray too much to observe background objects.

We propose a hypothesis about users having similar patterns in exploration with immersive VR environments. In order to verify this hypothesis, in this paper we established a preliminary user study experiment. 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. When there is no obvious significant object, users scan the environment in search of one. 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 immersive VR. This storyboard may also be used as a visualization tool to help designers evaluate their immersive VR scenes when explored by users.

2. Related Work

2.1. Visual Attention

Vision is a major way for human beings to obtain information in our world; the percentage of doing so even exceeds 90% in some situations [1], such as while driving. Knowing vision attention behaviors of users in interactive virtual environments can help designers understand the information exchange between users and the content. Eye movement plays a very important role in determining gazed areas or objects [2]. However, as a result of the inattentive blindness phenomenon [3], which was observed when viewers were concentrated on a task, the probability that the user may not notice details irrelevant to the task even within their foveal focus increases with the intensity of the task [4].

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 content and indicates users' vision attention behaviors. The subjective method and cognition research results are borrowed into objective methods to indicate the relationship between content features and observation patterns. These two categories always combine with each other. As in the subjective method, the specific features must be extracted to analyze their impact on user reactions, especially involving eye movement. Meanwhile, the subjective method is always used as a verification tool for the accuracy of the objective method [5].

The eye tracking system, which is able to directly record users' eye movements, is a typical subjective method for estimating visual attention. The eye movement can then be used as strong evidence to infer what is perceived by the user [6]. As the eye tracker system can only record the eyes' gazing directions, it still requires much more effort to establish what the user is observing [7]. Some works establish this relationship based on screen pixels, such as determining the pattern of which kinds of pixels affect the eye movement of users. Other works move forward to discover the relationship between semantic objects and their features in the contents with eye movement of subjects [5].

There are some pioneer studies that embed eye trackers into VR HMDs [8]. Even though such is far from mainstream in modern commercial VR HMDs, this makes obtaining eye movement possible in immersive VR environments. However,

¹ Website: <https://www.oculus.com/story-studio/films/lost/>

eye movement is only the first stage to use eye tracking systems. Raw data for eye mobility contains eye direction positions on-screen, which then must be translated into contents within the virtual scene. We argue that eyeball movements can only reflect subtle user gazing behavior in low-scale scenes with a passive mode, such as viewing a picture or video. Gaze behavior mainly comes from head motion, especially in large-scale scenes that are dominant in VR. Consider humans' gaze behavior in the real world. We seldom only move our eyeballs when seeking something. Instead, people rotate their heads to perform large-scale viewing and then use eye movement as a preliminary seeking method or a guide for following head movement. Furthermore, many eye movements are saccades, which are both voluntary and reflexive. There is virtually no visual information cognitively processed during a saccade [6]. The eyes are never completely still; rather, they always jitter, and small movements called tremors or drifts [9] occur, which are meaningless to final perception. Nevertheless, eye tracking-equipped VR HMD devices can also use the visual attention evaluation method proposed in this paper to make full use of the eye motion raw data in detailed levels while exploring immersive VR scenes.

Objective methods, on the other hand, can predict whether the user would pay attention to some objects or areas. The prediction result has to be evaluated by eye tracking or other subjective methods [5, 7, 10]. 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 [11]. 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 [10, 12]. These two types of methods are always combined to achieve a better result that better coordinates with human beings [10, 12-13].

Very recently, Vincent *et al.* addressed the problem of how users explore virtual environments and what constitutes saliency in immersive applications [14]. 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 [14]. A virtual object-based visual attention is obtained in our paper rather than the conventional region-based saliency map obtained in [14].

Our method in this paper can be categorized as a 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 [15-16]. A task is a key factor for guiding the exploring behavior of users in immersive VR environments. This finding is used to set up our user study scenes to make sure the task-related objects receive enough interest from the users. To the best of our knowledge, the method proposed in this paper is the first subjective way to estimate user attention in immersive VR environments without any extra hardware.

2.2. Applications of Visual Attention

Predicting users' attention behavior has been well researched in some areas like image processing, video analysis, and graphical user interface design. Many practical applications have taken advantage of the results of users' attention behavior research, such as user experience analysis [17-18], natural interaction design [19], user interface design [20], automatic recognition [13], and scene alignment [21].

Recently, user visual attention research has been involved in interactive virtual environment applications [5, 14, 22]. The main purpose is to help designers improve their work and enhance user experience. In papers [23] and [24], 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. This is our research goal as well. To evaluate users' visual attention of different objects in virtual scenes, designers can easily notice whether the users' exploration behavior is under their design plans.

In the immersive video field, Thomas *et al.* presented a visualization framework for analyzing head movements and gaze data for immersive 360° videos [22]. A specialized view similarity visualization method was provided, and it 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 is 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 [25]. Such systems mainly utilized the objective method to automatically extract the most important shots in contents and organize them into scripture boards [25]. 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 also help designers qualify their work in an intuitive way.

3. Immersive VR Exploring Feature

As discussed in paper [26], 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 [26]. 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 instead to gaze at objects that appear on the screen.

According to a recent state-of-art paper [27], quantitative evaluation is considered the cornerstone of the scientific discovery process. The evaluation can always be 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.

Therefore, 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 works, such as the recently published paper [28]. As inattention blindness exists, participants most likely will not remember everything that they just observed [3]. Furthermore, many errors might be introduced in using a post-process questionnaire system.

As interactions are so different between immersive and non-immersive VR scenes, there exist distinct differences of gazing behavior between these two types of applications. In order to find the difference and further find the gaze pattern in immersive VR environments, we set up a series of user study scenes running both in immersive and non-immersive modes to collect specific behavior data for analyzing.

4. 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. It has been proven that the task-related gaze behavior can dominate over saliency [16]. In each scene, there are at most three math problems displayed upon particular objects called task-related objects. The participants are required to solve these simple math problems and input their answers after the timer runs out or they press the quit button. We hypothesized that participants in immersive VR scenes tend to move their heads towards the task-related objects, while simply scanning with their own eyeballs with a stable virtual camera in non-immersive scenes. 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 on the gazing behavior. In order to diagnose the exact relationship between task-related objects and the users' viewing direction, we allowed them to keep translating in different trajectories through particular scenes. We hypothesized that the users would try to follow the moving objects in immersive scenes, but not always in non-immersive scenes.

We used the Unity 3D² game engine to establish the five scenes for the user study. Each scene will run in both immersive and non-immersive modes. In each user study scene, there are one or two game objects with two or three math problems for participants to solve. Each math problem involves only addition equations between two numbers less than ten (as seen from Figure 1(a)).

These five user study scenes are named *1static*, *2static*, *1move*, *2moveS*, and *2move* respectively. 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 *1static* contains only one task-related object, while scene *2static* contains two. In order to determine whether participants' view direction is strongly correlated to task-related objects, we set the objects to be active in the last three scenes. In the *1move* scene, there is only one moveable task-related object moving horizontally, back and forth. The *2moveS* scene contains two in up-and-down and left-and-right trajectories. The final scene *2move* contains two moveable 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 hypothesize 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.

4.1. User Study Configuration

We want to concentrate on gaze behavior because it 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 the 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 the first-person view.

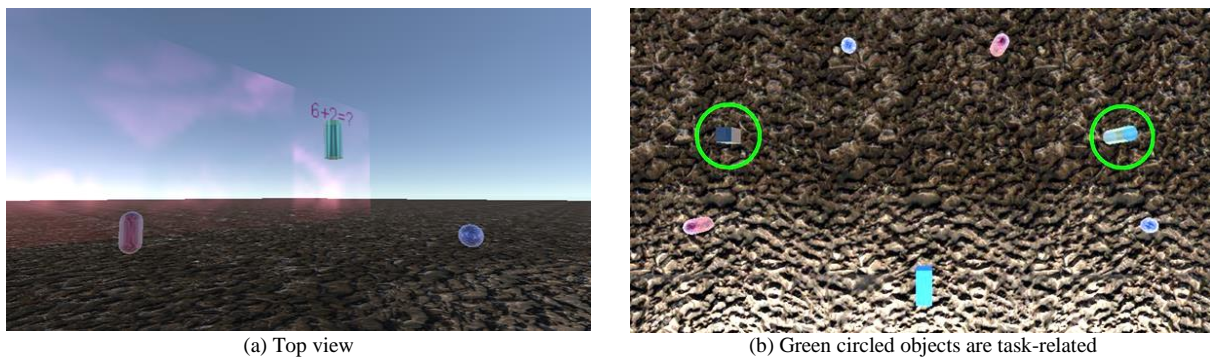


Figure 1. Shots captured from the user study scene

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 the 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.

The screenshots based on the visual attention evaluation equation proposed in Section 5 will be saved as well. If an object's accumulative attention exceeds 3, this means the user has gazed it for around three seconds, and the screen shot will be saved every time this criterion has been met.

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

² Website: <https://unity3d.com/>

4.2. 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.

They participated with informed consent under a protocol approved by the Institutional Review Board. They were notified that they should try their best to pass the test, in which the purpose is to calculate all math problems. After fully understanding the purpose and control method of the user study application, they were asked to play these scenes in both the VR modes.

Just as the hypothesis stated in the beginning showed, most participants acted differently in the immersive mode than in the non-immersive mode. In the immersive mode, they tended to focus on the task-related objects exclusively when they were trying to solve the math problems upon the objects. In the non-immersive mode, however, they tended 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 the real world, where they tend to keep the goal right in the center of their vision until reaching a limitation, and then user eye movement acts as a complement. In the 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.

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 *2moveS*. We can see clearly that in the 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 the 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. As he was focusing, the included angle fell below 10° . It is 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 the non-immersive mode.

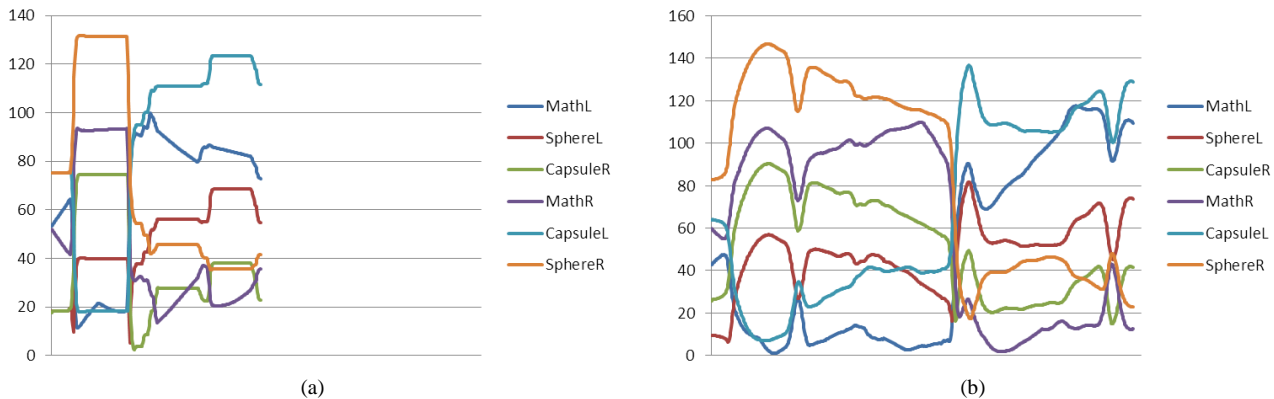


Figure 2. Included angles of a participant with different objects in scene *2moveS* using non-immersive (a) and immersive (b) methods, respectively; the X axis is time, and the Y axis is included angles (a) non-immersive mode (b) immersive mode

5. Object-based Attention Quantitative Equation

We can conclude from Section 4.2 that objects receiving the most user attention continuously appear in the center of the screen in the immersive VR mode. In the non-immersive mode, users have more freedom to use their own eyes, which is more convenient for gazing at 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, and 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 that

the logistic function fairly meets our requirements, as it emphasizes the center effect. The effect of the revised logistic function (Equation (1)) used in our experiment can be seen in Figure 3. From the figure, we can see that included angles falling below 15° have similar yet important attention values, while angles exceeding 20° indicate that the attention drops, approaching zero. As summarized from our user study results introduced in Section 4.2, included angles below 10° can be seen as focusing in immersive VR scenes. We add 5° as a tolerance, so the threshold degree is 15°. Therefore, the object-based visual attention value A of object O in an immersive VR scene can be calculated with Equation (1), a revised logistic function.

$$A(O) = \int \frac{1}{1 + e^{k(deg-d)}} dt \tag{1}$$

Where k denotes the steepness of the curve; here, we use 0.5. deg denotes the included angle degree between the 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, we use 15° as the default value. As summarized from our user study results introduced in Section 4.2, 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 [21].

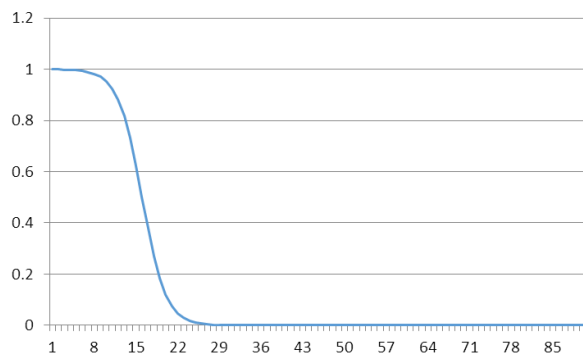


Figure 3. Revised logistic function effect used in our system; the X axis is the independent variable, which means the degree of the included angle used in our application, and the Y axis is the function value

From Figure 4, 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 the immersive VR mode, the two task-related objects received the highest attention values. This means that the participant always put them in the center of his or her view. In the non-immersive VR mode, the participant kept the view unmoving so much that the background object *CapsuleR* was put in the center of his or her view unintentionally, resulting in its highest attention value.

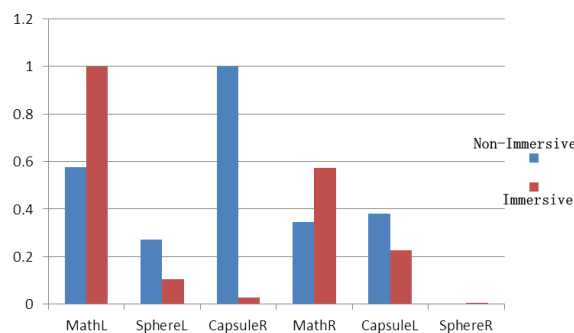


Figure 4. A user's normalized attention degree calculated by Equation (1) in scene *2moveS* using non-immersive and immersive modes

Using the object-based attention equation can uncover user perception experience in a quantitative manner. From the attention score, level designers may diagnose if users are following the intentional purpose that paying enough attention to particular objects will achieve. Incorporated with the included angle chart showed in Figure 2, level designers can have a clear view of the exploration process of users. This will be helpful to efficiently review their design quality without requiring time-consuming surveillance or a personal inspection.

6. Applications

Using the object-based visual attention evaluation method proposed in Section 5, 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 included angle metric and object-based attention equations 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 their scenes are well explored, followed by their design purposes.

We developed an automatic storyboard generation system using the object-based visual attention calculation equation proposed in Equation (1). It can further help designers understand how users explore their work and enrich the user experience as a complimentary result for the user.

In conventional films and television shows, cameras are controlled by directors so that the audience experiences the exact the same content passively. Audiences of immersive VR content, however, have much more freedom than conventional media, since they can control the camera to explore whatever they want, so every immersive VR audience may have their own viewing experience. Using storyboards, we can depict each user's personal exploring process. The challenge is how to decide the metric for key frames.

We use Equation (1) as a criterion to decide whether 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 4.2, 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 set up an immersive VR short film using resources from the cutting-edge demo Adam³, 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 guard character stays in the front left side of the user. Character Adam will walk straight starting from the door to the other side of the room. From then, character Lu will walk by on the same path. Character Sebastian follows, walking from the door to stop just in front of the user. He will then raise his hand, pointing to the user.

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. The accumulator is reset to zero as the following conditions happen:

- The current frame is saved as a key frame as the accumulator exceeds 3.
- The user moves his or her view away from a certain object for more than 0.5 seconds. This can prevent some background objects from receiving too much attention if they are located on the way to the searching path, so the users' view always falls inside it unintentionally.

As can be seen from Figure 5, six screen shots were captured by our personalized storyboard system. These key frames roughly illustrate the 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 5 can also help us indicate some design issues. Character Lu seems to be missed by the participant, as he did not receive much attention.

Our system successfully captured the key frames, making it easier to understand what the user has perceived from his or her immersive VR experience. Nevertheless, when particular viewer guide hints are lacking, the user tends to look around until something captures his or her eye. The personalized story board can collect users' experience data in visual attention without reviewing the entire video records. The user study scenes' results can be seen in Figure 6. As can be seen from the figure, all key frames of this participant are about task-related objects that verify that our user study design purpose is suitable to use tasks to attract the participants' attention. We showed these storyboard to the participants and asked their thoughts. Most of them were surprised that the results accorded with their own experience so well. They agreed that the storyboard can help them remember the most impressive moments during the VR experience.

³ Website: <https://unity3d.com/pages/adam>

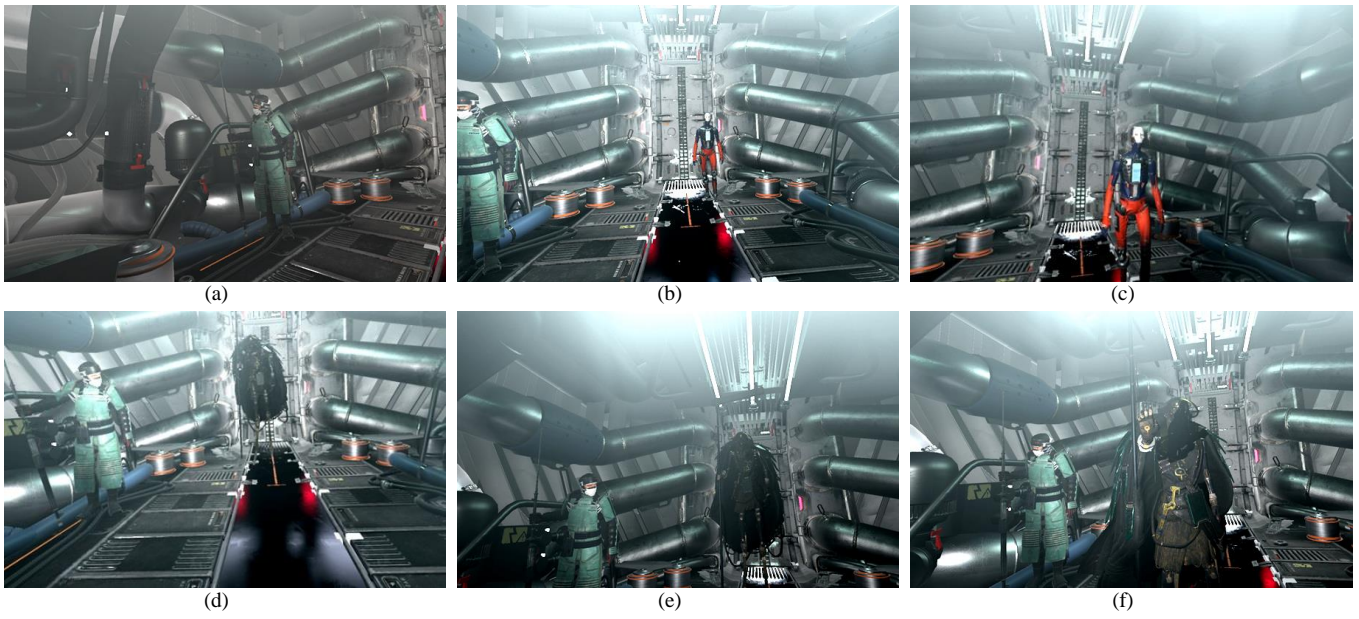


Figure 5. A participant's personalized experience storyboard captured by our system



Figure 6. A participant's personalized experience storyboard captured by our method for user study scenes

7. Conclusions

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 introduced in Section 4, and the storyboard generation system demo used in Section 6 can be found in the GitHub repository: <https://github.com/hanhonglei/VRViewGuide>

Acknowledgements

The authors would like to acknowledge the support provided by the China Scholarship Council and UNC Charlotte during Honglei Han's visit to UNC Charlotte.

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

References

1. M. Sivak, "The Information that Drivers Use: Is it Indeed 90% Visual?" *Perception*, Vol. 25, No. 9, pp. 1081-1089, 1996
2. K. Rayner, "Eye Movements and Attention in Reading, Scene Perception, and Visual Search," *The Quarterly Journal of Experimental Psychology*, Vol. 62, No. 8, pp. 1457-1506, 2009
3. A. Mack and I. Rock, "Inattentional Blindness," MIT Press, MA, 1998
4. D. J. Simons and C. F. Chabris, "Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events," *Perception*, Vol. 28, No. 9, pp. 1059-1074, 1999
5. R. J. Peters and L. Itti, "Applying Computational Tools to Predict Gaze Direction in Interactive Visual Environments," *ACM Transactions on Applied Perception*, Vol. 5, No. 2, pp. 1-19, 2008
6. A. T. Duchowski, "Eye Tracking Methodology: Theory and Practice," Springer-Verlag, London, 2007
7. M. Bernhard, E. Stavrakis, and M. Wimmer, "An Empirical Pipeline to Derive Gaze Prediction Heuristics for 3D Action Games," *ACM Transactions on Applied Perception (TAP)*, Vol. 8, No. 1, 2010
8. "Tobii Eye Tracking in Virtual Reality," (<https://vr.tobii.com/>, Last accessed on October 23, 2018)
9. R. J. Snowden, P. Thompson, and T. Troscianko, "Basic Vision : An Introduction to Visual Perception," Oxford University Press, Oxford, 2012
10. 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," *IEEE Transactions on Visualization & Computer Graphics*, Vol. 18, No. 3, pp. 356-368, 2012
11. L. Itti, "Quantifying the Contribution of Low-Level Saliency to Human Eye Movements in Dynamic Scenes," *Visual Cognition*, Vol. 12, No. 6, pp. 1093-1123, 2005
12. S. Lee, G. J. Kim, and S. Choi, "Real-Time Tracking of Visually Attended Objects in Virtual Environments and its Application to Lod," *IEEE Transactions on Visualization & Computer Graphics*, Vol. 15, No. 1, pp. 6-19, 2009
13. K. Yun, Y. Peng, D. Samaras, G. J. Zelinsky, and T. L. Berg, "Studying Relationships between Human Gaze, Description, and Computer Vision," in *Proceedings of Computer Vision and Pattern Recognition*, pp. 739-746, 2013
14. V. Sitzmann, A. Serrano, A. Pavel, M. Agrawala, D. Gutierrez, and G. Wetzstein, "Saliency in VR: How Do People Explore Virtual Environments?" *IEEE Transactions on Visualization and Computer Graphics*, Vol. 24, No. 4, pp. 1633-1642, 2018
15. V. Sundstedt, M. Bernhard, E. Stavrakis, E. Reinhard, and M. Wimmer, "Visual Attention and Gaze Behavior in Games: An Object-based Approach," Springer-Verlag, London, 2013
16. M. Land, N. Mennie, and J. Rusted, "The Roles of Vision and Eye Movements in the Control of Activities of Daily Living," *Perception*, Vol. 28, No. 11, pp. 1311-1328, 1999
17. G. N. Yannakakis, H. P. Martínez, and A. Jhala, "Towards Affective Camera Control in Games," *User Modeling and User-Adapted Interaction*, Vol. 20, No. 4, pp. 313-340, 2010
18. P. Burelli, "Virtual Cinematography in Games: Investigating the Impact on Player Experience," in *Proceedings of International Conference on the Foundations of Digital Games Society for the Advancement of the Science of Digital Games*, pp. 134-141, 2013

19. P. Majoranta and A. Bulling, "Eye Tracking and Eye-based Human-Computer Interaction," Springer-Verlag, London, 2014
20. T. Nakayama, H. Kato, and Y. Yamane, "Discovering the Gap between Web Site Designers' Expectations and Users' Behavior," *Computer Networks*, Vol. 33, No. 1-6, pp. 811-822, 2000
21. F. Alt, A. S. Shirazi, A. Schmidt, and J. Mennen, "Increasing the User's Attention on the Web: Using Implicit Interaction based on Gaze Behavior to Tailor Content," in *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, pp. 544-553, Copenhagen, Denmark, 2012
22. 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," in *Proceedings of Workshop on Eye Tracking and Visualization*, pp. 141-146, 2015
23. A. Picardi, P. Burelli, and G. N. Yannakakis, "Modelling Virtual Camera Behaviour through Player Gaze," in *Proceedings of the 6th International Conference on Foundations of Digital Games*, pp. 107-114, 2011
24. P. Burelli and G. N. Yannakakis, "Towards Adaptive Virtual Camera Control in Computer Games," in *Proceedings of International Symposium on Smart Graphics*, pp. 25-36, Berlin, Heidelberg, 2011
25. 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," *IEEE Transactions on Multimedia*, Vol. 7, No. 5, pp. 907-919, 2005
26. M. Slater, "Place Illusion and Plausibility Can Lead to Realistic Behaviour in Immersive Virtual Environments," *Philosophical Transactions of the Royal Society of London*, Vol. 364, No. 1535, pp. 3549-3557, 2009
27. C. Papadopoulos, I. Gutenko, and A. E. Kaufman, "Veevvie: Visual Explorer for Empirical Visualization, VR and Interaction Experiments," *IEEE Transactions on Visualization & Computer Graphics*, Vol. 22, No. 1, pp. 111-120, 2016
28. 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," *IEEE Transactions on Visualization & Computer Graphics*, Vol. 22, No. 4, pp. 1406-1414, 2016

Honglei Han joined the Communication University of China after receiving his M.S. degree in computer science in 2006. He is currently an associate professor. He received his Ph.D. from the Institute of Software at Chinese Academy of Sciences in 2015. He was a visiting scholar at the University of North Carolina at Charlotte. His research interests include computer graphics, virtual reality, and computer games.

Aidong Lu received her bachelor's and Master's degrees in computer science from Tsinghua University in 1999 and 2001, respectively, and her Ph.D. in electrical and computer engineering from Purdue University in 2005. She is currently an associate professor at the University of North Carolina at Charlotte. Her research interest is developing effective visualization approaches to improve visual communications in real-life applications and education.

Chanchan Xu obtained her Master's degree in computer application technology from Beijing Forestry University in 2013. She is currently a graduate student at the Communication University of China. Her research interests are computer graphics and virtual reality.

Unique Wells is a Master's degree candidate at the University of North Carolina at Charlotte. Her research interests are data visualization and virtual reality.