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](mailto:uwells@uncc.edu)

**Abstract:** A method to measure what and how deep the user can perceive in immersive virtual reality environment, based on the immersive head-mounted display, is proposed. A preliminary user study was conducted to verify that user gaze behavior has specific differences in immersive virtual reality environment compared with that in conventional non-immersive virtual reality environments, based on the non-immersive desktop screen. Gathered from the user study result for gaze behavior, the users in immersive virtual reality environment are more likely to adjust their head movement to let interesting objects center in their fields of the view, while in non-immersiver virtual reality environment the users tend to move their own eyes and control the avatar's head when necessary. Based on this finding, a quantitative method is proposed to measure the user's visual attention in immersive virtual reality environment. A so-called user personalized storyboard, based on the attention evaluation method, is proposed to capture the user's most perceived views as key frames which can depict users' exploration experience in immersive virtual reality environment.

**Keywords:** Gaze analysis, visual attention, virtual reality, interaction, eye tracking

# 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 experience them through rendered images which are actually the views from a virtual camera. With the evolution of modern three dimensional rendering engines and virtual reality Head Mounted Displays (HMDs), users can experience an immersive feeling which has been something pursued in the computer graphics field for decades. In this paper, we only discuss HMDs based immersive VR applications which is the most popular recently, not including CAVEs or other systems.

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 out. It is special in immersive VR that users can explore using natural interaction methods, which are much easier to utilize excessively. Therefore, if there lacks sufficient 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 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 level. Tasks can attract users' attention, and task-related content may serve as strong hints to guide users' exploration behavior. In immersive VR films, however, it's easy to get lost since there usually lacks specific tasks. The immersive VR short film named "Lost" produced by Story Studio [[1](#_ENREF_1)], 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 would guide the audiences' view to ensure they follow the narrative of the film.

Therefore, confirming whether or not 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. Especially true for an immersive VR video, the story cannot even continue effectively unless the user follows the view guidance designed by the film director. It is vital for immersive 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. This assists the directors with understanding their audiences’ experiences, which would then help give ideas on how to improve them.

According to a recent state-of-art paper [[2](#_ENREF_2)], 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. We propose a hypothesis involving 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.

We decided to use behavior analysis methods based on recorded participants' particular interactions, that better reflect the real-time exploration interests, instead of questionnaires as used in many other user interaction evaluation researches such as the recently published paper [[3](#_ENREF_3)]. As inattention blindness exists, participants most likely will not remember everything that they just observed [[4](#_ENREF_4)]. Furthermore, many errors might be introduced in using a post-process questionnaire system.

The analysis results verified our hypothesis that in immersive VR, viewers instinctively move their heads to gaze at objects they are most interested in. They adjust their own line of vision to focus on the most interesting object into their center of view. In the case studies the users move their heads, trying to focus on the moving significant objects, but when a new and more eye-catching object appears nearby, they move their heads to focus on that instead. 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 and redirecting the avatar's field of view when necessary in non-immersive VR.

From this finding, a quantitative method of user visual attention is derived to measure how deep the objects are perceived in virtual environments by the audience. We use the logistic function with a parameter of how focused the object is on scale with the quantitative math model.

Finally, the visual attention quantitative method is used for applications to capture the most impressive shots when subjects are exploring the immersive VR scenes. This so-called personality storyboard can be seen as a user exploring summarization to enrich his or her own experience with using immersive VR. This storyboard can also be used as a visualization tool to help designers evaluate their immersive VR scenes when explored by audiences.

# Related work

## 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 [[5](#_ENREF_5)], 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 [[6](#_ENREF_6)]. However, as a result of the inattentional blindness phenomenon [[4](#_ENREF_4)], 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 [[7](#_ENREF_7)].

The visual attention quantitative method 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 content 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 researches results are borrowed into objective methods to indicate the relationship between content features and observation patterns. These two categories, however, always combine with each other. As in the subjective method, the specific features have to be extracted to analyze their impact to user reactions, especially involving the eye movement. Meanwhile, the subjective method is always used as a verification tool for the accuracy of the objective method [[8](#_ENREF_8)].

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 [[9](#_ENREF_9)]. 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 [[10](#_ENREF_10)]. 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 [[8](#_ENREF_8)].

There are some pioneer studies that embed eye trackers into VR head mounted displays [[11-13](#_ENREF_11)]. Even though such is far from mainstream in modern commercial VR HMDs, this makes obtaining eye movement possible in immersive VR environments. However, 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 has to 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 which are dominant in VR. Consider humans' gaze behavior in the real world. We seldom only move our eye balls 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, and there is virtually no visual information cognitively processed during a saccade [[9](#_ENREF_9)]. The eyes are never completely still; rather they always jitter, when small movements called tremors or drifts [[14](#_ENREF_14)] 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.

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. In paper [[15](#_ENREF_15)], three criteria of evaluating objects' attention in non-immersive VR scenes were proposed: Center, occlude percentage, and distance. These are followed by user studies conducted to verify the efficiency of the attention quantitative method. More camera control research can be found in the specific survey paper [[16](#_ENREF_16)].

Objective methods, on the other hand, 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 [[17](#_ENREF_17)]. 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 [[18](#_ENREF_18), [19](#_ENREF_19)]. These two types of methods are always combined to achieve a better result which better coordinate with human beings [[19](#_ENREF_19), [20](#_ENREF_20)].

Objective methods 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 [[8](#_ENREF_8), [10](#_ENREF_10), [19](#_ENREF_19)].

To the best of our knowledge, even though the view guide in immersive VR environments is critical in a VR narrative, the fundamental research of gaining quantitative user attention in immersive VR environments similar with what we proposed in the paper is rare. Our method in this paper can be categorized into the subjective method, since it relies on users' head movements as a strong hint of gaze and attention, but differently from other subjective methods, we didn't use eye movements. Instead, we used head motion as the main criterion of gazing in immersive VR. The top-down idea in objective methods inspired us greatly, especially concerning the importance of task oriented gaze guiding[[21](#_ENREF_21)]. 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 would receive enough interest from the users.

## Applications of visual attention

How to predict 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 these researches' results, such as user experience analysis[[22](#_ENREF_22), [23](#_ENREF_23)], natural interaction design[[24](#_ENREF_24)], user interface design[[25](#_ENREF_25)], automatic recognition[[20](#_ENREF_20)], and scene alignment[[26](#_ENREF_26)].

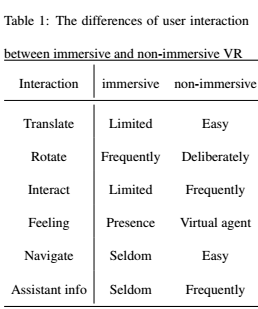
Recently, user visual attention research has been involved in interactive virtual environment applications[[8](#_ENREF_8)]. The main purpose is to help designers improve their work and enhance user experience. In paper[[27](#_ENREF_27)] and [[28](#_ENREF_28)], the authors established the relationship between user experience and user gaze data through virtual cameras. As they believe that user visual attention is very important to user experience in interactive virtual environments, they try to use the attention model to make sure 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 or not the users' exploration behavior is under their design plans.

We developed a software method to collect virtual objects' centerness data that clearly represents the user gaze interest. A visual attention estimation method is proposed based on gathering enough sample data. This method can help the designers assemble an overview from the data and recognize whether or not most users are guided by the intended design strategy. It has also been used in personalized storyboard generating applications. There have been many other user behavior abstraction systems proposed in recent years, like video summarization[[29](#_ENREF_29)]. Such systems mainly utilized the objective method to automatically extract the most important shots in contents and organize them into scripture boards[[29](#_ENREF_29)]. However, using the proposed quantitative method for visual attention estimation derived from the user study, our method can extract the most important objects, based on the user's own exploration behavior. Everyone has his or her own gaze trajectory in the same immersive VR scene. The trajectories may seem similar because of the specific design purpose of the scene, but there may also be many differences, depending on each personality. The personalized storyboard system proposed in this paper can not only improve user's experience as a complimentary result, but also help designers to qualify their work in an intuitive way.

# The features of exploring behavior in immersive VR environment

There are a number of substantial differences of user behaviors between immersive and non-immersive VR applications, as summarized in Table 1. In the immersive VR environment the users typically feel immersed into the virtual environment, as if they are actually in the simulated world, so they emulate behaviors and motions similar to ones portrayed in the real world. As shown in Table 1, for example, users move their heads, scanning the area they are in to try and capture objects they find interesting. 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 virtual avatar as the third person, controlling it to explore the virtual world instead of placing themselves in the virtual avatar's shoes.

Immersive VR equipments make users feel as though they are within the environment, but in a conventional non-immersive VR environment users explore the simulated world in an observer's view, with the full acknowledgement of being out of the virtual world. Users tend to move their heads more often in immersive VR environments 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.

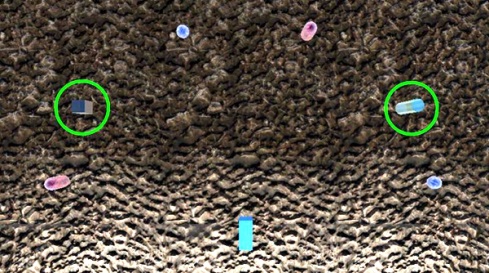
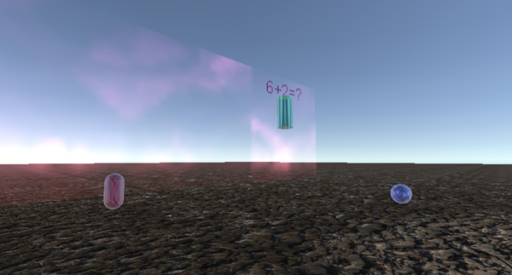


As interactions are so different between immersive and non-immersive VR scenes, there exists distinct differences of gazing behavior between these two types of applications. In order to find the difference and further find the gaze pattern in immersive VR environments, we set up a series of user study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

# 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. In order to achieve this goal, we set particular items as task-related objects, and the others as background objects, since it has been proven that task-related gaze behavior can dominate over saliency[[30](#_ENREF_30)]. In each scene, there are at most two 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 material similar to the background to make it slightly troublesome to 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' view direction, we let them keep translating in different trajectories through particular scenes. We hypothesized the users would try to follow the moving objects in immersive scenes, but not always in non-immersive scenes.

We used Unity 3D[[1]](#footnote-1) to establish the five scenes for the user study. Each scene will run in both immersive and non-immersive modes. In each study, there are particular numbers of game objects with math problems for participants to solve. Each math problem is only addition equations between two numbers less than ten (as seen from Figure 1(a).



(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

These five user study scenes are named 1*static*, 2*static*, 1*move*, 2*moveS*, and 2*move* respectively. They all have the same scene arrangement, as can be seen from Figure 1(b). Objects circled in green have math problems on top of them, and a camera is located at the bottom middle (see the blue box as 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 rooted task-related object, while Scene 2*static* contains two rooted task-related objects. In order to find out if users' 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 moving task-related objects 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. The purpose of adding background objects is to compare attention differences between them and task-related objects.

## 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 can use VR HMDs, latest Oculus Rift HMDs are used in our experiments, to explore the scene with a natural interactive manner. In non-immersiver mode, the user study scenes run using a common PC, and participants can navigate using a mouse as the first person controller that is common in many action PC games.

The field of view in non-immersive mode is 60°while in immersive one is 96°which are common in most applications. When an object falls out of user's field of view, they would be marked as culled, and their information would not be recorded.

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 users concentrate on tracing and focusing on the task-related objects. If the user solves all the problems early, they have the option of using the “Space” key to skip to the answer stage directly.

We record behavior data in every fixed frame along with gaze information involved with the objects, including task-related and background objects. These data are relative to position in the camera coordinates, as well as their projection position in screen coordinate.

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 user.

The screenshots based on the visual attention evaluation equation proposed in Section 4 will be saved as well. If an object's accumulative attention exceeds 3, this means the user has gazed it for around three seconds, the screen shot will be saved every time this criteria 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, half of the participants will experience non-immersive mode first, and the other half will experience immersive mode first.

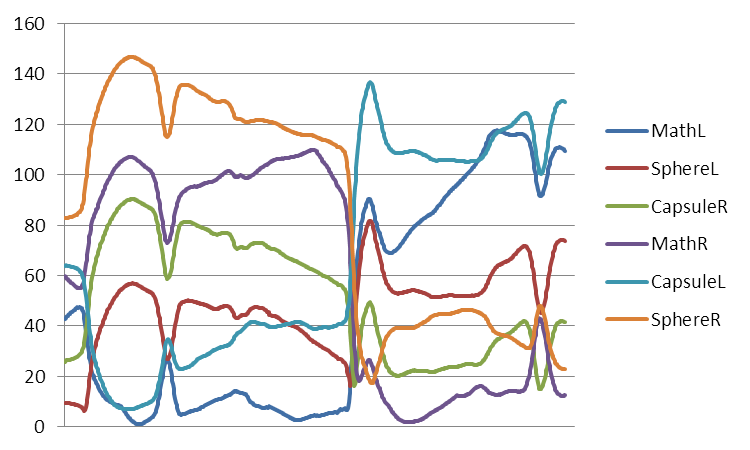
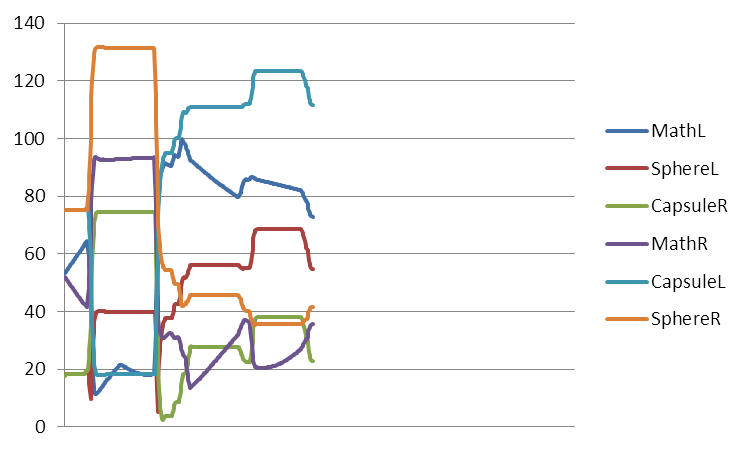
## Result analysis

We invited twenty-five participants who are all students and faculty members in our university, aging from 19 to 48 with 15 males and 10 females. Their background are diversified, and all familiar with how to navigate in non-immersive VR applications. But only 5 of them once experienced HMDs based immersive VR scenes.

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 both in immersive and non-immersive mode respectively.

Just as what the hypothesis in the beginning showed, 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 environment people act like they do in real world, in which they tend to keep the goal right in the center of their vision until reaching a limitation, then user eye movement as a complement. In non-immersive mode, using a mouse or some other interaction method to control the virtual camera would trigger a series of computer processes to refresh. Even though this process is efficient thanks to current computer capabilities, repositioning the camera can still cause instability that makes focusing on particular objects difficult. User interaction methods to move the virtual camera while solving the math problems would at least somewhat distract users' attention. 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 seen from Figure 2. The *X* axis represents the user experience time and the *Y* axis represents the included angle. In interactive virtual environments the gaze tends to be more focused on the center of the screen [[10](#_ENREF_10), [31](#_ENREF_31), [32](#_ENREF_32)], so the included angle can reflect the user's focus degree. A smaller degree means there is more received attention from the user. Figure 2 depicts a participant's exploration process in the user study scene *2moveS*. The participant is required to solve the math problems floating over two moving objects, named *MathR* and *MathL*, and the others in the scene are background objects. From Figure 2, 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)), on the other hand, the curves are not so regular compared with the ones in non-immersive mode, as users tended to move their heads the whole time. This participant tended to keep his head moving and tried to focus on the task-related object when he was solving the math problems upon it. The included angle when he was focusing fell below 10°. More user study results can be found in the files attached. It's worthy to point out that the participant got a 100% answer correct rate in this scene using both immersive and non-immersive modes.



(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

# 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 immersive VR mode. In non-immersive mode, users have more freedom to use their own eyes, which is more convenient for exploring the artificial scene on the monitor instead of using a mouse and keyboard to control the virtual camera.

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 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 can fairly meet 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. Included angles falling below 15°have similar yet important attention values, while angles exceeding 20°means 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°.

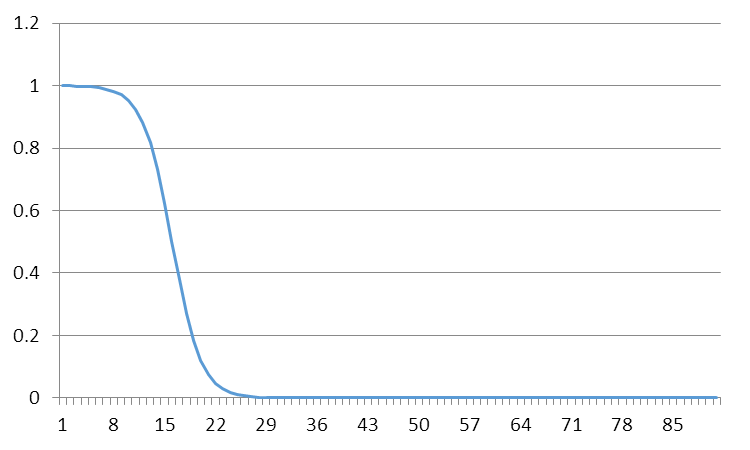


Figure 3. Revised Logistic function effect used in our system, *X* axis is independent variable, and *Y* axis is function value

The object based visual attention equation built on the Logistic function, used in our system, can be seen from Equation 1

where *deg* denotes the included angle degree between view direction and the vector pointing from the camera to the object. *k* denotes the steepness of the curve, here we use 0.5. *d* denotes the sigmoid curve's midpoint, which means that for the threshold of gazing included angles, we use 15°as the default value, deduced from our user study experiments. The attention value is accumulated along with time to prove the statement that the more subjects have interest in an object, the longer they will focus on it [[26](#_ENREF_26)].

From Figure 4, the apparent correlation between saliency levels of objects and their attention degrees can be easily seen. Included angle data used in the chart are from the same source in Figure 2. There, the two task-related objects get much higher scores than other background objects based on Equation 1. This is especially true in immersive VR scenes, but not always in non-immersive VR scenes. The participant got a perfect score for answers, which means the participant tried his or her best to gaze at the task-related objects. In immersive VR mode he or she really put the task-related objects in the center of the view, while in non-immersive VR mode the user kept the view unmoving so much that it unintentionally put object *CapsuleR* in the center of the view.

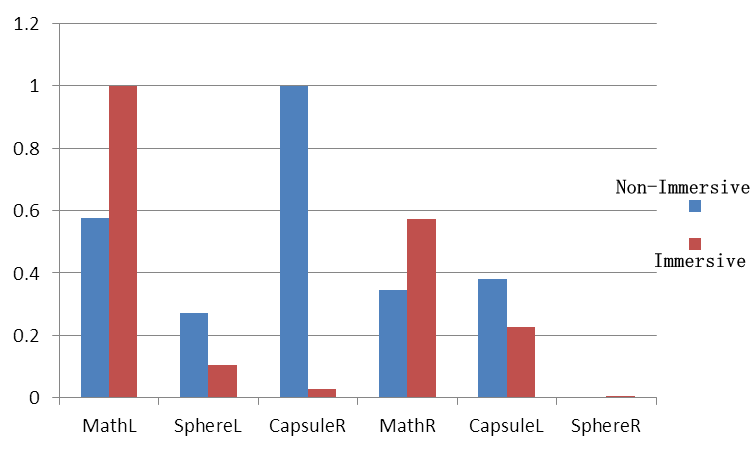


Figure 4. A user's normalized attention degree in scene *2moveS* using non-immersive and immersive modes respectively calculated by our method

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 having time consuming surveillance or a personal inspection.

# Applications

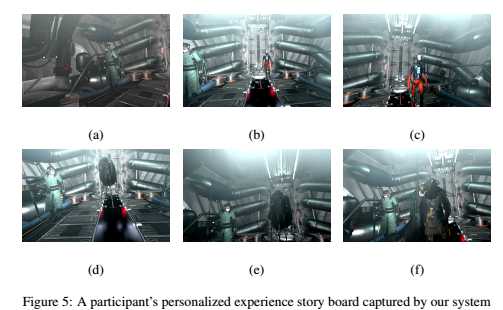
Using the object based visual attention evaluation method proposed in Section 5, designers can have an intuitive way to look into users' exploration patterns of their work. The series of quantitative charts or values used includes angle metric and object based attention equations that can be generated 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.

Through helping designers understand how users explore their work, and improving their work based on the understandings, they can enrich the user experience. We developed an automatic storyboard generation system using the object based visual attention calculation equation proposed in Equation 1.

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 if a shot can 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 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[[2]](#footnote-2) shared by Unity. There is an indoor realistic scene that lasts for 100 seconds, where four animated characters will appear in the scene. 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.



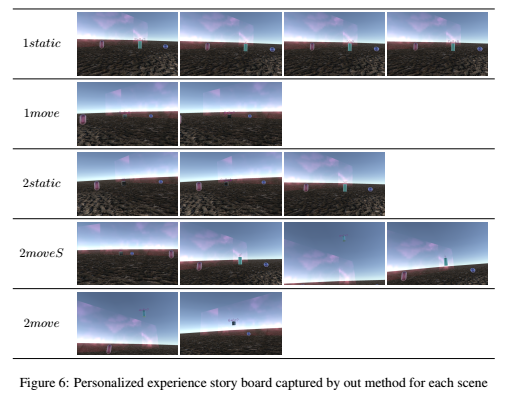
A script was assigned to the particular parts of each character and some background objects, in order to record their levels of attention. There is an accumulator parameter in the script to accumulate the attention value calculated by Equation 1 in each fixed frame. Once the accumulator exceeds the value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

1. The current frame has been saved as a key frame as the accumulator exceeds 3.

2. The user has moved 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 that 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 user's exploring process. In the very beginning, there was only a guard character in the scene. This user had no choice but to stare at him for a while. The user noticed that Adam came out of the door with a very strange appearance, so Adam attracted the user's attention (shown in the second and third frames). Finally, Sebastian walked in. He seemed like a boss, thus drawing a lot of attention from the user. The personalized story board generated by our system in Figure 5 can also help us indicate some design issues. Character Lu seems missed by the audience, as he didn't get enough attention from it.

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 there lacks particular viewer guide hints, 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 are about task-related objects that verify that our user study design purpose is suitable to use tasks to attract the participants' attention.



# Conclusion and future work

A quantitative object based attention evaluate method was proposed in this paper. Based on the results of the user study to diagnose the gaze pattern for task-related objects both in immersive and non-immersive VR modes, we found that head movement is vital, especially in immersive VR environment. According 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 use 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 for designers' review purposes.

Since only head movement is taken into account to evaluate the visual attention in immersive VR environment, the eye tracking system is not necessary. So our method can be easily embed into current immersive VR applications without bringing much calculation or equipment cost.

The position sensors in immersive VR equipments are supposed to be more stable in the near future. We should also allow translated degrees in immersive VR environments to be involved in user studies, and try to figure out their features in immersive VR environments.

The eye tracking system can also be incorporated into the visual attention evaluation process, especially when the user keeps his or her head still to gaze at two objects that are very close to each other. Eye movement can help distinguish which object the user is focusing on and put more weight onto it.

# Authors introduction:

1. Honglei Han, Ph.D. Associate professor. School of Animation and Digital Arts at Communication University of China.

Address: School of Animation and Digital Arts, Communication University of China, No.1 Dingfuzhuang East Street, Chaoyang District, Beijing, 100024, P.R. China.

Email: hanhonglei@cuc.edu.cn

Phone: 13691414966

2. Aidong Lu, Associate Professor. Department of Computer Science, University of North Carolina at Charlotte

Address: Woodward Hall 435F,UNC Charlotte, 9201 University City Blvd. Charlotte, NC 28223

Email:Aidong.Lu@uncc.edu

Phone: (704) 687-8570

# References

1. Story Studio, O. VR and LLC. Lost. 2017
2. C. Papadopoulos, I. Gutenko and A. E. Kaufman. VEEVVIE: Visual Explorer for Empirical Visualization, VR and Interaction Experiments. 2015
3. 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
4. A. Mack and I. Rock. Inattentional Blindness. 1998
5. M. Sivak. The information that drivers use: is it indeed 90% visual? 1996
6. K. Rayner. Eye movements and attention in reading, scene perception, and visual search. 2009
7. D. J. Simons and C. F. Chabris. Gorillas in our midst: Sustained inattentional blindness for dynamic events. 1999
8. R. J. Peters and L. Itti. Applying computational tools to predict gaze direction in interactive visual environments. 2008
9. A. T. Duchowski. Eye tracking methodology. 2007
10. M. Bernhard, E. Stavrakis and M. Wimmer. An empirical pipeline to derive gaze prediction heuristics for 3D action games. 2010
11. FOVE. 2017
12. SMI. VR and AR Eye Tracking. 2017
13. 7invensun. 7invensun. 2017
14. R. J. Snowden, P. Thompson and T. Troscianko. Basic vision : an introduction to visual perception. 2012
15. H. L. Han, X. U. Chan Chan, G. Z. Fei, X. Cui and W. C. Wang. Implicit measures of user attention in virtual environment navigation. 2014
16. M. Christie and P. Olivier. Camera control in computer graphics: models, techniques and applications. 2009
17. L. Itti. Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. 2005
18. S. Lee, G. J. Kim and S. Choi. Real-time tracking of visually attended objects in virtual environments and its application to LOD. 2009
19. 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
20. K. Yun, Y. Peng, D. Samaras, G. J. Zelinsky and T. L. Berg. Studying Relationships between Human Gaze, Description, and Computer Vision. 2013
21. V. Sundstedt, M. Bernhard, E. Stavrakis, E. Reinhard and M. Wimmer. Visual Attention and Gaze Behavior in Games: An Object-Based Approach. 2013
22. G. N. Yannakakis, H. P. Martínez and A. Jhala. Towards affective camera control in games. 2010
23. P. Burelli. Virtual cinematography in games: Investigating the impact on player experience. 2013
24. P. Majaranta and A. Bulling. Eye Tracking and Eye-Based Human–Computer Interaction. 2014
25. T. Nakayama, H. Kato and Y. Yamane. Discovering the gap between Web site designers' expectations and users' behavior. 2000
26. 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
27. A. Picardi, P. Burelli and G. N. Yannakakis. Modelling virtual camera behaviour through player gaze. 2011
28. P. Burelli and G. N. Yannakakis. Towards Adaptive Virtual Camera Control in Computer Games. 2011
29. 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
30. M. Land, N. Mennie and J. Rusted. The roles of vision and eye movements in the control of activities of daily living. 1999
31. A. Kenny, H. Koesling, D. Delaney, S. Mcloone and T. Ward. A preliminary investigation into eye gaze data in a first person shooter game. 2005
32. S. Yan and M. S. El-Nasr. Visual attention in 3D video games. 2006

1. Website: https://unity3d.com/ [↑](#footnote-ref-1)
2. https://unity3d.com/pages/adam [↑](#footnote-ref-2)