Analyzing Visual Attention via Virtual Environments

Haikun Huang[1]   Ni-Ching Lin[2]   Lorenzo Barrett[1]   Darian Springer[1]
Hsueh-Cheng Wang[2]   Marc Pomplun[1]   Lap-Fai Yu[1]
[1]University of Massachusetts Boston,   [2]National Chiao Tung University

Figure 1: *Humans rely on their visual attention to directional signage to navigate in an unfamiliar environment.*

Abstract

The widespread popularity of consumer-grade virtual reality devices such as the Oculus Rift and the HTC Vive provides new, exciting opportunities for visual attention research in virtual environments. Via these devices, users can navigate in and interact with virtual environments naturally, and their responses to different dynamic events can be closely tracked. In this paper, we explore the possibility of using virtual environments to study how directional signage may guide human navigation in an unfamiliar environment.

Keywords: Virtual Environments, Visual Attention, Visualization

Concepts: •**Computing methodologies → Perception; Virtual reality;**

1   Introduction

Humans navigate and interact with everyday environments primarily based on their visual attention. For example, a tourist walks to his or her destination in an unfamiliar subway station (Figure 1) by following the directional signage (e.g. text, signs, maps) he or she sees in the environment. Therefore, in order to optimize the flow and safety of human crowds, research on human visual attention is important and can lead to various practical applications to improve the design of public facilities (e.g., subway stations) and to increase commercial benefits through appropriate design of commercial premises (e.g., shopping malls).

The potential benefits brought about by visual attention research are huge, yet much of the existing research is still preliminary and in the laboratory stage. The major bottleneck is that most of the existing experiments are conducted in a highly controlled environment. For example, due to technical limitations, experiments on human visual attention are mostly done on 2D images, rather than in realistic 3D environments. Hence the experimental findings can hardly be generalized to support the development of practical tools for handling everyday scenarios. In recent years, there have been significant breakthroughs in virtual reality technology, which make virtual 3D environments practical for conducting human visual attention experiments. Using state-of-the-art virtual reality devices,

such as the Oculus Rift VR headset and the Virtuix Omni platform, one can navigate in virtual 3D environments like in the real-world. The visual perception in this type of virtual reality is highly realistic, and perceptual data can be collected for analyzing human visual attention and behavior that are almost identical to those in the real-world. In this project, we propose to conduct visual attention research in virtual reality and to use the analysis results to devise automatic 3D scene analysis tools to be used in everyday scenarios. Specifically, we aim at identifying the key visual features that influence human navigation and decision-making in everyday environments. We then train machine learning models to statistically encode these visual features. These models enable us to build 3D scene analysis tools for various applications.

2   Related Work

Visual attention is an important research direction in cognitive science to study and computationally model the behavior of humans. People rely on their visual attention to navigate in everyday environments. In order to direct their attention, they continually shift their gaze to retrieve information for decision making [Hwang et al. 2009; Hwang et al. 2011; Wu et al. 2014]. For example, by paying attention to directional signage (e.g., traffic signs), people can cross roads safely. Visual attention is known to be attracted by low-level features such as high contrasts, depictions of objects (e.g., cars, pedestrians, or text), and high-level scene context. The attraction of attention by low-level visual saliency has been extensively studied. Some particular classes of objects, however, are known to attract eye fixations independently of their low-level visual saliency. These include text [Wang and Pomplun 2012; Wang et al. 2012; Wang et al. 2014], people and other objects with special meaning.

In the past decades, visual attention studies were typically carried out using static stimuli (e.g., on a still 2D image). However, real-world scene inspection in everyday activity is generally dynamic. There has been investigation to track people's eye movements on a screen while they are watching video stimuli (sports and movies), but the scenario is still very different from real-world scene inspection, such as the 3D world perceived by a person walking on a street. Although wearable eye tracking systems have become available, it is still difficult to analyze the correspondence of eye movement data with the objects in real-world 3D environments captured by a scene camera. The recent advancement in virtual reality provides a new opportunity to study human visual attention in a much more realistic yet controllable manner, by tracking people's eye fixations in virtual 3D environments.

So far little attempt has been made on using virtual 3D environments to conduct visual attention research, or on using visual attention data to devise 3D scene analysis tools. Human visual attention data can provide very useful information for analyzing and design-
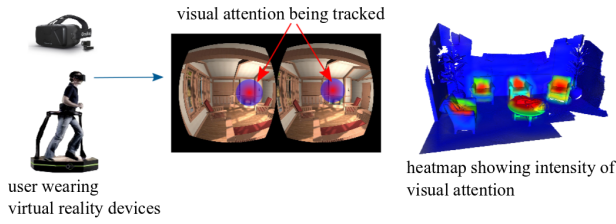
Figure 2: *Overview of our approach.*

ing 3D scenes, such that the scene designs automatically generated by graphics techniques [Yu et al. 2011; Yu et al. 2016] are optimized with respect to human ergonomics and visual attention. The generated designs can be put into practical uses in the real-world, for example, when planning the layout of airports or train stations and deciding where to place signs, maps, or other types of information relevant to navigation.

3 Approach

Virtual Reality Platform. We build a virtual reality platform on which a user can navigate freely in and view a virtual environment. Figure 2 shows our virtual reality platform, which is composed of the Virtuix Omni motion-tracking platform and the Oculus Rift headset. Refer to [Avila and Bailey 2014] for details of the devices. The Virtuix Omni tracks the current location of the user in the virtual environment. The Oculus Rift tracks the position and orientation of the user's head (camera), from which we can approximately compute the user's current position of visual attention. We collect the visual attention data from a number of users, and visualize the visual attention pattern by overlaying a heatmap over the virtual environment.

Virtual Environment. We construct the 3D model of the Kenmore subway station in Boston to conduct our experiments. Figure 3 shows some screenshots. The subway station has about 30 to 40 directional signage which tells people the correct directions to walk to a platform of the subway station or the exit of subway station. We choose to conduct our experiments in this station because this is a place where most pedestrians navigate with a particular goal (e.g. to take a train at a certain platform) in mind, and their navigation decision heavily relies on the instructions given by the directional signage.

We manually build the 3D model of the station. First, we do a visual inspection of the station, taking pictures and marking the locations of all the directional signage. Based on the station's indoor map, we build the 3D layout of the station, and add in all the realistic details such as chairs, trash bins, ticket machines and gates. All the 3D objects are textured similar to their real-world counterparts. We put directional signage in the virtual station according to their exact locations in the real station, so that we can study their effects on pedestrians' navigation.

Directional Signage. In a scene, there are different types of directional signage which hint the user about how to navigate to the exit of the subway station. In particular, we study the effects of *textual signs* and *arrows*, which are among the most common directional signage being used. *Textual signs* are words that hint about where a direction leads to, e.g. "To Exit". *Arrows*, which usually appear on the wall or on the ground, hint about where the pedestrian should go for a particular destination.

Measuring Visual Attention. We collect the movement trajectory and head posture of each user. We then compute the visual attention received by different regions of the virtual environment by ray-casting. The region hit by a ray receives visual attention, with an intensity proportional to the length of time span of the attention. To account for visual acuity, we consider objects within a

Figure 3: *The virtual Kenmore subway station showing different directional signs.*

conic range with the ray being at the central, and the intensity of visual attention drops following a Gaussian distribution. We visualize visual attention by overlaying a heatmap of its intensity onto the virtual environment.

4 Conclusion

We proposed a novel virtual reality platform for conducting visual attention experiments. In particular, we discussed how such platform can be used to study how human visual attention and navigation behavior is guided by the directional signage in the environment. We believe this platform is very useful for conducting many other visual attention experiments. In future, we will also incorporate an eye tracking device on our virtual reality platform to obtain more accurate positions of visual attention.

References

AVILA, L., AND BAILEY, M. 2014. Virtual reality for the masses. *IEEE computer graphics and applications*, 5, 103–104.

HWANG, A. D., HIGGINS, E. C., AND POMPLUN, M. 2009. A model of top-down attentional control during visual search in complex scenes. *Journal of Vision 9*, 5, 25.

HWANG, A. D., WANG, H.-C., AND POMPLUN, M. 2011. Semantic guidance of eye movements in real-world scenes. *Vision research 51*, 10, 1192–1205.

MAW, N. N., AND POMPLUN, M. 2004. Studying human face recognition with the gaze-contingent window technique. In *CogSci*.

WANG, H.-C., AND POMPLUN, M. 2012. The attraction of visual attention to texts in real-world scenes. *Journal of Vision 12*, 6, 26.

WANG, H.-C., LU, S., LIM, J.-H., AND POMPLUN, M. 2012. Visual attention is attracted by text features even in scenes without text. In *CogSci*.

WANG, H.-C., LANDA, Y., FALLON, M., AND TELLER, S. 2014. Spatially prioritized and persistent text detection and decoding. In *Camera-Based Document Analysis and Recognition*. Springer, 3–17.

WU, C.-C., WANG, H.-C., AND POMPLUN, M. 2014. The roles of scene gist and spatial dependency among objects in the semantic guidance of attention in real-world scenes. *Vision research 105*, 10–20.

YU, L.-F., YEUNG, S. K., TANG, C.-K., TERZOPOULOS, D., CHAN, T. F., AND OSHER, S. 2011. Make it home: automatic optimization of furniture arrangement. *ACM Trans. Graph. 30*, 4, 86.

YU, L.-F., YEUNG, S.-K., AND TERZOPOULOS, D. 2016. The clutterpalette: An interactive tool for detailing indoor scenes. *IEEE transactions on visualization and computer graphics 22*, 2, 1138–1148.