Gaze Analysis and Prediction in Virtual Reality

Zhiming Hu*
Peking University

Abstract

In virtual reality (VR) systems, users' gaze information has gained importance in recent years. It can be applied to many aspects, including VR content design, eye-movement based interaction, gaze-contingent rendering, etc. In this context, it becomes increasingly important to understand users' gaze behaviors in virtual reality and to predict users' gaze positions. This paper presents research in gaze behavior analysis and gaze position prediction in virtual reality. Specifically, this paper focuses on static virtual scenes and dynamic virtual scenes under free-viewing conditions. Users' gaze data in virtual scenes are collected and statistical analysis is performed on the recorded data. The analysis reveals that users' gaze positions are correlated with their head rotation velocities and the salient regions of the content. In dynamic scenes, users' gaze positions also have strong correlations with the positions of dynamic objects. A data-driven eye-head coordination model is proposed for realtime gaze prediction in static scenes and a CNN-based model is derived for predicting gaze positions in dynamic scenes.

Index Terms: Gaze analysis—Gaze prediction—Eye tracking; Visual saliency—saliency prediction—Convolutional neural network (CNN)—Virtual reality

1 Introduction

Virtual reality (VR) systems provide users with an opportunity to explore new worlds in an immersive setting. As VR technology develops, users' gaze information in VR becomes increasingly crucial and it has many important applications such as VR content design [9], VR content compression [9], eye movement-based interaction , gaze-contingent rendering [8], etc. Therefore, it is very meaningful to understand users' gaze behaviors in VR and to predict users' gaze positions. At present, the most commonly used solution for eye tracking is to utilize a hardware called an eye tracker. Eye trackers are currently expensive and are not widely used. In addition, eye trackers are mainly designed to calculate users' current gaze direction and cannot predict users' gaze position in the future. Under this background, many researchers focused on developing alternate methods for eye tracking and gaze prediction [9, 10].

Gaze analysis and prediction in virtual reality can be very challenging because users' gaze behaviors may perform differently under different conditions. There exist two mechanisms of visual attention: a bottom-up mechanism and a top-down mechanism [6]. This means human visual attention is influenced not only by the content of the scene, but also by the tasks assigned to them. Therefore, human gaze behaviors under task-oriented conditions may perform differently from that under free-viewing conditions. In addition, users' visual attention in dynamic scenes are more complicated than that in static scenes because moving stimuli will attract users' attention [3]. Considering the research difficulty, this work will be conducted in the following order:

- **Static Scenes under Free-Viewing Conditions**
- **Dynamic Scenes under Free-Viewing Conditions**
- **Task-Oriented Conditions**

2 Related Work

2.1 Gaze Prediction

Gaze prediction is an active area of vision research and there exist many gaze prediction models. In general, most of the models can be classified into top-down models and bottom-up models. Top-down models utilize high-level features of the scene like tasks and context to predict visual attention while bottom-up models focus on low-level image features. In the area of virtual reality, there is limited work on gaze prediction. Sitzmann et al. [9] collected users' eye tracking data in 360° static images and predicted saliency maps of the scenes. Xu et al. [10] analyzed users' gaze behaviors in dynamic 360° videos and proposed a deep learning-based model to predict gaze displacement. Koulieris et al. [7] proposed a classifier to predict gaze object categories in a video game. In contrast with prior works, this research focuses on 3D virtual scenes, including both static and dynamic scenes.

2.2 Gaze Behavior Analysis

Human gaze behaviors have been studied by many researchers. Itti [6] revealed that human gaze behaviors are controlled by 2 mechanisms: a top-down mechanism and a bottom-up mechanism. The 2 mechanisms are independent. Yarbus [11] discovered the coordinated movements between eyes and the head during gaze shifts. Einhäuser et al. [2] further reported that human eye-head coordination exists in natural exploration. Franconeri et al. [3] found that moving stimuli capture visual attention. In the field of virtual reality, Sitzmann et al. [9] found a latitudinal equator bias during observers' exploration of 360° static images. Xu et al. [10] revealed that, in dynamic 360° immersive videos, observers' gaze positions coincide with salient regions and moving stimuli. Inspired by the above-mentioned works, this research also analyzes the influences of head pose, scene content, and dynamic stimuli on gaze behaviors.

3 Current Research

Human gaze behaviors under task-oriented conditions are more complicated than that under free-viewing conditions because users' visual attention will be influenced by the tasks assigned to them. Therefore, for simplicity, this paper only focuses on free-viewing conditions and leave task-oriented conditions for future work. Specifically, gaze behaviors in static scenes and gaze behaviors in dynamic scenes are studied separately and 2 separate models are proposed for gaze position prediction.

3.1 Gaze Analysis and Prediction in Static Scenes

Gaze Data Collection: A total of 60 participants are asked to freely explore 7 static virtual scenes including both indoor and outdoor scenes (Fig. 1, top). During their exploration, the observers' gaze data are recorded by an eye tracker; their head pose data are collected from HTC Vive's head tracking system; and the realtime scenes viewed by the observers are recorded by a screen-recorder.

Gaze Behavior Analysis: Pearson's correlation coefficient is applied to measure the linear relationship between gaze positions and other variables and head rotation velocities are found to have

*e-mail: jimmyhu@pku.edu.cn

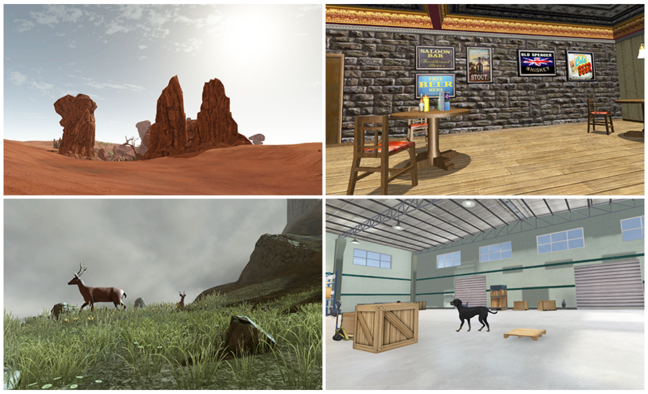

Figure 1: Some of the scenes used in the data collection process. Top: static virtual scenes. Bottom: dynamic virtual scenes. Some animals are randomly placed in the dynamic scenes and are used as dynamic objects.

strong linear correlations with gaze positions. A latency between eye movements and head movements is also observed. The saliency maps of the scenes viewed by the observers are extracted by the state-of-the-art SAM-ResNet saliency predictor [1] and the correlation between gaze position and saliency information is analyzed. The result indicates that the salient regions of the content are also correlated with users' gaze positions.

Gaze Prediction: Based on the above analysis, a data-driven eye-head coordination model is proposed for realtime gaze prediction in static scenes [5]. This model takes users' head pose information and the saliency information of the scenes as input to predict users' realtime gaze positions. This model outperforms the baselines and its effectiveness is validated in practical applications.

3.2 Gaze Analysis and Prediction in Dynamic Scenes

Gaze Data Collection: 43 users in total participate in the data collection process and they are asked to explore 5 dynamic virtual scenes (Fig. 1, bottom) under free-viewing conditions. Some animals like deer, dogs, etc. are randomly placed in the scenes and are utilized as dynamic objects. The animals are allowed to wander in the scene in a random manner. Users' eye tracking data are collected from an eye tracker; their head rotation velocities are obtained from HTC Vive; the scenes viewed by the users are recorded by a screen-recorder; the positions of the dynamic objects are recorded using a Unity script.

Gaze Behavior Analysis: Based on the collected data, statistical analysis is performed to reveal the correlations between users' gaze positions and other factors. The results indicate that the positions of dynamic objects are closely correlated with users' gaze positions and the nearer an object is, the higher correlation it will have. Observers' head rotation velocities are also found to have correlations with their gaze positions. The saliency maps of the scenes observed by the users are calculated using SAM-ResNet and the distributions of gaze positions on salient regions are analyzed. The results reveal that most of the gaze positions lie in the saliet regions of the scene.

Gaze Prediction: On the basis of the analysis, a CNN-based model is derived for predicting gaze positions in dynamic scenes [4]. This model combines the sequence of dynamic object positions, the sequence of head rotation velocities, and the saliency features of virtual scenes to predict gaze positions. This model can be applied to dynamic as well as static scenes. By setting the gaze positions at different time intervals as the model's targets, it can be trained to not only predict relatime gaze positions but also to predict gaze positions in the near future. When an eye tracker is available, this model is also capable of combining users' past gaze data to further improve its prediction performance. This model achieves significant improvement over prior method and its availability is verified in many applications.

4 Future Research

This research aims at analyzing and predicting users' gaze behaviors in virtual reality. However, the current work only focuses on free-viewing conditions and only reveals the influences of head pose, scene content, and dynamic stimuli. Considering the limitations of the current research, there is still plenty of room left for further exploration:

- **Task-Oriented Conditions:** Human gaze behaviors under task-oriented conditions are different from that under free-viewing conditions. Therefore, gaze analysis and prediction under task-oriented conditions deserve to be studied.

- **Other Factors:** Users' gaze behaviors may be influenced by other factors such as sound, users' gestures, users' behavioral habits, users' mental states, etc. These factors need to be explored in detail.

- **Application of Eye Tracking:** The application of eye tracking technology in virtual reality has not been well explored and it will be interesting to develop new applications.

- **Other Systems:** This research only focuses on virtual reality system and it has the potential to be extended to other systems like augmented reality system and mixed reality system.

References

[1] M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara. Predicting Human Eye Fixations via an LSTM-based Saliency Attentive Model. *IEEE Transactions on Image Processing*, 2018.

[2] W. Einhäuser, F. Schumann, S. Bardins, K. Bartl, G. Böning, E. Schneider, and P. König. Human eye-head co-ordination in natural exploration. *Network: Computation in Neural Systems*, 18(3):267–297, 2007.

[3] S. L. Franconeri and D. J. Simons. Moving and looming stimuli capture attention. *Perception & psychophysics*, 65(7):999–1010, 2003.

[4] Z. Hu, S. Li, C. Zhang, K. Yi, G. Wang, and D. Manocha. Dgaze: Cnn-based gaze prediction in dynamic scenes. *IEEE transactions on visualization and computer graphics*, 2020.

[5] Z. Hu, C. Zhang, S. Li, G. Wang, and D. Manocha. Sgaze: A data-driven eye-head coordination model for realtime gaze prediction. *IEEE transactions on visualization and computer graphics*, 25(5):2002–2010, 2019.

[6] L. Itti. *Models of bottom-up and top-down visual attention*. PhD thesis, California Institute of Technology, 2000.

[7] G. A. Koulieris, G. Drettakis, D. Cunningham, and K. Mania. Gaze prediction using machine learning for dynamic stereo manipulation in games. In *2016 IEEE Virtual Reality (VR)*, pp. 113–120. IEEE, 2016.

[8] A. Patney, M. Salvi, J. Kim, A. Kaplanyan, C. Wyman, N. Benty, D. Luebke, and A. Lefohn. Towards foveated rendering for gaze-tracked virtual reality. *ACM Trans. Graph.*, 35(6):179:1–179:12, Nov. 2016.

[9] V. Sitzmann, A. Serrano, A. Pavel, M. Agrawala, D. Gutierrez, B. Masia, and G. Wetzstein. Saliency in vr: How do people explore virtual environments? *IEEE Transactions on Visualization and Computer Graphics (IEEE VR 2018)*, 24(4):1633–1642, April 2018.

[10] Y. Xu, Y. Dong, J. Wu, Z. Sun, Z. Shi, J. Yu, and S. Gao. Gaze prediction in dynamic 360○ immersive videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5333–5342, 2018.

[11] A. Yarbus. Eye movements and vision. 1967. *New York*, 1967.