Cross-Reality Attention Guidance on the Light Field Display

Chuyang Zhang*
Keio University

Kai Kunze†
Keio University

Figure 1: Illustration of our visual attention guiding approach of the cross-reality system. (a) The VR user's view with a gaze visualizer when he focuses on the dragon in the front. (b) The view on the light field display corresponds to (a). (c) The VR user's view when he focuses on the bunny behind. (d) The view on the light field display corresponds to (c).

Abstract

In this work we present a cross-reality collaboration system with visual attention guidance that allows a user in virtual reality (VR) to share his view with the users in the real world. The VR user can remotely select the display content of the light field display oriented to multiple real-world users by manipulating the camera in the virtual world. Using a head-mounted display (HMD) integrated with an eye tracker, the VR user's gaze focus depth is estimated and then used to adjust the focal plane of the light field display. The dioptric blur on the areas out of the focal plane is used as a stimulus to naturally guide real-world viewers to where the VR user is currently focusing on in real time. In a high-quality and low-cost way, our system can improve collaboration in multi-user scenarios especially in which one host user is delivering information to other users such as in classrooms and museums.

1 Introduction

Cross-Reality (CR), a new paradigm that encourages users to freely explore interaction spaces of different immersiveness levels, makes it possible to balance the trade-off between perceived quality and system costs especially in many-user scenarios. To this end, several key challenges need to be tackled, including the seamless transition between different immersiveness levels, and context-aware interaction and collaboration as users may work through multiple systems. Since CR enables users to inter-operate between different types of display [2], reproducing a virtual environment in an AR context has been actively researched as a solution to provide CR experiences. Compared with near-eye displays, a light field display has a desktop-based set-up and is thus an alternative to present a virtual environment to many viewers simultaneously. It renders a coherent sequence of views to different viewing angles so people will get different views with depth cues from different viewing positions, which provides better visual quality than conventional desktop or handheld displays. However, for such light field displays it is difficult to give visual guidance to viewers due to their different perspectives. Traditional unnoticed attention guiding methods [1] (e.g., image filtering, image modulation) are mostly image-based methods in the screen space and could give inaccurate results when applied to light field displays. Besides, it is also difficult to be applied in multi-user scenarios as an imperceptible stimulus for one viewer could be noticeable for others depending on their positions and concentrations. Although overt guidance stimuli (e.g., arrows, cursors) appearing as a 3D object in the scene are globally accurate, such stimuli unrelated to the visual context will destroy the immersive transition of the CR system. Hence, we propose a real-time indirect attention guidance method for light field displays, exploiting the dioptric blur of the light field display as a guiding stimulus. When the VR user focuses at a certain depth in the scene, the display will update the focal plane so that the other region will be blurred naturally, guiding real-world viewers looking to the clear area where the VR user is currently focusing on without disturbing the transition between different immersiveness levels.

*e-mail: chuyang@kmd.keio.ac.jp
†e-mail: kai@kmd.keio.ac.jp

2 Our Approach

Our approach uses two separate subsystems rendering 3D content for the VR user and the real-world viewers respectively, as depicted in Figure 2. In the virtual environment, the user is free to manipulate the virtual camera's transformation. An eye-tracker integrated in the HMD headset estimates the user's gaze focus depth. The gaze is then input into a cross-reality re-mapper together with the virtual camera's transformation information. The re-mapper calculates the correct focal plane distance and passes it to the light field display. It also synchronizes the camera perspective by remapping the virtual camera's transformation in the VR system into the multi-view camera in the light field rendering system. In other words, the virtual camera in VR is a representation of the multi-view camera in the light-field rendering system enabling the VR user to manipulate the light field perspective remotely, whilst the VR user can also preview the overall perspective locally in the virtual environment. Compared with previous systems mapping virtuality to reality [2], our system provides a stereoscopic 3D view in a low-cost way and does not need users from both sides to be colocated.

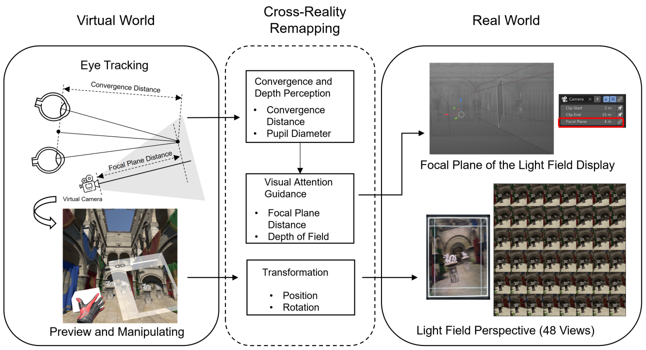

Figure 2: Overview of our system.

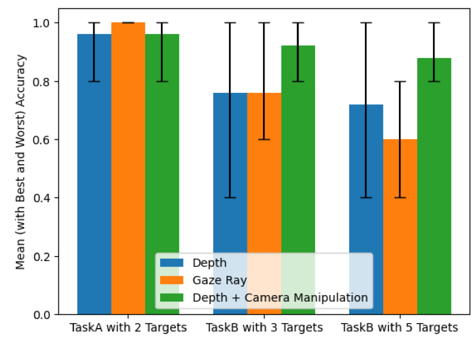

Figure 3: Result of the pilot study. The error bars represent the range of accuracies among the recognition to 5 user's focusing of the rounds

2.1 Indirect Attention Guidance

With the user's gaze focus depth in the VR scene, the focal plane of the light field display can be easily adjusted to the corresponding position in the light field system. However, it is difficult to continuously update the changed focal plane on the light field display due to performance issues. Therefore, we propose to discretely refresh the light field display at a variable rate with a custom model inspired by the depth of field (DoF). When the updated focus distance is still within the simulated DoF of the last frame, the re-mapper will neglect the updated focus distance and will not send updating commands to the light field renderer. Liu et al. propose a computational model by substituting human eye measurements into the classic DoF equation to simulate dynamic DoF with run time pupil diameter as an input [3]. We rewrite their model to make it controllable by replacing the pupil diameter d with d to the power of k:

$$D_F = \frac{497.29 D_S}{497.29 + \frac{3.23 - 0.14 D_s}{d^k}} \tag{1}$$

$$D_N = \frac{497.29 D_S}{497.29 - \frac{3.23 - 0.14 D_s}{d^k}} \tag{2}$$

Where D_F is the depth of field far limit, D_N is the near limit, D_S is the focal plane distance, d is the run time pupil diameter, and k is a value representing the depth filtering level. When k takes a larger value the far limit gets further and the near limit gets closer, which allows for spatially filtering out the unnecessary focus changes caused by the noise of the tracked convergence distance. Large k values allow us to use a smaller fixation time threshold for temporally filtering the tracking noise which reduces the latency for focal plane updating, while small k values can be selected when a smaller discrimination threshold is required for multiple targets that are close to each other in the scene. Note that the simulated depth of field only controls the focus updating and will not affect the displaying content of the light field display.

3 Pilot User Study

We conducted a pilot user study with 5 participants (4 males and 1 female, aged from 23 to 30) to evaluate the feasibility of our system. Participants are asked to wear an HMD integrated with an eye tracker and to complete 3 tasks. In each task, there are different numbers of targets set in a column in front of the subject, and these targets will be alternately highlighted in VR according to a randomly generated number sequence. It is noteworthy that the light field system will present the synchronized VR perspective but will not highlight the targets. Each task includes 3 rounds in which the subjects need to gaze at the sole highlighted target as it changes. In the first and second rounds, the focal plane of the light field display will be respectively updated according to the gaze focus depth and the hit position of the gaze ray, and the subjects are not allowed to manipulate the virtual camera in VR. An observer who cannot view the VR scene recognizes and records which target is highlighted and being focused by the subject in VR according to the focal plane of the light field display. In the third round, with focus depth used as a guide, the subjects are allowed to manipulate the camera to present the highlighted target to the observer.

We have analyzed the observer's recognition accuracy of each round by comparing the recorded sequences with pre-generated number sequences (ground truth). Figure 3 shows the mean accuracy of recognition by the observer for 5 subjects completing 3 rounds in each task, with each error bar indicating the range between the best and worst accuracy observed among the 5 subjects for each round of the task. The result provides an indication that, with the increasing of the targets, our approach using focus depth to guide the observer's attention achieves better accuracy than using gaze rays. Additionally, the accuracy is better when the VR user manipulates the camera to auxiliarily present the focused target to real-world viewers.

4 Conclusion & Future Work

In this work, we present a cross-reality collaboration system for a VR user to share his immersive environment through a light field display. We also propose a gaze-contingent visual attention guidance approach that can be applied to light field displays without disturbing the transition between different immersion levels. To address the system latency, we plan to develop customized light field rendering components capable of streaming the VR scene to the light field display fully seamlessly and real-timely. In the future, we will also conduct a comprehensive user study to investigate how it improves collaboration in many-user scenarios.

Acknowledgments

This work was supported by JST Moonshot R&D Program "Cybernetic being" Project (Grant number JPMJMS2013).

References

[1] R. Bailey, A. McNamara, N. Sudarsanam, and C. Grimm. Subtle gaze direction. *ACM Transactions on Graphics*, 28(4):1–14, Aug 2009. doi: 10.1145/1559755.1559757

[2] U. Gruenefeld, J. Auda, F. Mathis, S. Schneegass, M. Khamis, J. Gugenheimer, and S. Mayer. Vrception: Rapid prototyping of cross-reality systems in virtual reality. In *CHI Conference on Human Factors in Computing Systems*, p. 1–15. ACM, New Orleans LA USA, Apr 2022. doi: 10.1145/3491102.3501821

[3] J. Liu, C. Mantel, and S. Forchhammer. Perception-driven hybrid foveated depth of field rendering for head-mounted displays. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, p. 1–10. IEEE, Bari, Italy, Oct 2021. doi: 10.1109/ISMAR52148.2021.00014