Multi-user Gaze Tracking via Dynamic Image Mapping in 360 Immersive 3D Visualization Systems

Lingyan Ruan
University of Melbourne
Melbourne, Australia
lingyan.ruan@unimelb.edu.au

Bin Chen
University of Melbourne
Melbourne, Australia
bin.chen@unimelb.edu.au

Chi-Hang Chow
City University of Hong Kong
Hong Kong, China
codeycc2@gmail.com

Hiu-Lok Chung
City University of Hong Kong
Hong Kong, China
hlchung24@cityu.edu.hk

Pok-Yin Victor Leung
City University of Hong Kong
Hong Kong, China
victor.leung@cityu.edu.hk

Miu-Ling Lam[*]
School of Creative Media
City University of Hong Kong
Hong Kong, China
miu.lam@cityu.edu.hk

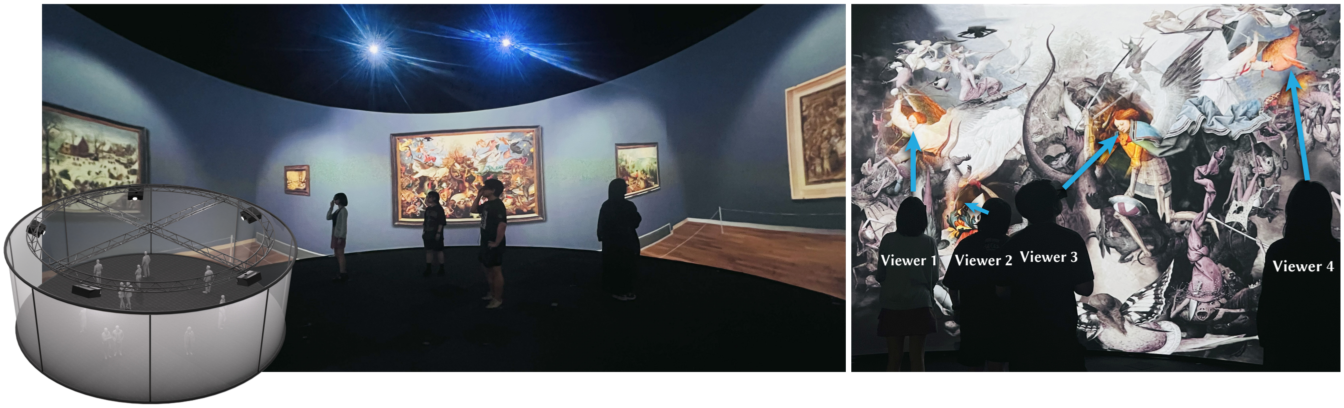

Figure 1: Left: We demonstrate gaze-tracking-based interaction in an art museum context within a large-scale, immersive 360° 3D visualization system, allowing users to move freely within the environment. Gaze tracking on the large display is powered by our custom-developed system. Right: Figurative elements in the artwork are highlighted and animated when users maintain their gaze fixations beyond a predefined dwell-time threshold. Our current prototype supports up to four users simultaneously without noticeable latency. Note that we demonstrate the stereo content in the supplementary video.

Abstract

Immersive panoramic 3D visualization systems enhance presence, enabling rich educational, cultural, and collaborative experiences. While gaze offers significant potential as a natural, hands-free interaction modality in such environments, existing solutions often depend on fixed camera setups, fiducial markers, or controller-based navigation, which limit user mobility, reduce gaze estimation accuracy, and introduce visual or physical distractions. We present a real-time, multi-user gaze-driven interaction system for panoramic 3D visualization. The system features a customized unit integrating binocular eye tracking, active stereoscopic glasses, and a scene camera, paired with a portable Android device. The scene camera enables precise gaze mapping onto panoramic 3D content through feature-based alignment with the 360° footage, while the Android device handles real-time gaze estimation and synchronization to ensure low-latency interaction. This approach delivers accurate, responsive gaze mapping for multiple freely moving users without the need for additional tracking infrastructure. Two interactive museum applications demonstrate the system's capability to support fluid, collaborative, and unobtrusive gaze-based interaction in dynamic cultural and educational installations.

CCS Concepts

• **Hardware → Emerging tools and methodologies**; • **Human-centered computing → Systems and tools for interaction design**; *Displays and imagers.*

ACM Reference Format:
Lingyan Ruan, Bin Chen, Chi-Hang Chow, Hiu-Lok Chung, Pok-Yin Victor Leung, and Miu-Ling Lam. 2025. Multi-user Gaze Tracking via Dynamic

---

[*] Corresponding author

SA Technical Communications '25, Hong Kong, Hong Kong

ACM ISBN 979-8-4007-2136-6/25/12
https://doi.org/10.1145/3757376.3771419

Image Mapping in 360 Immersive 3D Visualization Systems. In *SIGGRAPH Asia 2025 Technical Communications (SA Technical Communications '25), December 15–18, 2025, Hong Kong, Hong Kong*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3757376.3771419

1 Introduction

Immersive 360° displays are widely utilized in museums, science centers, and cultural installations, where they present high-resolution content that enhances presence and promotes spatial engagement. Yet, interaction with such displays remains challenging due to their large scale. Previous approaches have employed handheld devices such as Flysticks [Tcha-Tokey et al. 2017] and infrared laser pointers [Benko and Wilson 2010], but these often disrupt immersion. Another line of work [McGinity et al. 2007] has explored non-invasive tracking with infrared camera arrays that reconstruct voxel-based representations of participants. While effective for capturing gross body movements, the relatively low resolution (about 4.2 cm per voxel) limits precise interaction. Gaze offers a natural and intuitive modality, widely studied in head-mounted displays and large public displays, but rarely applied to fully immersive 360° environments. Such spaces introduce additional challenges: they are typically dark, requiring robust low-light tracking; interactions must be independent of user position and head orientation; systems must accommodate multiple simultaneous users with minimal latency; and they should avoid reliance on visual markers.

To address these challenges, we present a novel active gaze-based interaction system that enables multiple users to interact with 360° content while freely moving and rotating their heads in low-light conditions. The system integrates an infrared eye tracker, a wide field of view scene camera, and a feature matching algorithm that continuously aligns the displayed 360° video with the scene camera feed to map gaze accurately onto the content. All components including the eye tracker, scene camera, and lightweight active shutter 3D glasses are mounted on a custom frame designed for ergonomic comfort. A portable processing unit performs gaze estimation, calibration, and real time processing without requiring markers or additional devices.

To evaluate the effectiveness and robustness of the proposed system, we implemented two application scenarios. In the first, users activated character animations within a panoramic painting by dwelling on specific visual elements. In the second, we developed a collaborative, user-aware narrative generation system in which an AI model produced descriptive content based on objects simultaneously attended to by multiple users. These scenarios demonstrate how active gaze-based interaction can broaden the expressive capacity and foster collaborative engagement in immersive 360° displays. Please refer to the project page for more details: https://sweb.cityu.edu.hk/miullam/GazeTrack360.

2 Related Work

2.1 Interaction Techniques for 360° Displays

Large scale 360° immersive displays such as CAVE [Cruz-Neira et al. 2023], CAVE2 [Febretti et al. 2013], and AVIE [McGinity et al. 2007] provide expansive visual coverage and a heightened sense of presence, offering advantages beyond conventional displays [Juan and Pérez 2009]. Their vast scale and immersive field define distinct interaction paradigms that pose unique challenges for perception, navigation, and control. Early methods relied on handheld devices such as laser pointers, with computer vision algorithms tracking the projected spot to enable position independent pointing and selection on panoramic content. Markerless full body tracking, exemplified by the AVIE 360° stereoscopic theater [McGinity et al. 2007], used infrared camera arrays to reconstruct volumetric user representations for real time gesture and positional input, supporting multi user and equipment free interaction in cultural and narrative contexts. Other approaches include tracked handheld controllers such as the flystick [Tcha-Tokey et al. 2017], which allow pointing, selection, and manipulation, and Kinect based hand gesture recognition [Leite et al. 2017], where static gestures are mapped to commands for object manipulation and navigation. Despite these developments, gaze, often regarded as a natural and unobtrusive interaction modality, has rarely been explored in this setting. This work addresses this gap by introducing a gaze based, marker free, multi user interaction framework for large scale immersive displays.

2.2 Gaze-Based Interaction with Large Scale Displays

Gaze interaction on large public displays such as wall-sized monitors and projection systems faces challenges similar to those in immersive 360° environments. Existing approaches can be broadly categorized into remote (fixed-camera) and wearable eye-tracking systems. The former relies on vision-based methods that estimate gaze direction using fixed cameras placed in front of participants, while the latter employs wearable devices that capture gaze from an egocentric perspective. Since our method is more closely related to the latter, this section focuses on prior work in that category.

Wearable eye trackers (e.g., Tobii Pro Glasses, Pupil Labs) provide higher accuracy after a one-time calibration and support more expressive interactions. For example, GazeProjector [Lander et al. 2015] integrated head-mounted eye trackers with an auxiliary camera to achieve seamless gaze mapping across multiple displays. Koch et al. [Koch et al. 2025] combined Pupil Labs Invisible glasses, projection mapping, and object detection to visualize multiple users' gaze points in real time on a tabletop projection surface, enabling intuitive collaboration and shared visual attention during group tasks, but the system requires physical markers on the tabletop surface. SocialEyes [Saxena et al. 2025] introduces a system for synchronizing, mapping, and analyzing multi-user mobile eye-tracking data in real-world social environments such as concerts, classrooms, and film screenings. However, its setup confined seated participants with limited movement. Building on this line of work with wearable eye trackers, we employ a lightweight customized system featuring an integrated scene camera and active shutter, optimized for high accuracy in low-light, large-scale 360° immersive environments. This configuration enables real-time gaze-based interaction, supports multi-user collaboration, and allows participants to move freely while engaging with 3D visualizations on a cylindrical display without the need for fiducial markers.

3 System

Our system comprises customized hardware and software, as shown in Fig. 2. A wearable eye-tracking device with an integrated scene camera enables stable gaze estimation, which is mapped to the

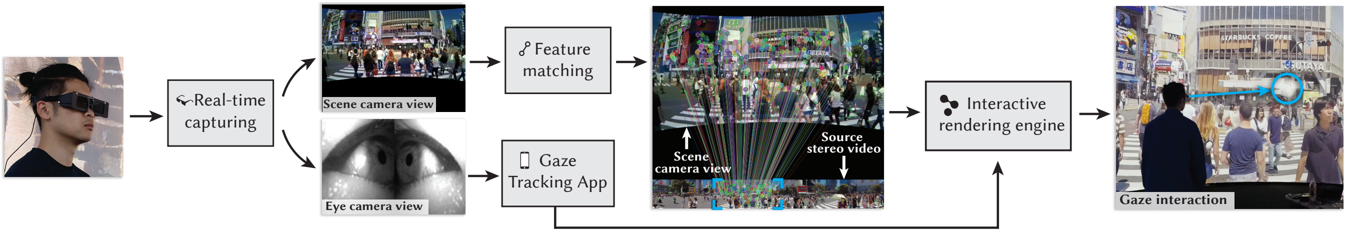

Figure 2: The flowchart of the gaze-based interaction system. Users wear a customized eye-tracking device with an integrated scene camera, enabling stable gaze estimation within the scene view. The estimated gaze is then mapped onto the cylindrical 3D visualization display using a feature-matching algorithm, enabling interactive rendering based on the user's gaze position.

cylindrical 3D display via a feature-matching algorithm for real-time interactive rendering.

3.1 Hardware Setup

Our system consists of cylindrical-shape stereoscopic display (see Fig. 1) and stereo 3D glasses equipped with an eye-tracking module, and portable Android devices, as shown in Fig. 3. The display is similar to that described in [McGinity et al. 2007] but measures 9.5 m in diameter and 4 m in height, illuminated by five Barco F80 projectors strategically arranged to uniformly cover the cylindrical surface, delivering a resolution of up to 12800 × 3200. Stereoscopic disparity is achieved using active shutter technology, which alternately blocks and transmits images to each eye, enabling audiences to perceive 3D content without relying on polarization. We employ Volfoni Edge RF active-shutter glasses synchronized with the visualization system, into which the Pupil Labs Neon eye-tracking module [Baumann and Dierkes 2023] is integrated. This module includes two infrared eye cameras and a wide field-of-view RGB scene camera capturing contextual video at 60 Hz with a resolution of 1600 × 1200 pixels and a field of view of 103° × 77°. All components, including the eye-tracking module and cameras, are embedded within a lightweight custom 3D-printed thermoplastic frame (Fig. 3). The gaze processing device is an Android-based platform (Moto Edge 40 Pro with Qualcomm Snapdragon) that wirelessly handles real-time gaze data processing and synchronization.

4 Image Processing

The software architecture consists of three layers designed for real-time performance and scalability. The Gaze Data Processing Layer captures raw eye images using infrared cameras and estimates gaze direction through pupil detection and corneal reflection analysis, while the RGB scene camera records the user's first-person view for gaze mapping. The Mapping Layer employs XFeat [Potje et al. 2024], a lightweight feature extraction and matching model, to detect keypoints across frames and computes a homography matrix via OpenCV to transform gaze coordinates from the scene camera to the display surface (see Fig. 2). Calibration is performed using the Neon Companion application (Pupil Labs) through a brief one-time procedure lasting under 30 seconds to ensure alignment between gaze data and displayed content. The Visualization and Interaction Layer leverages TouchDesigner [Derivative Inc. 2023] for real-time rendering of interactive panoramic content and integrates a synchronization module that merges gaze inputs from multiple users, enabling collaborative interaction. Our feature matching approach works well with stereoscopic video content, as shown in the supplementary material, by overlapping the stereo pair with a disparity-based offset and using the result as the reference image.

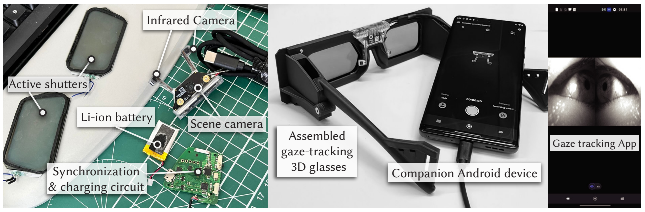

Figure 3: We customized a lightweight pair of 3D glasses by integrating a wearable eye tracker, active shutters, a scene camera, and battery-powered synchronized charging circuits. Gaze estimation and synchronization with the display are performed on an Android device.

5 Interaction

In this section, we demonstrate two gaze-based interaction applications on the display, both employing dwell-time as the primary input modality, where sustained fixation on a target for a predefined duration triggers interaction. The applications are developed on *The Fall of the Rebel Angels* [1] , from the Royal Museums of Fine Arts of Belgium, digitized by the Google Cultural Institute.

In the first application, fixation on a figurative element beyond the dwell-time threshold highlights the object and triggers an animation, as shown in Fig. 1. The system supports concurrent interactions and is demonstrated with up to four participants. The second application explores collaborative interaction augmented with generative AI. When multiple figurative elements are jointly selected, the system leverages a large language model (ChatGPT) to generate narrative text contextualizing the chosen objects. This functionality is demonstrated with two participants in the supplementary video. It should be noted that, although the demonstration showcases four users, the system is capable of accommodating a larger number of participants, as it supports simultaneous multi-user interaction without noticeable latency. We evaluated the processing times for gaze estimation, feature matching, and display remapping on a Supermicro X13DEG-QT workstation running Windows 11. The system achieved an average latency of approximately 80 ms, which

---

[1] Available at: https://artsandculture.google.com/story/the-fall-of-the-rebel-angels-royal-museums-of-fine-arts-of-belgium/9gXx-oPTgMeLKg?hl=en

remained stable with up to four concurrent users. This efficiency is attributed to the use of batch processing and feature matching on downsampled inputs (640 × 480), identified as the minimum resolution preserving high matching accuracy. These results indicate that the system maintains real-time responsiveness and can scale effectively to support additional users beyond those tested.

6 Application Scenarios

In this section, we outline several potential application scenarios that demonstrate the versatility and practicality of our system.

6.1 Scientific Collaboration and Data Analysis

One important application of our system is collaborative exploration of complex scientific data. In this scenario, multiple researchers engage with large-scale panoramic visualizations, such as volumetric climate models, molecular structures, or astrophysical simulations. Unlike traditional setups that rely on laser pointers or handheld devices, our system enables researchers to highlight and annotate regions of interest through natural gaze-based interaction. The ability to synchronize gaze points across users facilitates shared attention and accelerates group decision-making. This approach is particularly valuable in multidisciplinary teams, where rapid consensus on complex visual information is critical.

6.2 Gaze Communication in Telepresence

Remote collaboration frequently suffers from a lack of gaze awareness, which makes it difficult for participants to interpret one another's attention, intention, and focus. Our system has the potential to address this limitation by integrating real-time gaze tracking with gaze correction techniques [Kononenko and Lempitsky 2015; Zhang et al. 2022] to restore mutual gaze perception. This capability can facilitate more natural eye contact, strengthen conversational grounding, and reduce ambiguity in distributed teamwork.

7 Discussion

Our prototype demonstrates the feasibility of real-time gaze interaction in large-scale 360° immersive displays, currently relying on dwell-time selection as the primary input modality. Although we reported system latency for up to four simultaneous users, a more comprehensive evaluation, including controlled user studies, is required to assess usability, comfort, and engagement in such expansive environments. The collaborative generative AI demonstration, where multiple users jointly selected objects to trigger AI-generated narrative, remains an early proof of concept. While it illustrates potential of combining gaze interaction with AI-driven content, systematic evaluation is needed to better understand its effectiveness and impact on collaboration. Future work will extend interaction beyond dwell-time and conduct controlled studies to examine scalability and integration with advanced AI applications.

8 Conclusion

Our system presents a lightweight and scalable solution for real-time, marker-free gaze interaction within large-scale immersive 360° environments. By integrating stereoscopic visualization with eye tracking and efficient mobile-based processing, it overcomes the limitations of conventional interaction methods that often rely on handheld devices. The architecture has been demonstrated through both individual and collaborative applications, including interactive engagement with cultural heritage content and AI-augmented multi-user experiences. Evaluation results confirm that the system maintains low-latency responsiveness even with multiple simultaneous users, supporting natural and intuitive gaze-based interaction. These findings highlight the potential of our approach to enhance education, training, cultural experiences, and entertainment in large-scale immersive displays, advancing gaze-aware interaction beyond traditional desktop or headset-based systems.

Acknowledgments

This work was supported in part by a grant from the Innovation and Technology Commission of the Hong Kong SAR, China (ITS/326/21FP).

References

Chris Baumann and Kai Dierkes. 2023. Neon accuracy test report. *Pupil Labs* 10 (2023).

Hrvoje Benko and Andrew D Wilson. 2010. Multi-point interactions with immersive omnidirectional visualizations in a dome. In *ACM International Conference on Interactive Tabletops and Surfaces*. 19–28.

Carolina Cruz-Neira, Daniel J Sandin, and Thomas A DeFanti. 2023. Surround-screen projection-based virtual reality: the design and implementation of the CAVE. In *Seminal Graphics Papers: Pushing the Boundaries*. 51–58.

Derivative Inc. 2023. TouchDesigner. https://derivative.ca/. Accessed: 2025-08-17.

Alessandro Febretti, Arthur Nishimoto, Terrance Thigpen, Jonas Talandis, Lance Long, JD Pirtle, Tom Peterka, Alan Verlo, Maxine Brown, Dana Plepys, et al. 2013. CAVE2: a hybrid reality environment for immersive simulation and information analysis. In *The Engineering Reality of Virtual Reality 2013*, Vol. 8649. SPIE, 9–20.

M Carmen Juan and David Pérez. 2009. Comparison of the levels of presence and anxiety in an acrophobic environment viewed via HMD or CAVE. *Presence: Teleoperators and virtual environments* 18, 3 (2009), 232–248.

Maurice Koch, Tobias Rau, Vladimir Mikheev, Seyda Öney, Michael Becher, Xiangyu Wang, Nelusa Pathmanathan, Patrick Gralka, Daniel Weiskopf, and Kuno Kurzhals. 2025. Group Gaze-Sharing with Projection Displays. Article 92.

Daniil Kononenko and Victor Lempitsky. 2015. Learning To Look Up: Realtime Monocular Gaze Correction Using Machine Learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Christian Lander, Sven Gehring, Antonio Krüger, Sebastian Boring, and Andreas Bulling. 2015. Gazeprojector: Accurate gaze estimation and seamless gaze interaction across multiple displays. In *Proceedings of the 28th annual ACM Symposium on User Interface Software & Technology*. 395–404.

Diego Q Leite, Julio C Duarte, Luiz P Neves, Jauvane C De Oliveira, and Gilson A Giraldi. 2017. Hand gesture recognition from depth and infrared Kinect data for CAVE applications interaction. *Multimedia Tools and Applications* 76, 20 (2017), 20423–20455.

Matthew McGinity, Jeffrey Shaw, Volker Kuchelmeister, Ardrian Hardjono, and Dennis Del Favero. 2007. AVIE: a versatile multi-user stereo 360 interactive VR theatre. In *Proceedings of the 2007 workshop on Emerging Displays Technologies: Images and Beyond: the Future of Displays and Interaction*. 2–es.

Guilherme Potje, Felipe Cadar, André Araujo, Renato Martins, and Erickson R Nascimento. 2024. Xfeat: Accelerated features for lightweight image matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2682–2691.

Shreshth Saxena, Areez Visram, Neil Lobo, Zahid Mirza, Mehak Khan, Biranugan Pirabaharan, Alexander Nguyen, and Lauren K Fink. 2025. SocialEyes: Scaling Mobile Eye-tracking to Multi-person Social Settings. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Article 751.

Katy Tcha-Tokey, Emilie Loup-Escande, Olivier Christmann, and Simon Richir. 2017. Effects on user experience in an edutainment virtual environment: comparison between CAVE and HMD. In *Proceedings of the European Conference on Cognitive Ergonomics*. 1–8.

Jichao Zhang, Jingjing Chen, Hao Tang, Enver Sangineto, Peng Wu, Yan Yan, Nicu Sebe, and Wei Wang. 2022. Unsupervised high-resolution portrait gaze correction and animation. *IEEE Transactions on Image Processing* 31 (2022), 5272–5286.