Towards Mixed Reality AI Docents: Egocentric Smart Glasses with Vision and LLM Interaction

Jongyoon Lim*
Sogang University

Jusub Kim†
Sogang University

Sangyong Kim‡
Sogang University

Yongsoon Choi§
Sogang University

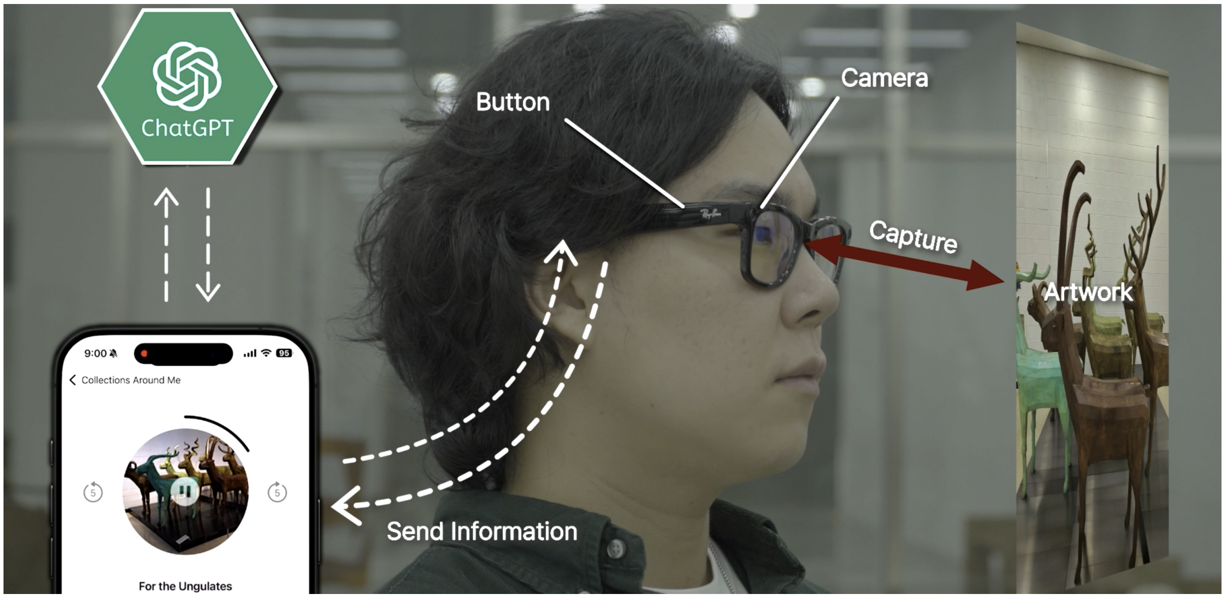

Figure 1: Overall system architecture of the proposed AI docent framework.

ABSTRACT

We present a novel AI-driven docent framework as a foundational step toward Mixed Reality (MR) museum interaction. Traditional docent systems, such as QR-based or audio devices, offer limited, one-way experiences that require active device handling. In contrast, our system enables egocentric and bidirectional interaction using smart glasses equipped with a camera, microphone, and speaker, integrated with BLE-based localization, image recognition, and large language models (LLMs) for contextual dialogue. This demonstration implements and evaluates two core components of our envisioned MR docent system: spatial awareness and conversational intelligence. While augmented reality (AR) display and visual wayfinding remain future directions, the current prototype supports hands-free, context-aware, and personalized interaction with nearby artworks while minimizing smartphone operation. We conducted a preliminary user study with 16 participants at a museum in Seoul, Korea. Results indicated high user satisfaction, particularly in the accuracy of image-based recognition, the usefulness of AI explanations, and the seamless integration of spatial context. Reported usability ratings averaged 6.5/7 across key metrics such as goal achievement, information clarity, and interface intuitiveness.

Keywords: Mixed Reality, Smart Glasses, AI Docent, BLE Localization, Image Recognition, LLM.

*e-mail: limjy.kor@gmail.com
†e-mail: jusub@sogang.ac.kr
‡e-mail: nicosj@hanmail.net
§e-mail: yongsoon@sogang.ac.kr

1 INTRODUCTION

Modern museums are increasingly integrating digital technologies to enhance visitor engagement, but most existing systems remain limited in interactivity and contextual responsiveness. Traditional solutions such as audio guides or QR-based apps require frequent device handling and offer only passive, one-way information delivery [4]. These limitations reduce immersion and create cognitive friction during the exploration of exhibitions [5].

In response, we propose an egocentric and AI-powered docent framework that minimizes smartphone interaction while enabling spatially contextualized, bidirectional engagement with artworks. Our system leverages smart glasses equipped with a camera, microphone, and speaker to deliver hands-free experiences. It integrates three core modalities: (1) BLE-based localization for identifying nearby artworks [2], (2) image recognition for visual understanding [3], and (3) large language model (LLM)-driven dialogue for real-time conversational responses [4].

This work serves as a modular implementation within a larger national Mixed Reality (MR) initiative for museum storage environments in Korea. While the broader MR vision includes AR display-based wayfinding and spatial augmentation, this study focuses on two foundational pillars: spatial awareness and conversational intelligence. These components collectively reframe the museum docent experience—from static, user-initiated querying to a dynamic, context-aware interaction cycle.

To validate the feasibility and usability of our approach, we conducted a user study with 16 participants at a museum in Seoul. The results revealed strong satisfaction in key areas such as the system's ability to recognize artworks accurately, the relevance of

AI-provided explanations, and the intuitiveness of the egocentric interaction paradigm. These findings support our belief that conversational and location-aware AI agents can meaningfully enhance cultural experiences, setting the stage for future integration of AR-based wayfinding and visual overlays.

2 System Overview

Our system integrates three core components—BLE-based localization, image recognition, and LLM-driven interaction—within a hands-free, egocentric smart glass interface. These components work in tandem to create a seamless flow from artwork detection to conversational explanation, forming the foundation for a future Mixed Reality (MR) museum experience.

BLE-based Localization. Each exhibit is paired with a Bluetooth Low Energy (BLE) beacon. The visitor's smartphone detects the Received Signal Strength Indicator (RSSI) values from nearby beacons to estimate proximity. To improve spatial accuracy, we incorporate fixed exhibit ordering and beacon triangulation along predefined visitor pathways. This localization enables the system to contextually narrow the candidate artworks for recognition and dialogue.

Camera-based Image Recognition. The smart glasses camera captures images of artworks upon user command (e.g., 'Take a picture'), which are then transmitted to the back-end server via a companion smartphone app. To enhance recognition precision, BLE-determined proximity is used to filter the recognition model's search space, minimizing false matches in dense exhibit environments.

LLM-driven Conversational Interaction. Using the microphone built into the smart glasses, visitors can verbally ask questions about the recognized artworks. The audio is transmitted to the smartphone, transcribed into text, and processed by a fine-tuned LLM. The system then generates a spoken response delivered back through the glasses' speaker, enabling real-time, bidirectional, and personalized interaction.

System Integration. The current prototype does not yet incorporate AR-based visual overlays, but the spatial and conversational modalities serve as essential precursors to such functionality. Together, they offer a lightweight form of MR by blending physical proximity, visual cues, and AI-driven context awareness without relying on explicit UI navigation. This modular structure allows for future extension toward AR-based wayfinding and immersive visual feedback.

3 Demonstration Details

The demonstration showcases a full user journey through the museum from a first-person perspective, illustrating how our system enables seamless interaction without reliance on handheld devices. As shown in fig. 1, each step reflects the integration of location, vision, and language technologies through smart glasses.

Proximity Detection via BLE. As the visitor walks through the exhibition space, BLE beacons installed near each artwork are detected by the visitor's smartphone. Based on signal strength and beacon ordering logic, the system infers which artworks are in close proximity and sets the contextual scope for further processing.

Hands-Free Image Capture. When the visitor focuses their gaze on an artwork, they can issue a voice command such as 'Take a picture'. The smart glasses capture an image and transmit it through smartphone to the back-end server. BLE data limits the recognition of candidates, improving speed and precision.

Artwork Recognition. The server uses the incoming image and proximity data to identify the artwork. This hybrid approach increases robustness [2,3], especially in spaces where multiple exhibits are close together.

Conversational Interaction with the AI Docent. After the artwork is identified, the visitor may ask follow-up questions verbally (e.g., "What is the story behind this piece?"). The microphone captures the audio, which is transcribed and sent to a fine-tuned LLM. The response is converted to speech and played through the speaker of the smart glasses.

Personalization and Recommendation (Planned Feature). While the current system logs user behavior such as dwell time and query patterns, personalized recommendation is not yet implemented. Future extensions will explore personalized artwork suggestions and theme-based route planning based on user interests, exploration history, and interaction frequency.

Video Showcase and Evaluation Context. A video recorded in a real museum storage shows user interaction with smart glasses. A study with 16 participants evaluated usability, and feedback informed design improvements. The video illustrates not only the technical process but also the experiential flow of real use [5].

4 Future Work

While the current prototype demonstrates effective integration of spatial sensing and conversational AI, several directions remain to complete the envisioned Mixed Reality docent framework.

AR Display and Visual Wayfinding. Future versions will integrate AR displays into smart glasses, providing overlays such as highlights, navigation, and annotations. This will enhance spatial orientation and enable more intuitive, visually enriched guidance [1].

Interoperability with Broader Museum Networks. We plan to expand compatibility with diverse museum infrastructures by adopting standardized metadata schemas and supporting cross-institutional content retrieval. This will allow the docent system to operate across exhibitions and contribute to a shared cultural knowledge base.

System Integration and Edge Optimization. Currently, the system operates through a loosely coupled architecture involving the smart glasses, smartphone, and cloud services. Future work includes tighter integration via shared APIs and possible edge computing optimization to reduce latency and improve responsiveness during real-time interactions.

Adaptive Personalization. Future work will add user modeling and feedback to refine recommendations and dialogue.

These developments aim to realize a fully embodied MR experience—where spatial intelligence, visual recognition, and conversational AI coalesce through a single, wearable interface.

Acknowledgments

This research was supported by the Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency (KOCCA) grant funded by the Ministry of Culture, Sports and Tourism in 2024 (Project Name: Development of real-virtual linked digital open storage and integrated management system for collections, Project Number: RS-2023-00225305).

References

[1] C. Green. Augmenting heritage: An open-source multiplatform ar application. *arXiv preprint arXiv:2310.13700*, 2023.

[2] R. Mautz. Indoor positioning technologies. Technical report, ETH Zurich, 2012.

[3] L. Meyer et al. Algorithmic ways of seeing: Using object detection to facilitate art exploration. In *Proc. CHI*, 2024.

[4] S. Varitimiadis et al. Implementing an ai chatbot platform for museums. In *Proc. ICCICMS*, 2020.

[5] R. Zhang et al. Immersive technology in hybrid museum exhibitions. *Digital Creativity*, 2024.