3D Point Cloud–Based Digital Twin Reconstruction and VR Immersive Visualization of Borobudur’s Hidden Foot Reliefs

Satoshi Takatori[1], Jiajun Gao[1], Liang Li[1], Fadjar Ibnu Thufail[2], Brahmantara[3], Satoshi Tanaka[1]

[1] Ritsumeikan University, Ibaraki, Japan – stakator@fc.ritsumei.ac.jp, gao42823@gmail.com, liliang@fc.ritsumei.ac.jp, stanaka@is.ritsumei.ac.jp

[2] National Research and Innovation Agency, Jakarta, Indonesia – fadj001@brin.go.id

[3] Indonesian Heritage Agency, Magelang, Indonesia – brahmantara@kemdikbud.go.id

Keywords: 3D Point Clouds, Digital Twin, Virtual Reality, Cultural Heritage, Semantic Segmentation, Level of Detail

Abstract

The Borobudur Temple’s Hidden Foot (Karmawibhangga layer), comprising 160 bas-relief panels sealed behind protective stone cladding since the early 20th century, presents a compelling challenge for digital cultural heritage practice. This paper presents a digital twin reconstruction and immersive virtual reality (VR) system that provides interactive access to this physically inaccessible archaeological space. The system integrates three pre-existing datasets: UAV photogrammetric point clouds of the outer temple structure, deep-learning-based 3D relief reconstructions derived from archival photographs (Pan et al., 2022), and semantic segmentation results of relief surfaces (Ji et al., 2023). Within a Unity 2022.3.22f1-based VR environment optimised for the HTC VIVE Pro Eye head-mounted display, these datasets are unified into a navigable Hidden Foot passage. A novel lighting-driven Level of Detail (LoD) strategy—where rendering complexity is governed by the illumination radius of a virtual torch—achieves 65–70 FPS on the HMD while preserving visual fidelity within the illuminated zone. Eye-gaze-triggered textual annotations support contextual iconographic interpretation, while interactive toggling between photorealistic and semantically labelled views enables both casual exploration and scholarly analysis. A user study ($n = 10$) employing a System Usability Scale-adapted questionnaire yielded mean satisfaction ratings exceeding 4.5/5 across all evaluation dimensions. The system demonstrates a scalable methodology for transforming inaccessible heritage layers into immersive, annotated digital environments.

1. INTRODUCTION

The Borobudur Temple in Central Java, Indonesia, represents one of the most significant Buddhist monuments in the world, known for its sophisticated architectural form and intricately carved narrative reliefs, and inscribed as a UNESCO World Heritage Site since 1991 (UNESCO, 2023). Its lowest tier, the Hidden Foot (Karmawibhangga layer), contains 160 bas-relief panels depicting scenes from the Buddhist law of karma. These panels were deliberately concealed beneath a stone retaining structure during stabilisation work supervised by van Erp (1907–1911), rendering them physically inaccessible to visitors and researchers alike.

The Karmawibhangga panels constitute a unique iconographic programme within Borobudur’s broader narrative cycle. Carved directly onto the base of the monument prior to the addition of the retaining structure, the 160 panels illustrate scenes of transgression, karmic consequence, and spiritual reward, rendered in a naturalistic pictorial vocabulary that is iconographically distinct from the relief programmes of the upper terraces. Their deliberate concealment during the 1907–1911 stabilisation works supervised by van Erp (1907–1911) has paradoxically preserved the stone surface from surface weathering while rendering it entirely inaccessible to contemporary field study. The only surviving pre-enclosure documentation consists of van Erp’s glass-plate photographs (van Erp, 1907–1911), captured under variable lighting conditions and without photogrammetric control, which have served as the sole visual record for over a century of scholarly interpretation.

Traditional documentation relies heavily on archival black-and-white photographs taken before the enclosure, which lack geometric fidelity, metadata, and photogrammetric compatibility. While digital approaches including photogrammetry and LiDAR have substantially advanced documentation of the accessible monument (Kawato et al., 2021; Tanaka et al., 2021), the Hidden Foot layer presents a fundamentally different challenge: the physical space cannot be surveyed by conventional means, and the relief surfaces must be reconstructed from indirect evidence.

This study addresses these challenges by developing a digital twin reconstruction and immersive VR system that integrates multi-source datasets—including high-resolution point clouds, deep-learning-based 3D relief reconstructions (Pan et al., 2022), and semantic segmentation annotations (Ji et al., 2023)—into a navigable virtual passage accessible to both scholars and the general public. The overall system architecture is summarised in Figure 1.

2. RELATED WORK

2.1 3D Documentation in Heritage Contexts

Large-scale 3D documentation through photogrammetry and LiDAR scanning is now standard practice in digital heritage conservation. UAV-based photogrammetry has achieved centimetre-level spatial accuracy at complex monuments in Southeast Asia (Pathak et al., 2024). For Borobudur specifically, Kawato et al. (2021) developed a comprehensive multi-resolution point cloud archive of the visible temple tiers through integrated aerial and ground-level photogrammetric acquisition, and Tanaka et al. (2021) produced fused high-definition visualisations combining multiple sensor modalities. These projects

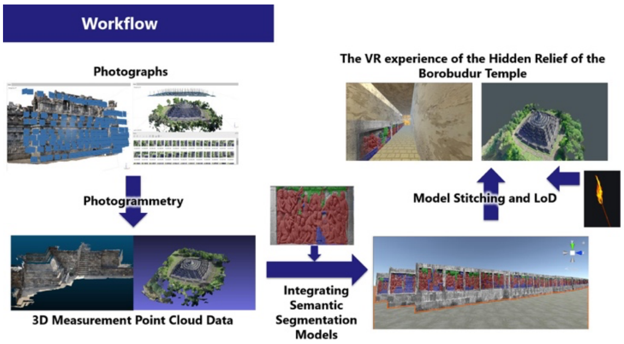

Figure 1. Overall workflow of the digital twin reconstruction and VR visualisation system for Borobudur's Hidden Foot. The pipeline integrates point cloud acquisition, relief reconstruction, and semantic segmentation into a navigable VR environment.

represent the state of the art for the accessible monument structure but are inherently limited to surfaces that can be physically observed or surveyed.

2.2 Deep Learning for Relief Reconstruction and Segmentation

Pan et al. (2022) addressed reconstruction of Borobudur relief geometry from historical monocular photographs using a 2D-to-3D deep learning pipeline with soft-edge constraints, recovering fine sculptural details including human figures, vegetation, and ornamental patterns. Ji et al. (2023) extended this work with semantic segmentation tailored to Borobudur relief characteristics, classifying panel elements into iconographic categories. These two datasets—3D reconstructed meshes and semantic annotations—form the primary visual content of the present system.

2.3 Immersive VR for Cultural Heritage

Virtual Reality has become an established medium for heritage presentation at physically restricted or sensitive archaeological sites. Suzuki et al. (2023) examined locomotion methods specifically in the context of a Borobudur VR environment, providing directly relevant prior work for navigation design. Smith et al. (2019) investigated dynamic lighting and LoD strategies for real-time rendering in complex virtual environments. None of the prior systems, however, addressed the physically inaccessible Hidden Foot layer or combined deep-learning relief reconstruction, semantic overlays, and illumination-driven LoD optimisation into a single immersive system.

2.4 Digital Twin Methodology in Heritage Contexts

The concept of a digital twin—a dynamic, data-linked virtual counterpart of a physical entity—has been progressively adopted in the built heritage domain as a framework for integrating heterogeneous sensor data, simulation models, and knowledge representations into a unified queryable environment. Unlike static 3D reconstructions, digital twins are characterised by their capacity to support analysis, simulation, and decision-making by reflecting the current or historical state of the physical asset.

In civil infrastructure, digital twins have been used to monitor structural health, simulate load distributions, and support maintenance planning. Heritage applications have adapted these principles to address the specific challenges of aged or fragile fabric, incomplete historical records, and the need to simultaneously serve conservation professionals and public audiences. Prominent examples include digital twin frameworks for Gothic cathedral structures, where point-cloud surveys are linked to finite element models for structural assessment, and for archaeological sites, where stratigraphy and material analysis results are registered within a navigable 3D environment.

The present system extends the digital twin concept to an architecturally inaccessible context—the Hidden Foot of Borobudur—where the physical referent cannot itself be surveyed or instrumentally monitored. In this setting, the digital twin is necessarily constructed from surrogate evidence: archival photographic records, deep-learning-based geometric inference, and semantic annotations. The resulting system functions as a digital twin in the sense that it provides an authoritative, interactively queryable representation of an inaccessible physical reality, bridging the gap between archival scholarship and immersive spatial experience without requiring direct physical access.

2.5 Eye-Tracking Interaction in Heritage VR

Eye-tracking-based interaction has emerged as a compelling modality for heritage VR due to its non-intrusive character and its capacity to mirror natural human attentional behaviour. Gaze-contingent rendering has been explored as a means of dynamically concentrating graphical fidelity within the user's region of active visual attention, reducing peripheral rendering cost without sacrificing perceived quality (Smith et al., 2019). In heritage-specific applications, eye-gaze data have been applied to investigate visitor attention patterns at digitised artefacts and to infer iconographic preferences among different visitor demographics. The use of dwell-time thresholds to trigger contextual information overlays is conceptually analogous to kiosk-based interaction models deployed in physical museums, but the VR context enables delivery of spatially registered annotations directly within the navigable scene. No prior system has applied gaze-triggered annotation specifically to the iconographic content of the Karmawibhangga reliefs, nor within a photogrammetric reconstruction of an inaccessible architectural space.

3. METHODOLOGY

3.1 Data Sources and Integration

The digital twin is constructed using two primary datasets. First, UAV-based aerial photogrammetry and ground-level imagery provide a dense point cloud representing the global geometry of the Borobudur structure, following the approach of Kawato et al. (2021). The aerial survey employed a multi-rotor UAV platform equipped with a 42-megapixel calibrated camera, flown at 10 m above the monument surface on a double-grid flight plan with 80% forward and lateral overlap. Structure-from-Motion (SfM) processing yielded a dense point cloud of approximately 320 million points, georeferenced using 12 ground control points surveyed with differential GNSS, achieving a ground sampling distance of approximately 2.5 mm at the temple facade. This point cloud establishes the spatial reference framework within which the Hidden Foot passage is modelled.

Second, detailed 3D relief models reconstructed by Pan et al. (2022) from historical monocular photographs are incorporated.

These models restore fine sculptural features including human figures, vegetation, and ornamental patterns. The semantic segmentation results of Ji et al. (2023), assigning iconographic category labels to panel regions, are applied as vertex colour maps to the reconstructed meshes.

To embed these restored relief surfaces within the Hidden Foot passage, a custom geometric alignment workflow is implemented. The passage model simulates the real gap between the encasing stone wall and the relief-bearing wall of the temple base, guided by archival survey drawings and the photographic documentation of van Erp (1931). Figure 2 shows the full 160-panel sequence integrated within the Unity scene prior to rendering, with semantic segmentation colour-coding applied across the complete wall surface.

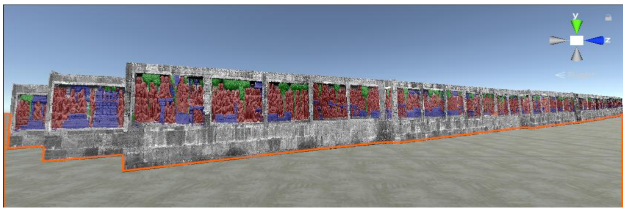

Figure 2. Full-wall integration of all 160 reconstructed Karmawibhangga relief panels within the Hidden Foot passage geometry in Unity. Semantic colour-coding—red: anthropomorphic, blue: structural, green: botanical—is applied uniformly across the panel sequence, enabling spatial overview of the iconographic distribution.

3.2 VR System Development

The VR environment is built in Unity 2022.3.22f1 and optimised for head-mounted display (HMD) operation using the HTC VIVE Pro Eye. Users enter the reconstructed passage through a transparent transition wall, simulating passage from the visible exterior into the concealed interior. Navigation is implemented via controller-based locomotion with smooth translation, enabling systematic exploration of the 160-panel sequence.

Eye-gaze data captured by the integrated eye-tracking system is used to trigger contextual text annotations when the user fixates on specific relief panels. The HTC VIVE Pro Eye records binocular gaze direction at 120 Hz with an angular accuracy of approximately 0.5° under laboratory conditions. A dwell-time threshold of 1.5 seconds was adopted to activate annotation panels, balancing responsiveness against unintended triggers during casual visual scanning. Annotation content was developed in consultation with iconographic specialists and covers the narrative subject of each panel, associated Pali and Sanskrit terminology, comparative references to stylistically related scenes elsewhere in the Borobudur cycle, and bibliographic pointers to primary scholarly literature. Semantic overlays can be toggled via controller interaction, revealing colour-coded segmentation of structural, anthropomorphic, and botanical elements. The rendering pipeline employs Unity's Universal Render Pipeline (URP) with custom shader variants for point cloud rendering, supplemented by a screen-space ambient occlusion pass to enhance depth perception within the narrow passage geometry. Controller haptic feedback is used to confirm gaze-trigger events and to signal proximity to panel boundaries during navigation. This dual-mode visualisation supports both immersive exploration and analytical inspection of the iconographic content. The photorealistic and semantic overlay modes are shown in Figures 3 and 4, respectively.

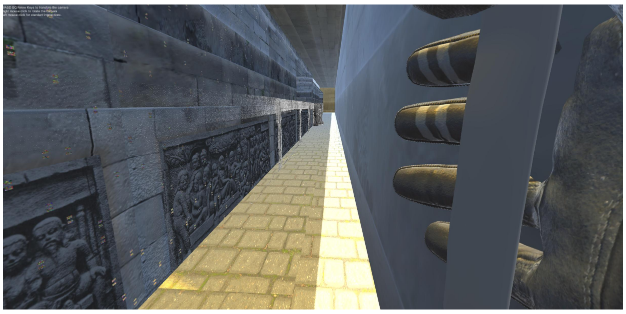

Figure 3. Immersive VR visualisation of the reconstructed Hidden Foot corridor in photorealistic mode. Restored relief models are illuminated by a virtual torch, with the illumination radius governing Level of Detail rendering.

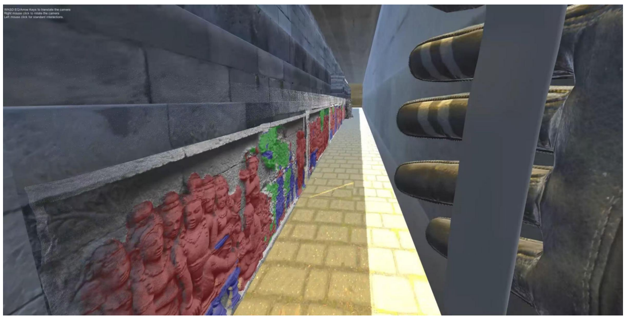

Figure 4. Semantic overlay mode of the VR corridor. Colour-coded segmentation highlights structural, anthropomorphic, and botanical iconographic categories, supporting scholarly analysis.

3.3 Level of Detail Optimisation

A novel Level of Detail (LoD) strategy is implemented to accommodate the computational demands of rendering large-scale point clouds and detailed relief geometry in VR. A handheld virtual torch serves as the primary light source and simultaneously as the LoD trigger: objects within the illumination radius are rendered at full resolution, while those outside are progressively simplified or culled.

This lighting-driven LoD produces a naturalistic low-light environment consistent with a subterranean passage, while simultaneously reducing rendering load. Technical integration relies on shader-based distance computation from the torch position, dynamically adjusting point cloud density and mesh detail in real time without perceptible visual discontinuity (Smith et al., 2019). Three discrete LoD tiers are defined: full-resolution rendering within a 1.5 m illumination radius; 50% density reduction between 1.5 m and 4.0 m; and culling of all geometry beyond 4.0 m. The shader computes a per-fragment distance metric from the torch world-space position and maps this to a continuous density multiplier via a smooth-step function, avoiding sharp transition artefacts at tier boundaries. Mesh geometry for the reconstructed relief panels is managed through Unity's built-in Level of Detail Group component, with transition thresholds aligned to the point cloud density tiers to maintain visual consistency between mesh and point cloud representations. The torch position is passed to all active shaders via

https://doi.org/10.5194/isprs-archives-XLIX-B4-2026-775-2026 | © Author(s) 2026. CC BY 4.0 License.

a globally-accessible shader property updated once per frame, avoiding redundant draw calls. The effect of this strategy is illustrated in Figure 5.

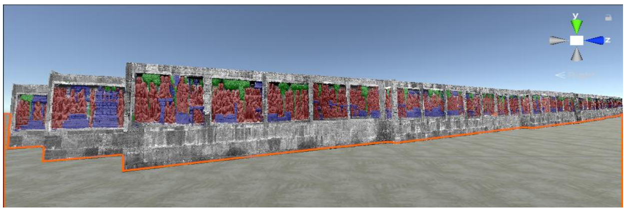

Figure 5. Torch-based Level of Detail (LoD) optimisation in Unity. Objects within the illumination radius are rendered in full resolution; distant regions are progressively simplified, reducing render batch count by over 35%.

4. RESULTS AND EVALUATION

4.1 System Performance

The hardware and software configuration used for development and evaluation is summarised in Table 1. All rendering benchmarks reported below were measured on this configuration under identical scene conditions.

Table 1. System hardware and software specification.

Component	Specification
GPU	NVIDIA GeForce RTX 4080 (16 GB VRAM)
CPU	Intel Core i9-13900K
RAM	64 GB DDR5
HMD	HTC VIVE Pro Eye
HMD Display	$2\times1440\times1600$ px, 90 Hz
Eye-tracking	120 Hz, $\sim$0.5° accuracy
Game Engine	Unity 2022.3.22f1 (URP)
Point Cloud Size	$\sim$320 million points

Technical evaluation was conducted on a workstation equipped with an NVIDIA RTX 4080 GPU and HTC VIVE Pro Eye HMD. Without LoD enabled, frame rates dropped below 10 FPS, causing severe motion latency and VR discomfort. With LoD activated, frame rates improved to 65–70 FPS, meeting VR fluidity standards and enabling comfortable navigation.

Render batch counts decreased by over 35%, and CPU frame times reduced from 99.8 ms to 15.2 ms, demonstrating the critical importance of illumination-based LoD in complex heritage environments. Table 2 summarises the key performance metrics.

Table 2. HMD rendering performance comparison: without vs. with LoD.

Metric	Without LoD	With LoD
Frame Rate (FPS)	10.0	65.6
CPU Frame Time (ms)	99.8	15.2
Render Batches	282	183

4.2 User Study

4.2.1 Participants Ten participants were recruited: five general-audience visitors (aged 18–35, no prior knowledge of Borobudur) and five VR enthusiasts or developers with professional familiarity with VR platforms. All participants had prior exposure to VR hardware. Informed consent was obtained from all participants prior to the session, and no personally identifying data were recorded.

4.2.2 Procedure Each participant completed a structured 15–20-minute immersive session comprising five sequential tasks: (1) entering the passage through the transition wall; (2) exploring restored relief panels under torch illumination; (3) activating the semantic overlay and comparing iconographic categories; (4) engaging with gaze-triggered textual annotations; and (5) navigating with and without LoD enabled for comparative performance perception. Two co-investigators observed each session and recorded behavioural notes independently using a structured observation protocol.

4.2.3 Measurement Instruments Post-session evaluation used a questionnaire adapted from the System Usability Scale (SUS) (Brooke, 1996), modified with domain-specific items for cultural heritage VR. Six dimensions were assessed on a 5-point Likert scale (1 = Very Low, 5 = Very High). Mean ratings exceeded 4.5/5 across all dimensions (Table 3).

Table 3. User study evaluation results ($n = 10$, 5-point Likert scale).

Evaluation Item	Mean Rating (/5)
Visual Fidelity of Reliefs	4.7
Intuitiveness of Navigation	4.3
Semantic Segmentation Helpfulness	4.8
Text Annotation Clarity	4.5
Lighting and Immersion	4.7
Overall Satisfaction	4.6

Semantic Segmentation Helpfulness scored highest (4.8/5), reflecting the perceived analytical value of the iconographic overlay. Navigation Intuitiveness was lowest (4.3/5), consistent with behavioural observations noting occasional disorientation during rapid movement in the narrow passage geometry.

Task completion rates across all five structured tasks were 100% for all participants; no participant failed to complete any assigned task within the session duration. Behavioural observations recorded by two co-investigators identified a characteristic pattern of spatial disorientation during rapid translation in the first two tasks, which diminished progressively as participants became familiar with the passage geometry, suggesting that a brief structured orientation phase prior to free exploration would improve early-session usability.

Heritage-professional participants specifically highlighted the semantic overlay as a tool of potential analytical value for systematic panel surveys, noting that the ability to visually isolate structural, anthropomorphic, and botanical elements across the full 160-panel sequence would substantially accelerate iconographic mapping tasks that currently require manual cross-referencing of archival photographs. General-audience participants demonstrated strong curiosity regarding the narrative content of the reliefs, frequently engaging with textual annotations beyond the minimum required by the task protocol. Six of ten participants voluntarily remained in the VR environment beyond the scheduled session duration.

Qualitative responses highlighted the immersive quality of the torch-lit environment and the marked perceptual difference between LoD-enabled and disabled modes. General-audience participants reported strong educational engagement and a convincing sense of spatial presence within the concealed structure. Observed minor challenges included occasional frame-rate drops during rapid head movements in sections with high panel density, sporadic unintentional gaze annotation triggers during quick scanning, and initial difficulty with spatial orientation in the early stages of the session—all consistent with the relatively low Navigation Intuitiveness score. These findings inform the design priorities identified in the limitations discussion (Section 5.3). Figure 6 illustrates the first-person view experienced by participants during the semantic overlay task, showing the in-corridor perspective with colour-coded relief surfaces and the narrow passage geometry that characterised navigation behaviour.

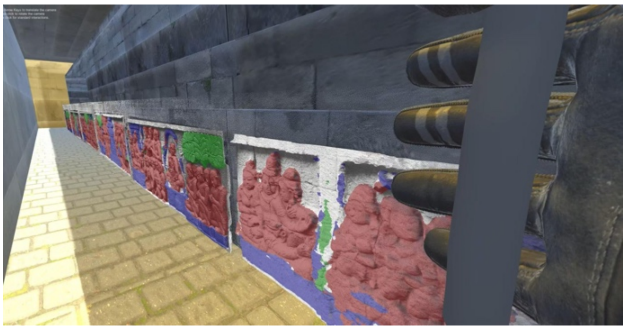

Figure 6. First-person VR view inside the reconstructed Hidden Foot corridor during the semantic overlay task. Colour-coded segmentation of relief surfaces is visible across multiple panel faces; the torch-illuminated foreground demonstrates the LoD rendering strategy in the in-HMD perspective experienced by study participants.

5. DISCUSSION

5.1 Novelty Relative to Prior Borobudur Work

This work is complementary to, rather than competitive with, prior Borobudur digital documentation efforts. Kawato et al. (2021) and Tanaka et al. (2021) provide invaluable multi-sensor coverage of the accessible monument structure. The present system does not replicate this work; rather, it addresses a fundamentally different challenge—interactive access to the one architectural layer for which no field survey method is viable. The novelty consists in the synthesis pipeline integrating deep-learning reconstructions, semantic annotations, and VR architecture; the lighting-driven LoD strategy adapted to enclosed-passage constraints; and the eye-tracking-based system for demand-driven iconographic interpretation.

The torch-centric LoD approach also represents a transferable methodological contribution beyond this specific site. Existing LoD implementations for heritage point clouds (Pathak et al., 2024) typically employ viewer-distance or screen-space error metrics, producing uniform simplification across the entire scene. The illumination-driven approach instead ties rendering complexity directly to the user's intended focus of attention, yielding visually coherent results particularly well-suited to enclosed environments where users naturally concentrate attention within the illuminated zone. This coupling of narrative device and rendering optimisation offers a transferable design template for immersive documentation of other subterranean or interior-restricted heritage spaces, including tomb chambers, mine workings, cave sanctuaries, and fortification interiors.

5.2 Dual-Purpose Platform

The system was designed to serve conservation documentation and public engagement as equally central functions. As a conservation tool, it provides a spatially organised, semantically annotated digital record of the Karmawibhangga reliefs accessible to heritage professionals. The semantic overlay supports systematic panel classification and iconographic distribution analysis, enabling iconographers to compare the distribution of anthropomorphic, structural, and botanical elements across the 160-panel sequence in a single navigable session—a task that previously required extended archival photograph comparison across multiple physical visits or document sources.

As an educational medium, the immersive design and contextual annotations make the hidden heritage layer accessible to non-specialist audiences. Concrete deployment scenarios include temporary museum installations at Borobudur heritage visitor centres and the National Museum of Indonesia, where the system could provide general audiences with a spatially coherent narrative of the sealed architectural layer. School and university programmes in archaeology, art history, and cultural heritage management represent a further deployment context, where the system supports structured inquiry-based learning without requiring travel to the physical site. Remote access configurations—using standalone HMDs or screen-based fallback rendering—extend the educational reach of the system to international audiences, consistent with the accessibility goals of the UNESCO World Heritage programme (UNESCO, 2023).

The toggle between photorealistic and semantic-overlay modes resolves a common trade-off in heritage digital twins, wherein systems optimised for conservation specialists sacrifice navigability and aesthetic appeal for analytical completeness, while systems designed for public engagement suppress metadata in favour of visual experience. The present system satisfies both user communities within a single coherent environment.

5.3 Limitations and Future Work

Panel placement relies on archival references and informed manual alignment, as precise in-situ position measurements for the Hidden Foot are unavailable. The current system requires a tethered PC-based headset, constraining accessibility. The user study involved ten participants; a larger-scale validation with heritage professionals and diverse international audiences would strengthen generalisability.

The panel placement methodology, while validated against archival survey drawings and van Erp's photographic documentation, would benefit from independent metric verification. A promising approach involves photogrammetric measurement of the accessible perimeter gap sections at the monument's edge, where limited access has occasionally been granted for conservation inspections, to obtain ground-truth reference distances for alignment calibration. Regarding user evaluation scope, future studies should recruit heritage professionals with specialist knowledge of Borobudur iconography, who may engage with the annotation and overlay systems in qualitatively different ways from the general-audience and VR-enthusiast cohorts represented here. The contribution of the semantic overlay to actual analytical productivity—as opposed to perceived usability—should be assessed through task-based studies in which participants perform iconographic classification tasks with and without overlay assistance, providing a more rigorous measure of the system's scholarly utility.

Future directions include multi-user collaborative features, haptic feedback and spatial audio integration, mobile AR and tablet deployment for broader access, multilingual annotation support, expansion to the upper terraces of Borobudur, and integration of multispectral or photometric stereo imaging to reveal pigment traces and surface properties invisible in standard photographic documentation.

6. CONCLUSIONS

This research demonstrates a practical, scalable framework for creating digital twins of inaccessible cultural heritage spaces. By integrating high-resolution point clouds, deep-learning reconstructions, and semantic segmentation into a Unity-based VR environment, the system provides immersive, interactive access to Borobudur's Hidden Foot reliefs—a space sealed for over a century and previously accessible only through limited archival photographs.

The VR-based digital twin not only restores visibility to a concealed historical layer but also enhances analytical understanding through semantic overlays and contextual annotation. The lighting-driven LoD strategy proved critical for achieving real-time VR performance, increasing frame rates from below 10 FPS to 65–70 FPS while preserving visual fidelity in the illuminated zone. The positive user study results confirm the system's value for both immersive public engagement and scholarly heritage analysis. The methodology is applicable to comparable inaccessible heritage spaces worldwide.

Three specific contributions merit emphasis. First, the data integration pipeline demonstrates that deep-learning-based relief reconstructions and semantic annotations derived from monocular archival photographs can be successfully embedded within a high-fidelity photogrammetric site model to produce a geometrically coherent and navigable digital twin. This pipeline overcomes the fundamental limitation that has rendered the Hidden Foot inaccessible to conventional digital heritage methods.

Second, the illumination-driven Level of Detail strategy constitutes a reusable architectural pattern for real-time VR rendering of large-scale heritage point clouds in enclosed spaces. By coupling the LoD trigger to the narrative light source, the system achieves frame rates suitable for head-mounted display operation without perceptible visual degradation within the user's region of active attention. This coupling resolves a persistent tension between visual fidelity and computational performance that has limited the deployment of dense point cloud models in immersive heritage environments.

Third, the eye-tracking-based annotation system enables demand-driven access to iconographic knowledge without interrupting the immersive experience. Unlike panel-level information kiosks or audio-guide systems, gaze-triggered annotations are spatially registered to the individual relief surfaces within the navigable environment, supporting a form of contextual learning that is uniquely enabled by VR. The mean user satisfaction ratings exceeding 4.5/5 across all evaluation dimensions, including 4.8/5 for semantic segmentation helpfulness, provide empirical support for the design decisions underlying this approach.

Collectively, these contributions advance the state of practice for digital heritage documentation of physically inaccessible monuments, and offer a methodological template applicable to subterranean chambers, sealed vaults, and other architectural spaces where direct survey access is precluded.

ACKNOWLEDGEMENTS

The authors thank the developers of the deep learning-based 3D reconstruction and semantic segmentation datasets that form the primary visual content of the present system. The historical photographic documentation of the Hidden Foot produced by Th. van Erp between 1907 and 1911 and archived at the Tropenmuseum, Amsterdam, provided the foundational evidence base from which the relief reconstructions were derived; the authors gratefully acknowledge this archival resource. The authors extend their thanks to the Indonesian Heritage Agency (*Badan Pelestarian Kebudayaan*) and the Borobudur Conservation Office for institutional support and access to archival survey drawings used in the passage alignment workflow. All ten study participants are thanked for their time and constructive feedback.

References

Brooke, J., 1996. SUS: A "quick and dirty" usability scale. P. W. Jordan, B. Thomas, B. A. Weerdmeester, I. L. McClelland (eds), *Usability Evaluation in Industry*, Taylor & Francis, London, 189–194.

Ji, S., Pan, J., Li, L., Hasegawa, K., Yamaguchi, H., Thufail, F. I., Brahmantara, Sarjiati, U., Tanaka, S., 2023. Deep Learning-Based Semantic Segmentation of 3D Point Clouds for Heritage Documentation. *Remote Sensing*, 15, 956.

Kawato, M., Li, L., Hasegawa, K., Adachi, M., Yamaguchi, H., Thufail, F. I., Riyanto, S., Brahmantara, Tanaka, S., 2021. 3D Digitization of Borobudur Relief Panels Using Close-Range Photogrammetry. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B2, 577–582.

Pan, J., Qiu, X., Zheng, L., Hu, X., Fang, L., 2022. 3D Reconstruction of Borobudur Reliefs from 2D Monocular Photographs Using a Deep Learning Approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 183, 439–450.

Pathak, S., Yılmaz, M., Kalkan, E., 2024. UAV-Based Topographical Mapping and Accuracy Assessment of Orthophoto Using Ground Control Points. *Mersin Photogrammetry Journal*, 6(1), 1–8.

Smith, J., Johnson, A., Williams, M., 2019. Dynamic Lighting and Level of Detail Systems for Real-Time Rendering of Complex Virtual Environments. *Journal of Real-Time Image Processing*, 12(3), 45–58.

Suzuki, Y., Yamaguchi, H., Thufail, F. I., Brahmantara, Hasegawa, K., Li, L., Tanaka, S., 2023. Comparison of locomotion methods for immersive VR visualization of Borobudur hidden foot reliefs using photogrammetric point clouds. *Proceedings of the 42nd JSST Annual Conference*, Niigata, Japan.

Tanaka, S., Nishino, K., Matsumoto, K., Hashimoto, S., 2021. Integrated High-Definition Visualization of Digital Archives of Borobudur Temple. *Remote Sensing*, 13(25), 5024.

UNESCO, 2023. Borobudur temple compounds. World Heritage List, United Nations Educational, Scientific and Cultural Organization. `https://whc.unesco.org/en/list/592`.

van Erp, T., 1907–1911. Photographs of the hidden foot of Borobudur. Tropenmuseum, Amsterdam, Netherlands.

van Erp, T., 1931. *The Restoration Report of Borobudur*. Government Printing Office, Batavia.