

Historical Building Information Modelling and Generative Artificial Intelligence Integration for 360 Degree Cultural Heritage Visualization: A Case Study of Pompeii

Özdemir Batuhan ^{1*}, Günaydın Esra¹, Duran Zaide ¹, Selbesoğlu Mahmut Oğuz¹

¹ Geomatics Engineering, Istanbul Technical University, Istanbul, 34469, Türkiye, ozdemirbat20@itu.edu.tr, esgunaydin@itu.edu.tr, duranza@itu.edu.tr, selbesoglu@itu.edu.tr

Abstract

Cultural heritage documentation traditionally prioritizes metric accuracy through three dimensional (3D) digital modelling and spatial data acquisition. While 3D models effectively represent structural geometry, they often lack experiential richness and immersive storytelling capacity. Beyond technical efficiency, immersive visualization environments play a significant role in supporting education, enhancing sustainability in digital heritage preservation, contributing to site protection, and providing remote access for individuals who are unable to physically visit archaeological locations. In this context, this study proposes a hybrid framework integrating Historical Building Information Modelling (HBIM) based modelling with Generative Artificial Intelligence (GAI) to develop 360 degree panoramic immersive visualizations. The ancient city of Pompeii was selected as the case study due to its historical significance and its relationship with the volcanic environment of Mount Vesuvius. A geometry based HBIM model was constructed to ensure volumetric consistency and spatial coherence, and its rendered outputs were integrated into a generative workflow using the Google Flow platform, guided by structured prompts and photographic references. The workflow follows a geometry constrained approach, where HBIM renders act as spatial references while real world images provide environmental context. Different generative video production approaches were evaluated using multiple input types, and the results indicate that approaches guided by reference images provide higher spatial consistency compared to less constrained generation. This study contributes by proposing a geometry constrained generative workflow and a 360 degree immersive visualization pipeline for cultural heritage applications, demonstrating that integrating HBIM geometry with AI improves spatial stability and environmental realism compared to unconstrained generation approaches.

Keywords: Historical Building Information Modelling, Generative Artificial Intelligence, Cultural Heritage, Immersive Visualization, Google Flow

Introduction

In recent years, the digital documentation and visualization of cultural heritage sites have emerged as a significant research area. Three dimensional (3D) models developed within the field of geomatics engineering enable the digital archiving of historical structures while preserving their geometric accuracy. These models, particularly through Historical Building Information Modelling (HBIM) approaches, serve as effective tools in the conservation and management of cultural heritage assets (Lovell et al., 2023). However, such digital models are generally limited to static representations and often fail to provide an experiential environment for users. Today, the objective is not only to document cultural heritage but also to transform it into interactive and educational environments. In this context, immersive visualization technologies offer significant potential by allowing users to experience historical spaces remotely (Murphy et al., 2009). At the same time, the development of generative artificial intelligence (GAI) technologies has made visual content creation faster and more accessible in digital environments. Artificial Intelligence (AI) based systems can synthesize images and videos to reinterpret historical spaces. Furthermore, the integration of AI with immersive

environments enables the development of more interactive digital experiences (Luck and Aylett, 2000). However, when these systems operate without geometric references, architectural inconsistencies, perspective distortions, and scale related issues may occur, creating reliability concerns, particularly in cultural heritage applications where accuracy is essential.

This study investigates how a geometry based HBIM model can be integrated with generative artificial intelligence to develop a hybrid visualization workflow that preserves metric accuracy while providing an immersive user experience. The archaeological site of Pompeii was selected as the case study due to its preserved urban structure and historical continuity. Unlike previous studies that address these technologies independently, this research adopts a geometry constrained approach, where the HBIM model is used as a spatial reference to guide generative processes. Different generative video production approaches are evaluated using multiple input types to examine their impact on spatial consistency and visual realism. In this context, the study contributes by proposing a geometry constrained generative workflow, providing a comparative evaluation of generative approaches, and developing a 360 degree immersive visualization pipeline for cultural heritage applications.

Data and Method

a) Study Area and Data

The study area is the archaeological site of Pompeii, located in southern Italy near the Bay of Naples. Pompeii was a Roman city that was suddenly buried under volcanic ash and pumice following the eruption of Mount Vesuvius in 79 AD. This event preserved much of the city's urban structure, architectural forms, and spatial organization, creating a unique archaeological environment in which the settlement remained largely frozen in time (Biancardo et al., 2020). For this reason, Pompeii provides an important reference context for digital heritage documentation and reconstruction studies. Previous research has demonstrated the application of multi-resolution and reality-based 3D modelling methodologies in the documentation of the Pompeii Forum, integrating laser scanning and photogrammetric techniques (Guidi et al., 2008).

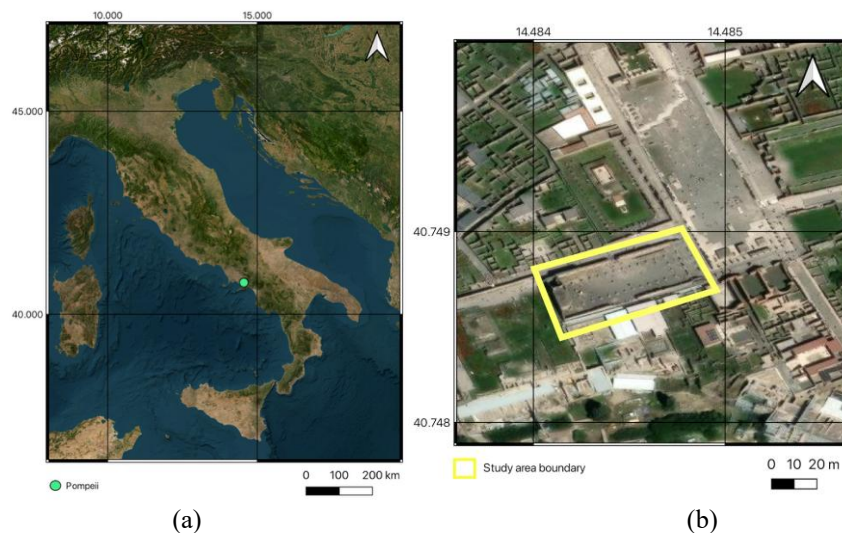


Fig. 1. a) Geographic location of Pompeii in Italy and (b) Temple of Jupiter in Pompeii.

Despite its level of preservation, the site is currently exposed to environmental degradation and intensive tourism. Therefore, digital documentation approaches such as digital twin modelling and HBIM are increasingly used as tools for sustainable conservation and management (Lovell

et al., 2023). In this study, a Digital Twin model of selected architectural elements of Pompeii was developed at approximately LOD 200 level, preserving overall volumetric proportions and spatial alignment without detailed ornamental features, as shown in Figure 2a. The interior spatial configuration and axial organization of the model are illustrated in Figure 2b. To evaluate the geometric reliability of the model, selected architectural elements were compared with point cloud measurements in both horizontal and vertical directions.



Fig. 2. a) Overall digital twin model, b) Interior perspective view.

The comparison indicates that the differences between the digital twin model and the point cloud measurements remain within decimetre-level tolerances. Horizontal deviations range between approximately -0.161 m and 0.080 m, while vertical deviations remain between -0.028 m and 0.018 m (Table 1). These values demonstrate that the model preserves overall geometric proportions and provides a reliable spatial reference for the subsequent generative workflow.

Table 1. Point cloud and digital twin measurement comparison

Structure	Element	Point Cloud Horizontal (m)	Point Cloud Vertical (m)	Model Horizontal (m)	Model Vertical (m)	Vertical Difference (m)	Horizontal Difference (m)
Pompeii	Active Column	0.799	4.542	0.875	4.549	-0.007	-0.076
Pompeii	Passive Column	0.794	1.614	0.955	1.626	-0.012	-0.161
Pompeii	Inner Wall	12.902	12.611	12.876	12.622	-0.011	0.026
Pompeii	Outer Wall	25.820	4.227	25.885	4.209	0.018	-0.065
Pompeii	Crepis	10.061	1.789	10.042	1.783	0.006	0.019
Pompeii	Drum	0.518	0.495	0.438	0.523	-0.028	0.080

The visual dataset consisted of rendered images derived from the digital twin model and contemporary photographs of the Pompeii site. The rendered images provided geometric references, while the photographs contributed material textures and environmental context for the generative process.

b) Generative Artificial Intelligence

Generative Artificial Intelligence enables the automated creation of visual, textual, and audiovisual content through data driven synthesis processes (Clark et al., 2025). These systems are increasingly used in digital visualization and immersive environments (Gupta et al., 2023). However, in cultural heritage applications, unconstrained generative models may produce architectural distortions, perspective inconsistencies, and unstable scene compositions, particularly when geometric accuracy is required. In this study, generative AI was used as a

visual synthesis tool to transform HBIM based renderings and real world photographs into video sequences for immersive visualization. Rather than relying on purely prompt based generation, the proposed workflow follows a geometry constrained approach. HBIM render images provide geometric structure, spatial proportions, and perspective alignment, while contemporary photographs of Pompeii contribute material textures and environmental context. In addition, structured prompts were used to define architectural elements, lighting conditions, atmospheric properties, and camera motion, allowing controlled generation of dynamic visual content. This approach enables multi source input conditioning, where visual references act as spatial anchors and prompts support scene composition and motion control. The overall workflow of the proposed system is illustrated in Figure 3, which presents the sequential pipeline from site analysis and HBIM based digital twin modelling to render production, integration of real world imagery, prompt design, generative video synthesis, and final output selection. As shown in Figure 3, the workflow is structured to ensure that geometric information derived from the digital twin remains the primary reference throughout the generative process, thereby reducing spatial inconsistencies.



Fig. 3. Workflow of the digital twin and generative AI-based VR reconstruction process.

The generative visualization stage was conducted using the Google Flow platform. Prompt design was developed by considering both semantic description and geometric preservation, and prompts were iteratively refined to improve spatial continuity, reduce visual inconsistencies, and maintain alignment with the HBIM geometry. This iterative process, as conceptually reflected in Figure 3, plays a critical role in controlling the balance between generative flexibility and geometric accuracy. Since video generation requires temporal coherence in addition to semantic alignment (Pan et al., 2017), two different generative video production strategies were evaluated in order to analyze their impact on spatial consistency and visual realism.

The first strategy, materials based generation, relies primarily on prompt driven descriptions of material properties, environmental conditions, and scene composition. While this approach allows flexible and creative outputs, it may lead to spatial inconsistencies, as the system tends to reconstruct the scene based on semantic interpretation rather than strict geometric constraints. The second strategy, frame guided generation, uses HBIM render images and real world photographs as direct visual references. This approach preserves architectural geometry more effectively, ensures stable background continuity, and improves temporal coherence across frames. As a result, it produces outputs that are more suitable for heritage visualization tasks requiring geometric fidelity. The comparative results of these two strategies are summarized in Table 2, which clearly demonstrates the difference in output efficiency and usability between the materials based and frame guided approaches.,

Table 2. Comparison of generative video production modes.

Generation Mode	Input Type	Total Videos Generated	Selected Videos
Materials to Video	Prompt + Digital Twin render + real photograph	9	1
Frames to Video	Prompt + Digital Twin renders + real photographs	13	6

As presented in Table 2, the materials to video mode produced 9 video sequences, of which only 1 was considered usable, whereas the frames to video mode generated 13 sequences, with 6 selected for integration into the final VR environment. This indicates that visually guided generation significantly improves both output reliability and spatial consistency. In addition to quantitative evaluation, representative prompt structures used in both approaches are provided in Table 3, illustrating how semantic descriptions, environmental parameters, and motion instructions were formulated to guide the generative process.

Table 3. Prompt design for generation approaches.

Generation Approach	Prompt Description
Materials based generation	Ancient Roman temple ruins in Pompeii with weathered stone columns and aged brick walls, realistic material textures, archaeological environment, natural daylight illumination, clear sky, cinematic lighting, wide angle perspective, human figure for scale
Frame guided generation	Ancient Roman temple ruins in Pompeii, maintain original architectural structure, realistic lighting conditions, smooth camera movement, stable background, no geometric distortion, high visual consistency

As shown in Table 3, prompts were carefully designed to balance descriptive richness and structural control, ensuring that generated scenes remained consistent with the underlying HBIM geometry while still enabling dynamic visual storytelling. In addition to the implemented workflow, alternative open source generative models such as AnimateDiff and Stable Video Diffusion were also considered. AnimateDiff supports image guided generation combined with prompt based control, while Stable Video Diffusion primarily relies on image driven generation with more limited prompt flexibility. A comparative evaluation of these platforms is presented in Table 4, highlighting differences in input requirements, computational complexity, controllability, and output stability.

Table 4. Comparison of generative AI tools.

Feature	Google Flow	AnimateDiff	Stable Video Diffusion
Multi-source Input	Supported (images, prompts, materials)	Text only (partial support)	Image only
Geometry-guided Generation	High	Low	Medium–Low
Temporal Consistency	High	Low	Low
Editability (Iterative Refinement)	High	None	None
Video Generation and Download Time	Fast (instant)	Very long (~7 days)	Moderate (~1 hour)
HBIM Workflow Compatibility	High	Insufficient	Insufficient

A comparative evaluation of generative video production platforms is presented in Table 4, highlighting differences in workflow integration, controllability, editability, and processing efficiency. The results indicate that Google Flow provides significant advantages over open source alternatives such as AnimateDiff and Stable Video Diffusion, particularly for geometry constrained cultural heritage visualization. One of the main reasons for selecting Google Flow in this study is its ability to support multi source input integration within a single environment. Unlike open source models, which typically rely on limited input structures, Google Flow enables the combined use of HBIM render images, real world photographs, and structured

prompts. This allows both visual and semantic control over the generative process, ensuring geometric consistency while maintaining flexibility in scene composition.

In addition, Google Flow offers a fully integrated and immediately usable workflow, eliminating the need for time consuming setup processes. As shown in Table 4, open source models often require lengthy preparation stages, including downloading and configuring model files, which can take more than one hour before use. In contrast, Google Flow allows direct access and rapid iteration. Its high level of editability enables efficient refinement of prompts and visual inputs without restarting the workflow, making the production process significantly faster and more flexible. Overall, these advantages demonstrate that Google Flow provides a more practical and reliable solution for HBIM based generative visualization tasks.

Results

This study presents a hybrid framework that integrates HBIM with generative AI for the production of 360 degree immersive cultural heritage visualizations. The results demonstrate that the use of HBIM based geometry as a spatial reference significantly improves the stability and consistency of generative outputs. By constraining the generative process with structured visual inputs, the proposed approach reduces geometric distortions and enhances the overall reliability of reconstructed environments. The comparative evaluation of different generative strategies highlights that frame guided generation provides higher spatial consistency and temporal stability compared to materials based generation. This finding emphasizes the importance of geometry constrained workflows in applications where architectural accuracy is essential. In addition, the comparison between integrated platforms and open source approaches indicates that while open source models offer flexibility and customization, they generally require more complex configuration and provide less stable results in geometry sensitive scenarios.

The proposed framework enables the representation of both static heritage visualization and dynamic historical scenarios within immersive environments, supporting educational applications and remote access to cultural heritage sites. However, several limitations were identified, including the reliance on prompt quality, the presence of visual inconsistencies in background elements, and the absence of fully interactive navigation within the generated environments.

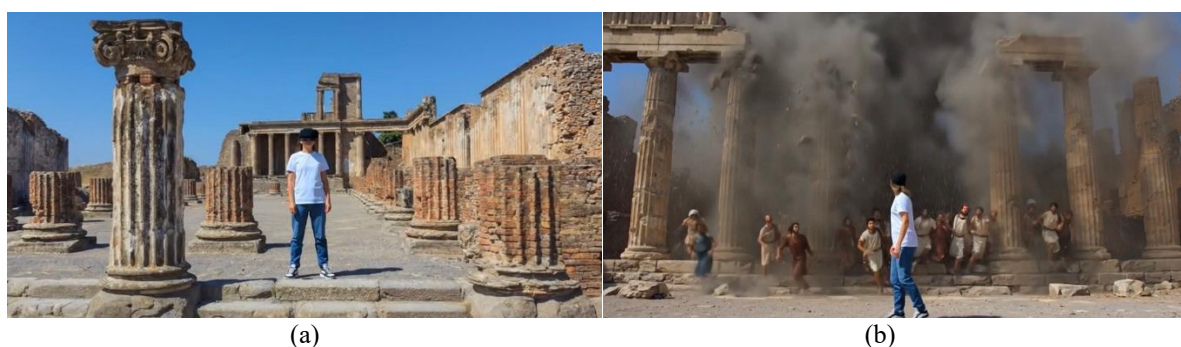


Fig. 4. a) Contemporary VR view of Pompeii and (b) Volcanic eruption VR reconstruction.

Conclusion

The results show a clear difference between the generative workflows evaluated within the Google Flow environment. The frame guided generation approach produced a significantly higher proportion of usable outputs compared to the materials based generation approach, with selection rates of approximately 46% and 11%, respectively. This difference is mainly due to the use of HBIM render images and real world photographs as direct visual references in the

frame guided approach, which constrains the generative process and preserves geometric structure. In contrast, the materials based approach relies more heavily on prompt driven interpretation, allowing greater visual flexibility but leading to spatial inconsistencies in terms of architectural alignment, perspective, and background continuity. The use of an HBIM based model improves the reliability of the generative process by providing clear geometric boundaries and acting as a spatial reference that helps maintain correct proportions and alignment. However, some visual inconsistencies were still observed, including partially missing background elements, instability in secondary structures, and variations in human figures, indicating the limitations of prompt dependent generation and the need for stronger structural constraints.

In addition to the comparison of generation strategies, the evaluation of different platforms highlights that integrated systems such as Google Flow provide higher spatial consistency, stronger temporal coherence, and greater ease of use compared to open source alternatives. While open source models such as AnimateDiff and Stable Video Diffusion offer flexibility and customization, they typically require additional configuration, GPU based processing, and do not provide fully integrated web based environments. Moreover, Google Flow enables more advanced control over scene composition, prompt refinement, and temporal continuity, resulting in more stable and reliable outputs, particularly in geometry constrained workflows.

Overall, the integration of HBIM geometry with generative artificial intelligence presents a promising direction for 360 degree immersive cultural heritage visualization. The proposed framework contributes by introducing a geometry constrained generative workflow and providing a comparative evaluation of different generation approaches and platforms. This approach supports both static heritage representation and dynamic historical reconstruction scenarios, enhancing accessibility, educational value, and sustainability of cultural heritage in digital environments. Future studies may focus on improving prompt control strategies, integrating real time user interaction, and evaluating the proposed workflow using open source generative frameworks. Further research may also investigate higher level modelling approaches and more detailed geometric representations to enhance both visual accuracy and experiential quality.

Data Availability

The VR demonstration video generated within the scope of this study is publicly available at: https://www.youtube.com/watch?v=2kXVWG_wqr4

References

- Clark, J., Barton, B., Albarqouni, L., Byambasuren, O., Jowsey, T., Keogh, J., ... & Jones, M. (2025). Generative artificial intelligence use in evidence synthesis: A systematic review. *Research Synthesis Methods*, 1-19.
- Lovell, L. J., Davies, R. J., & Hunt, D. V. (2023). The application of historic building information modelling (HBIM) to cultural heritage: a review. *Heritage*, 6(10), 6691-6717.
- Gupta, A., et al. (2023). Generative AI for virtual environments and digital content production. *Computers & Graphics*, 112, 45–58.
- Biancardo, S. A., et al. (2020). Digital survey and analysis of the archaeological site of Pompeii. *Journal of Cultural Heritage*, 45, 1–10.
- Pan, Y., Qiu, Z., Yao, T., Li, H., & Mei, T. (2017, October). To create what you tell: Generating videos from captions. In *Proceedings of the 25th ACM international conference on Multimedia* (pp. 1789-1798).

Murphy, M., McGovern, E., & Pavia, S. (2009). Historic building information modelling (HBIM). *Structural Survey*, 27(4), 311–327.

Guidi, G., Remondino, F., Russo, M., Rizzi, A., Voltolini, F., Menna, F., ... & Benedetti, B. (2008). A multi-resolution methodology for archeological survey: the Pompeii forum. In *Proceedings of the 14th International Conference on Virtual Systems and Multimedia (VSMM 2008)* (pp. 51-59). DEU.

Luck, M., & Aylett, R. (2000). Applying artificial intelligence to virtual reality: Intelligent virtual environments. *Applied Artificial Intelligence*, 14(1), 3–32.