ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Innovation of Digital Inheritance Education Mode of Traditional residential architectural Cultural heritage in immersive virtual Environment

Dongdong Zhang[1,*]

[1] School of Art and Design,Zhengzhou University of Industrial Technology,Zhengzhou, 451100, Henan, China

SUMMARY: *This paper constructs an immersive virtual inheritance system for the digital inheritance of traditional residential architectural cultural heritage, which integrates real image acquisition, point cloud reconstruction, component level semantic annotation, virtual space organization, interactive behavior perception and cultural content mapping into a unified computing framework to realize the collaborative expression of architectural form, decorative details, structural relationships and regional cultural narrative. Based on 286 traditional residential buildings, 4120 sets of scene interaction samples and 13840 semantic annotation data, the system completes 3D model training, space mapping and interaction testing. The results show that the completeness of 3D reconstruction is 94.8%, the texture fidelity is 93.5%, the accuracy of cultural content mapping is 92.6%, the average interactive response delay of the head-worn terminal is controlled at 76ms, and the task completion rate is 89.7%. The success rate of scene loading is 97.1%, the node hit rate is 91.4%, and the content traceback rate is 86.0%. The system performs stable in space identification, knowledge understanding and process recording. This method can enhance the ability of space identification, knowledge understanding and process recording while maintaining architectural authenticity and cultural coherence, which provides a computable, deployable and scalable implementation path for digital inheritance education of traditional residential architectural cultural heritage in immersive virtual environment.*

Povzetek: Ta članek gradi potopitveni virtualni sistem dediščinskega prenosa, ki združuje zajem slik, rekonstrukcijo oblaka točk, semantično označevanje, prostorsko organizacijo in kulturno preslikavo za digitalni prikaz tradicionalne stanovanjske arhitekturne dediščine. Poskusi kažejo, da je sistem dosegel 94,8 % popolnost rekonstrukcije, 92,6 % natančnost preslikave in odzivni čas 76 ms. Sistem izkazuje dobro stopnjo avtentičnosti, interaktivnosti in zmožnosti uvajanja.

KEYWORDS: *Immersive virtual environment; Traditional residential architecture; Digitization of cultural heritage; Virtual interaction*

1 Introduction

The cultural heritage of traditional residential buildings not only preserves the construction skills, spatial ethics and life memory, but also carries the structural logic, decorative vocabulary and regional cultural connotation. In the face of the update of display media and the change of knowledge transmission scene, it is difficult to fully present the spatial hierarchy, structural relationship and cultural context of traditional dwellings by relying on the expression of floor

---

*m15736770285@163.com

https://doi.org/10.65102/is2026192

Published: 30 April, 2026

plans, static photos and field visits. With the help of 3D modeling, semantic organization and interactive perception, the immersive virtual environment integrates architectural ontology, component information, narrative content and learning behavior into the same digital space, which provides a stable and reliable computer implementation basis for the digital inheritance education of traditional residential architectural cultural heritage. The path is not to simply move architectural materials into the virtual interface, but to make architectural cognition, cultural understanding and interactive learning interlinked expression in a unified system through data reorganization, relationship mapping and behavior feedback.

In recent years, related research has continuously promoted the digitization of cultural heritage into a stage of higher precision, stronger interaction and fine-grained organization. Boboc et al. reviewed the application of augmented reality in cultural heritage and pointed out that mobile terminal overlay, spatial registration and information annotation have become important supports for digital heritage display [1]. De Paolis et al. constructed a virtual reality augmented scheme around the castle scene, and proposed the realization idea of connecting tangible heritage and intangible cultural content with immersive spatial narrative [2]. Hulusic et al. studied the touchable interactive interface oriented to educational environment, and showed that the design of interactive devices would directly affect the operation coherence and cognitive engagement in the virtual heritage system [3]. Suvari et al. used augmented reality to complete the virtual restoration of lost cultural components and proved that component-level digital reconstruction could enhance the integrity of historical scene expression [4]. Lucifora et al. discussed the influence of virtual reality on heritage understanding from the perspective of cognition, and proposed that perspective switching and context substitution play a positive role in cultural interpretation [5]. Alatrash et al. constructed an immersive communication framework for engineering heritage, so that the interpretation process in the virtual environment has a complete task chain design [6].

At the level of system modeling and scene organization, Innocente et al. summarized the XR technical framework in the field of cultural heritage and pointed out that a stable data coupling relationship should be established between immersive devices, spatial content and user behavior [7]. Lucchi proposed to introduce digital twin into the scene of heritage construction automation, indicating that real-time data linkage can enhance the life cycle management ability of heritage objects [8]. Izaguirre et al. conducted research on heritage exploration based on VR, emphasizing the organizational value of spatial experience and knowledge discovery [9]. After quantitative analysis of VR, AR and MR Literature, Zhang et al. pointed out that immersion technology is shifting from a single display tool to a composite cognitive medium [10]. Rafeiro et al. proposed the expression mode of interweaving physical and digital around the immersive learning of lost architectural heritage [11]. Rodriguez-Garcia et al. systematically sorted out the path of 3D reconstruction of cultural heritage in immersive virtual reality, and believed that high-quality models, scene optimization and interaction mechanisms jointly determine application performance [12]. Wang et al. combined computer vision with immersive assessment for visual perception analysis of industrial heritage, reflecting the development trend of collaborative integration of vision algorithms and virtual display [13].

The research on digital reconstruction and system deployment for architectural heritage is also continuously deepening. Tini et al. proposed Scan-to-HBIM-to-VR integration process, which connected scanning data, building information model and virtual walkthrough into a continuous working chain [14]. Penjor et al. summarized the challenges and gaps of HBIM in heritage protection and pointed out that semantic accuracy, object classification and cross-platform compatibility still determine model usability [15]. Ramtohul et al. reviewed the augmented reality system of cultural heritage, and believed that in addition to visual superposition effect, interaction logic and information scheduling also affect the quality of practical application [16]. Buragohain et al. discussed the digitization of cultural heritage from the perspective of the metaverse, and proposed that multi-user sharing, content update and virtual autonomy mechanism would become the direction of the next stage [17]. Corrales-Serrano et al. verified the usability, learning and emotion scales in legacy virtual reality, which provided a quantitative basis for interactive application evaluation [18]. Zheng et al. studied the digital preservation process of architectural heritage based on virtual reality technology, emphasizing the collaborative relationship between modeling accuracy and display stability [19]. Iacono et al. completed the VR deployment of the museum scene, showing the reusable modeling way of small heritage space in the immersive system [20]. In order to compare the technical characteristics and application differences of related research, Table 1 summarizes the representative results in recent years.

The existing achievements have formed a multi-layer technology chain from 3D reconstruction, XR presentation, HBIM organization to interactive evaluation, but there are still several aspects to be improved in the digital inheritance education scene of traditional residential architectural cultural heritage. Architectural form reconstruction, cultural knowledge organization and learning behavior recording often belong to different modules, and the unified data closed loop has not yet formed. Many systems can complete the scene walkthrough, but do not further deal with the dynamic mapping between component semantics, cultural nodes and task feedback, so that the inheritance content remains at the visible level. Rendering load, response delay and interaction consistency in cross-terminal deployment also lack synchronous measurement, which is not conducive to system scalability. Based on this, this paper constructs an immersive virtual heritage system, which integrates traditional residential heritage modeling, virtual space reconstruction, cultural content mapping and interactive behavior analysis into a unified computing framework, and verifies the comprehensive performance and application efficiency of the system in scene construction, knowledge expression and interactive application through multi-terminal comparative test and ablation analysis.

Table 1: Comparison of immersive cultural heritage digitization studies in recent years.

Reference	Application Scenario	Core Technology	Spatial Reconstruction Capability	Interaction Organization Capability	Evaluation Dimensions	Computational Cost
De Paolis et al. [2]	Castle heritage presentation	VR scene modeling, immersive storytelling	Strong	Moderate	Scene immersion, cultural expression	Moderate
Innocente et al. [7]	Cultural heritage XR framework	XR system integration, device coupling	Moderate	Strong	Framework adaptability, interaction consistency	Moderate
Rodriguez-Garcia et al. [12]	Immersive heritage reconstruction	3D reconstruction, scene optimization	Strong	Moderate	Model quality, runtime performance	High
Tini et al. [14]	Architectural heritage documentation	Scan-to-HBIM-to-VR	Strong	Moderate	Modeling accuracy, workflow continuity	High
Corrales-Serrano et al. [18]	Heritage education application	VR interaction evaluation, scale validation	Moderate	Strong	Usability, learning, emotion	Moderate
Zheng et al. [19]	Digital preservation of architectural heritage	VR modeling, presentation deployment	Strong	Moderate	Preservation accuracy, presentation stability	Moderate

There is still room for deepening the existing research in the immersive digital inheritance of traditional residential architectural cultural heritage. Some studies focus on 3D modeling and virtual display, which can complete architectural form reproduction and scene browsing, but the linkage processing between component semantics, cultural nodes and learning behavior links is insufficient, and cultural content mostly stays at the static presentation level. Other research focuses on interactive experience and user perception, but lacks a unified computing logic for the mapping relationship between architectural spatial structure, construction level and task organization, which makes it difficult to form a continuous closed loop of scene browsing, knowledge triggering and feedback recording. In addition, most of the existing results are mainly evaluated based on immersion and usability, and quantitative indicators are seldom used, which is not conducive to method comparison and system replication.

Around the above technical boundaries, this paper treats the digital inheritance education of traditional residential architectural cultural heritage in the immersive virtual environment as a complete computer scene, and verifies the system construction effect. Based on this goal, we formulate the following technical Settings: 1) multi-view image reconstruction and component-level semantic parsing can be performed simultaneously to provide hierarchical geometric and cultural information for traditional residential buildings; Secondly, after introducing cultural content mapping and interactive behavior perception mechanism on the basis of 3D scene, the system can establish stable correspondence between architectural components, regional cultural information and learning task nodes. 3) Through multi-terminal deployment and quantitative testing, the system can form stable technical performance in terms of scene loading success rate, cultural content mapping accuracy, task completion rate and interactive response delay.

Based on the above Settings, this paper constructs an immersive virtual inheritance system for digital inheritance education of traditional residential architectural cultural heritage, which integrates heritage modeling, virtual space reconstruction, knowledge node organization, interactive feedback and result evaluation into a unified process. Based on the three-dimensional space reconstruction, this paper hierarchically labels the courtyard units, component types, decorative parts and cultural content, and establishes an interactive link for inheritance application through node trigger, path guidance and task record.

This paper further forms a technical path from spatial reconstruction to content mapping to interactive evaluation, connects the geometric modeling results with cultural knowledge nodes, realizes the orderly embedding of cultural content in the virtual scene, and constructs an interactive mechanism that can record the process of eye stay, node visit, path selection and task completion. It makes the inheritance activity change from static browsing to dynamic process which can be analyzed and played back. This method enhances the spatial expression ability and content organization ability of traditional residential architectural cultural heritage in immersive virtual environment, and also provides system support for the computer implementation of digital inheritance education mode.

2 Methods

2.1 Framework of virtual inheritance system

The framework of virtual inheritance system is the core operation basis of digital inheritance education of traditional residential architectural cultural heritage in immersive virtual environment. The framework takes multi-source real data access, three-dimensional space reconstruction, cultural semantic organization, interactive behavior perception and result feedback evaluation as the main line, and integrates the courtyard pattern, component relationship, decoration information and regional cultural knowledge of traditional houses into the unified calculation process. The front-end of the system receives image sequences, point cloud segments, component annotations, text data and task labels, and then completes registration, modeling, coding and mapping in a unified coordinate space, so that the virtual scene not only assumes the function of geometric rendering, but also supports knowledge triggering, path guidance and learning record synchronously. In this way, building entities, cultural content and user behavior are placed into the same closed loop, and a continuous linkage between scene browsing, node access and content traceback can be formed.

As shown in Figure 1, the system consists of multi-source data acquisition layer, spatial reconstruction layer, cultural mapping layer, interactive response layer and evaluation feedback layer. The multi-source data acquisition layer is responsible for obtaining the external facade, internal structure, decoration details and historical text information of traditional dwellings. The spatial reconstruction layer completes geometric alignment, mesh generation and hierarchical scene organization. The cultural mapping layer establishes the node index according to the component attribute, spatial location and knowledge label. The interactive response layer is responsible for path scheduling, event triggering and state updating. The evaluation feedback layer counted the task completion rate, node hit rate, response delay and content backtracking performance, and sent the results back to the mapping module and scheduling module.

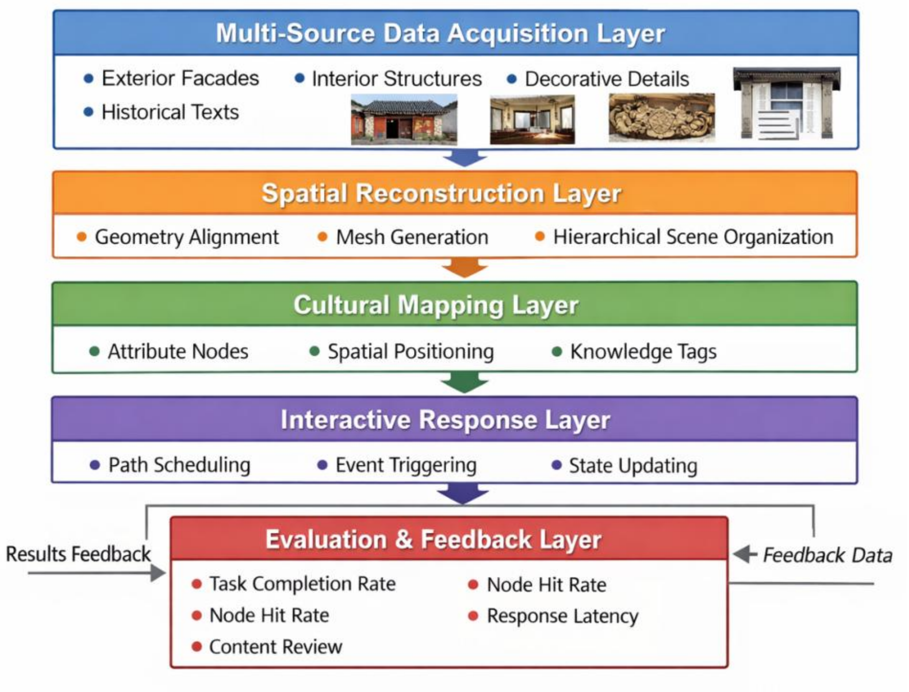

Figure 1: Framework of the virtual inheritance system.

In order to ensure the stable alignment of multi-source data in the same virtual space, the system adopts the coordinate mapping method with rigid transformation and scale normalization, and the standardized expression of scene points is shown in Equation (1):

$$\tilde{p}_i = \frac{Rp_i + t - \mu}{\sigma} \tag{1}$$

Here, p_i represents the original space point, $\tilde{p}_i$ represents the normalized scene point, R represents the rotation matrix, t represents the translation vector, μ represents the point concentration center, and σ represents the global scale factor. Equation (1) is used to unify the spatial data coordinates from different sources and provide a consistent geometric basis for subsequent scene reconstruction and semantic annotation.

In the component representation stage, it is difficult to support the complete expression of traditional dwelling cultural information only by relying on geometric contours. Therefore, the system encodes geometric features, texture features and cultural attributes into a fusion vector, whose calculation form is shown in Equation (2):

$$z_i = \lambda_1 g_i + \lambda_2 u_i + \lambda_3 m_i, \quad \lambda_1 + \lambda_2 + \lambda_3 = 1 \tag{2}$$

Here, g_i represents the geometric feature vector of the i-th component, u_i represents the texture and color features, m_i represents the cultural semantic vector, z_i represents the fused component representation, and λ_1, λ_2, λ_3 represent the weight coefficients. Equation (2) enables the system to simultaneously process component morphology, surface decoration, and cultural meaning within the same representation space, providing computable input for subsequent content mapping.

Considering that the display value of different components in the inheritance scenario is not the same, the system further introduces the node priority evaluation function, which is used to dynamically determine the trigger order of cultural nodes, and its form is shown in Equation (3):

$$q_i = \alpha \, \mathrm{Sim}(z_i, r) + \beta \frac{1}{d_i + 1} + \gamma c_i \tag{3}$$

Where q_i represents the priority score of the node, $\mathrm{Sim}(z_i, r)$ represents the similarity between the semantic vector of the component and the semantic vector r of the current task, d_i represents the spatial distance between the node and the current location of the user, c_i represents the cultural information density of the node, and α, β, γ represent the adjustment coefficient. Equation (3) makes the system preferentially present more relevant, closer and more informative cultural content in the path advancement process, thus enhancing the consistency of scene browsing and knowledge learning.

In the task organization phase, the system also needs to judge the matching strength between the cultural node and the learning task. To this end, this paper adopts the mapping strategy of semantic correlation and structural correlation, and its calculation expression is shown in Equation (4):

$$s_{ij} = \frac{z_i^T y_j}{||z_i|| \, ||y_j||} + \rho a_{ij} \tag{4}$$

Here, s_{ij} represents the matching score between the i-th cultural node and the j-th task, y_j represents the task semantic vector, a_{ij} represents the structural association marker between the node and the task, and ρ represents the association enhancement coefficient. Equation (4) is used to drive node binding, task push and path guidance, so that the operation sequence of users in the virtual space is coordinated with the unfolding sequence of cultural content.

In order to weaken the interference of instantaneous operation fluctuation on the interactive output of the system, the interactive response layer uses the time smoothing mechanism to modify the event stream, and its state update form is shown in Equation (5):

$$r_t = \theta r_{t-1} + (1 - \theta)(\kappa_1 e_t + \kappa_2 n_t + \kappa_3 f_t) \tag{5}$$

Here, r_t represents the system response state at time t, r_{t-1} represents the response state at the previous time, e_t represents the current interaction event intensity, n_t represents the node hit value, f_t represents the feedback correction amount, θ represents the smoothing coefficient, κ_1, κ_2, κ_3 represent the event contribution weight. Equation (5) can suppress the state jitter caused by sudden operation, so that the head-mounted terminal and the desktop terminal can maintain a relatively stable response rhythm during the switching process.

Under the support of the above mechanisms, the virtual heritage system framework integrates the spatial reconstruction, cultural expression and interactive analysis of traditional residential architectural cultural heritage into a single running link, allowing users to complete continuous operations such as browsing, recognition, understanding and backtracking in the virtual environment. The framework enhances the extensibility of scene construction, the interpretability of content organization and the traceability of interactive feedback. It also provides a unified interface for subsequent heritage modeling, virtual space reconstruction, cultural content mapping and experimental verification.

2.2 Traditional residential heritage modeling and virtual space reconstruction

Traditional residential heritage modeling and virtual space reconstruction are the key links in the virtual heritage system. Its goal is to generate 3D scenes that can be used for immersive browsing, knowledge triggering and task guidance under the premise of maintaining the consistency of architectural form, structural hierarchy and cultural clues. In the acquisition stage, the collaborative input method of multi-view images, depth clips and component labeling is used to continuously record the entrance of the courtyard, the turning point of the corridor, the node of the roof truss, the details of the doors and Windows, and the decorative parts. In order to ensure the complete coverage of the model, the external images are collected according to the surround path, and the overlap rate is controlled at more than 75%. In the indoor narrow space, the segmented filling strategy is used to reduce the influence of occlusion on the matching stability. For the areas with obvious texture changes such as wood, brick and painting, the system additionally saves high-resolution samples for subsequent detail restoration and semantic correction. Figure 2 shows the process of traditional residential heritage modeling and preprocessing. Firstly, the original image and the depth segment are time-stamped aligned, and then distortion correction, feature screening, component segmentation, noise suppression and hole filling are completed.

Figure 2: Traditional residential heritage modeling and preprocessing process.

Before entering the feature matching and dense reconstruction, the system needs to evaluate the availability of images from different viewpoints to weaken the interference of blurred, uneven exposure and insufficient overlapping samples on the modeling quality. To this end, this paper introduces the image quality weighting function, which is used to calculate the contribution weight of each input image in the reconstruction process, and its expression is:

$$\omega_i = \frac{\exp(\tau_1 c_i + \tau_2 o_i + \tau_3 l_i)}{\sum_{j=1}^{N} \exp\left(\tau_1 c_j + \tau_2 o_j + \tau_3 l_j\right)} \tag{6}$$

Here, ω_i represents the reconstruction weight of the i-th image, c_i represents the image sharpness score, o_i represents the overlap with neighboring views, l_i represents the illumination consistency coefficient, τ_1, τ_2, τ_3 represent the adjustment parameters, and N represents the total number of input images. Equation (6) is used to complete the quality assessment and contribution allocation of image samples before reconstruction, so as to improve the stability of subsequent feature matching and spatial reconstruction.

In the feature organization stage, the system no longer only relies on a single image feature, but synchronously introduces geometric cues, texture cues and component labels. Then, in view of the characteristics of beer-column handover, cornice turning and repeated panes in traditional houses, the system uses the local consistency constraint to complete the candidate matching screening. Let the feature sets of the i-th image and the$j$-th image be f_i andf_j, respectively, and the matching score function be defined as:

$$m_{ij} = \omega_1 \exp\left(-\frac{||f_i - f_j||_2^2}{\sigma_f^2}\right) + \omega_2 \cos(n_i, n_j) \tag{7}$$

Here, m_{ij} represents the feature matching score, ω_1 and ω_2 represent the weight coefficients, σ_f represents the descriptor scale parameters, and n_i and n_j represent the local neighborhood direction vectors. Equation (7) considers both descriptor distance and neighborhood structure consistency, thereby weakening mismatching caused by repeated patterns.

After the sparse pose recovery is completed, the system updates the camera parameters and 3D point positions through the reprojection optimization with weights. Let the observed pixel be u_{ik} and the projection function be $\pi(\Theta_i, X_k)$, then the objective function is written as:

$$\min_{\Theta,X} \sum_{i=1}^{N} \sum_{k=1}^{M} \rho\left(||u_{ik} - \pi(\Theta_i, X_k)||_2^2\right) + \lambda ||LX||_2^2 \tag{8}$$

Here, Θ_i represents the i-th camera parameter, X_k represents the k-th 3D point, $\rho(\cdot)$ represents the robust loss function, L represents the adjacency constraint matrix, and λ represents the smoothing coefficient. Equation (8) maintains local geometric continuity while suppressing outlier errors.

Figure 3 illustrates the virtual space reconstruction process. Firstly, the initial skeleton is generated from the sparse pose recovery, and then the dense reconstruction is performed according to the boundary and spatial partition of the components. After that, semantic tags are introduced to classify the roof, wall, beam frame, door and window, and decorative units hierarchly.

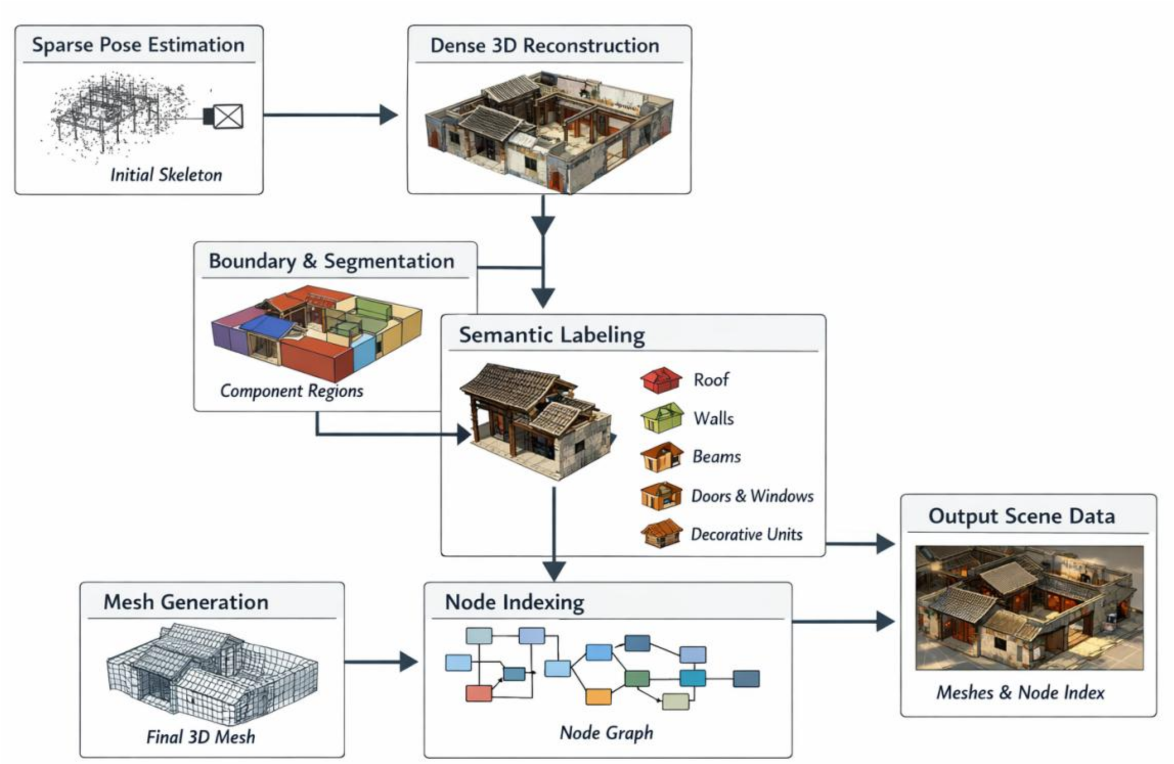

Figure 3: The virtual space reconstruction process.

Based on this, we further introduce a semantic-guided mesh fusion strategy to incorporate both geometric adjacency and cultural labels into the mesh generation process. Let the grid cell fusion weight be $\mathrm{w_m}$, the geometric similarity $\mathrm{g_m}$, and the semantic consistency $\mathrm{c_m}$, which can be expressed as:

$$\mathrm{w_m} = \alpha \mathrm{g_m} + \beta \mathrm{c_m} + \gamma \frac{\mathrm{o_m}}{\sum_{\mathrm{r}=1}^{\mathrm{K}} \mathrm{o_r}} \tag{9}$$

Here, $\mathrm{o_m}$ represents the number of observed coverage of the m-th grid cell, and α, β, γ represent the fusion coefficients. Equation (9) can avoid merging components only according to geometric proximity, so as to maintain the independent expression of detail units such as the bird, the dougan and the door hairpin in the virtual scene. After the reconstruction is completed, the system binds the 3D model to the knowledge node, and the stay, click and look back behaviors triggered by the user during the roaming process are written into the scene graph in the reverse direction, which is used to adjust the model loading order and the cultural content expansion level. The virtual space constructed in this way not only has high form restoration ability, but also has strong content organization ability of serving digital inheritance education.

2.3 Virtual interaction mechanism and cultural content mapping method

Virtual interaction mechanism and cultural content mapping method determine the operation quality of digital inheritance education of traditional residential architectural cultural heritage in immersive virtual environment. This method does not simply attach cultural information to the surface of the 3D model, but organizes the courtyard structure, component semantics, behavior trajectory and task goal into a unified computable relationship network. Firstly, the system builds a cultural node set according to spatial location, component category and knowledge tag, and then uses interactive events to drive node activation, content expansion and path adjustment, so that browsing behavior, knowledge triggering and result feedback are synchronized. Figure 4 shows the virtual interaction mechanism and the cultural content mapping process. The graph takes the scene node as the center, and connects the task semantics, the user state and the feedback cache, forming a closed loop structure of "input-matching-triggering-recording-back".

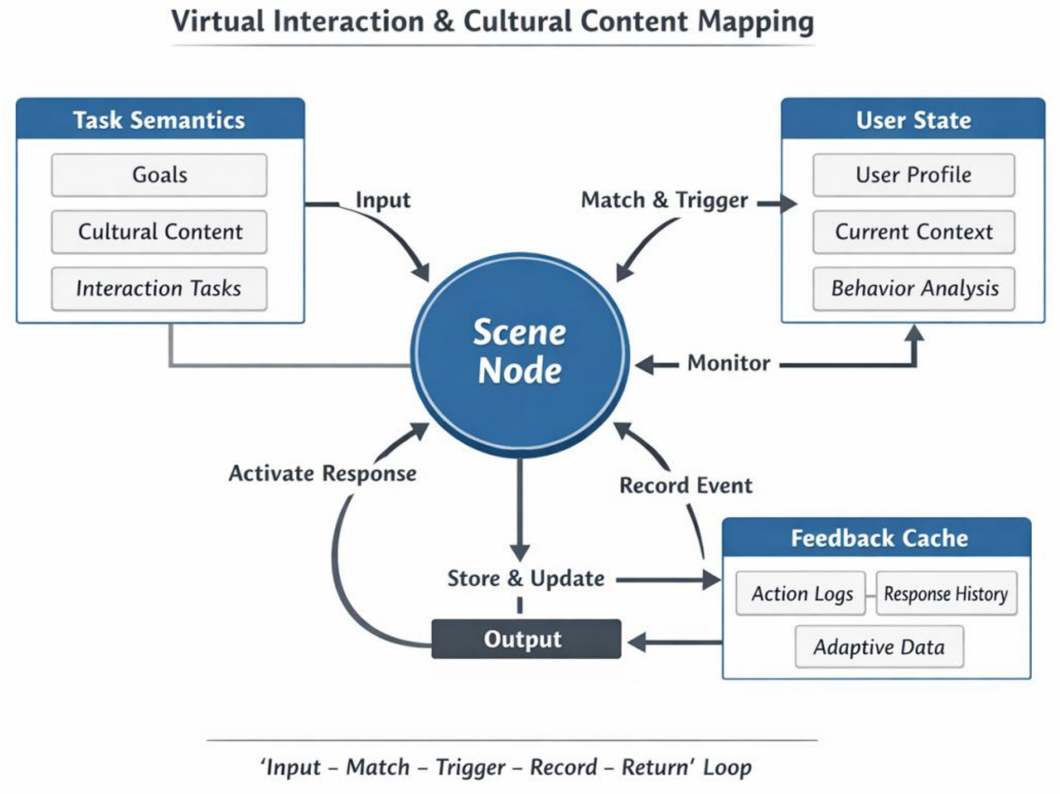

Figure 4: Virtual interaction mechanism and cultural content mapping process.

In order to measure the semantic proximity between scene nodes and tasks, this paper defines a node matching function as shown in Equation (10):

$$s_i = \frac{z_i^T y}{||z_i||\,||y||} + \lambda_1 \exp\left(-\frac{d_i^2}{\sigma_d^2}\right) \tag{10}$$

Here, s_i represents the matching score between the i-th cultural node and the current task, z_i represents the semantic vector of the node, y represents the semantic vector of the task, d_i represents the spatial distance between the node and the current position of the user, λ_1 is the regulation coefficient, and σ_d is the distance scale parameter. Equation (10) is used to comprehensively evaluate semantic relevance and spatial proximity.

In the interactive triggering phase, the system synthetically calculates the activation strength of gaze stay, movement direction and historical access state, and its expression is shown in Equation (11):

$$a_i = \alpha_1 \frac{g_i}{g_{max}} + \alpha_2 \cos(\theta_i) + \alpha_3 (1 - h_i) \tag{11}$$

Here, a_i represents the node activation strength, g_i represents the current gaze length, g_{max} represents the upper bound of normalization, θ_i represents the Angle between the gaze direction and the node normal direction, h_i represents the historical visit frequency, and α_1,α_2,α_3 are the weight coefficients. Equation (11) is used to determine whether the node enters the trigger queue in the current interactive state.

When multiple cultural nodes satisfy the trigger condition at the same time, the system needs to dynamically allocate the priority of content push. To this end, this paper introduces the node scheduling weight function, whose form is shown in Equation (12):

$$w_i = \beta_1 s_i + \beta_2 a_i + \beta_3 c_i + \beta_4 \frac{1}{r_i + 1} \tag{12}$$

where w_i represents the node scheduling weight, c_i represents the cultural information density, r_i represents the interval between the last time the node was visited and the current moment, and β_1,β_2,β_3, and β_4 are the fusion coefficients. Equation (12) is used to determine the order of different cultural nodes in the content push sequence.

After cultural content enters the display queue, it needs to be consistent with the spatial position. Let the mapping coordinate of the i-th cultural content block in the scene be q_i, and its spatial projection relationship be shown in Equation (13).

$$q_i = K(Rx_i + t) \tag{13}$$

Here, x_i represents the 3D coordinates of the cultural nodes, R and t represent the rotation matrix and translation vector of the current viewpoint, K represents the projection matrix, and q_i represents the mapping position of the content block in the display plane. Equation (13) ensures that the text, image or voice labels are synchronized and corresponding to the position of the component.

In order to ensure the hierarchy of the content presentation process, the system organizes text, image, voice and component description into multi-modal expression units, and uses the fusion function to calculate the comprehensive display value, which is in the form of Equation (14):

$$m_i = \sum_{k=1}^{4} \gamma_k u_{ik}, \quad \sum_{k=1}^{4} \gamma_k = 1 \tag{14}$$

Here, m_i represents the comprehensive display value of the i-th node, u_{ik} represents the contribution of the k-th type of modal information, and γ_k represents the modal weight. Equation (14) is used to control the expansion ratio of text, image, voice and component description.

In the behavior feedback phase, the system cumulatively encodes click, stay, look back, and task completion, and generates an interaction state vector. In order to weaken the influence of contingency operation on feedback judgment, this paper adopts the time smooth update strategy, whose calculation method is shown in Equation (15):

$$r_t = \eta r_{t-1} + (1 - \eta)(\delta_1 e_t + \delta_2 v_t + \delta_3 u_t) \tag{15}$$

Here, r_t represents the feedback state vector at time t, r_{t-1} represents the state at the previous time, e_t represents the current event encoding, v_t represents the task completion signal, u_t represents the content review information, η represents the smoothing coefficient,δ_1,δ_2,δ_3 represent the weight parameters. Equation (15) is used to stabilize the continuous judgment of the system on the user interaction state.

With the support of the above mechanism, cultural content mapping is no longer dependent on fixed scripts, but dynamically generated according to node relationships, task semantics and real-time behavior. This method makes the gatehouse, beam frame, courtyard, window lattice and decorative components in the traditional residence form an ordered knowledge link in the virtual environment, and also enables users to obtain continuous interactive support in the process of browsing, recognition, understanding and review, so as to provide a unified interface for content organization, task promotion and effect evaluation in the digital heritage education scene.

2.4 Experimental Design

In order to verify the applicability of the virtual inheritance system in the digital inheritance education of traditional residential architectural cultural heritage, 286 traditional residential buildings were selected as real scene samples, and 4120 groups of interaction samples and 13840 semantic annotation data were constructed. At the acquisition end, the SONY A7M4 camera and the iPad Pro laser scanning module were used to obtain the image and depth information. The original image resolution was controlled at 3840×2160, and the overlap rate was maintained at more than 75%. All the data were preprocessed, modeled and labeled under Windows 11 environment, and the 3D reconstruction was performed using RealityCapture and Blender processes. The interactive module was deployed on the Unity platform.

The dataset was divided into training set, validation set and test set according to 70%, 15% and 15%, and the building type, component category and random seed were recorded during the division process to ensure the reproducibility of the experimental results. The cultural content samples cover six types of objects, including gatehouses, beam frames, courtyards, window lattice, painted and brick carvings. Texts, images and voice materials are uniformly numbered into the semantic library. In the training phase, rotation, brightness perturbation, random cropping and noise injection enhancement strategies are used.

The experimental platform is configured with Intel Core i9-13900K processor, 64GB memory and NVIDIA RTX 4090 graphics card, and the framework is PyTorch. The number of model training rounds is set to 120, the batch size is 12, AdamW is selected as the optimizer, the learning rate is set to 0.0005, and the attenuation is carried out by cosine annealing strategy. The evaluation phase focuses on the completeness of 3D reconstruction, the accuracy of cultural content mapping, the task completion rate and the average response delay. At the same time, the node hit rate, the knowledge backtracking rate and the cross-terminal stability are recorded. Five-fold cross validation was used in the experiment, and scene-level confidence intervals were combined to calculate the system performance under different terminal and different task conditions.

3 Results and discussion

3.1 Construction results of virtual scenes of cultural heritage of traditional residential buildings

Based on 286 traditional houses, 4120 sets of interaction samples and 13840 semantic annotation data, this paper statistics the scene construction results of the virtual inheritance system. The results show that the three levels of the overall courtyard, the interior of the hall and the decorative details form a stable spatial reconstruction and content embedding structure. The completeness of the overall scene of the courtyard is 95.6%, the interior of the hall is 94.3%, and the decorative details are 93.8%, indicating that after the fusion of multi-view reconstruction and semantic constraints, the main structure and detail components can maintain continuous expression in a unified coordinate system. As shown in Figure 5, the difference between the completeness and texture fidelity of different scene types is small, and the overall fidelity is stable between 92.9% and 94.1%.

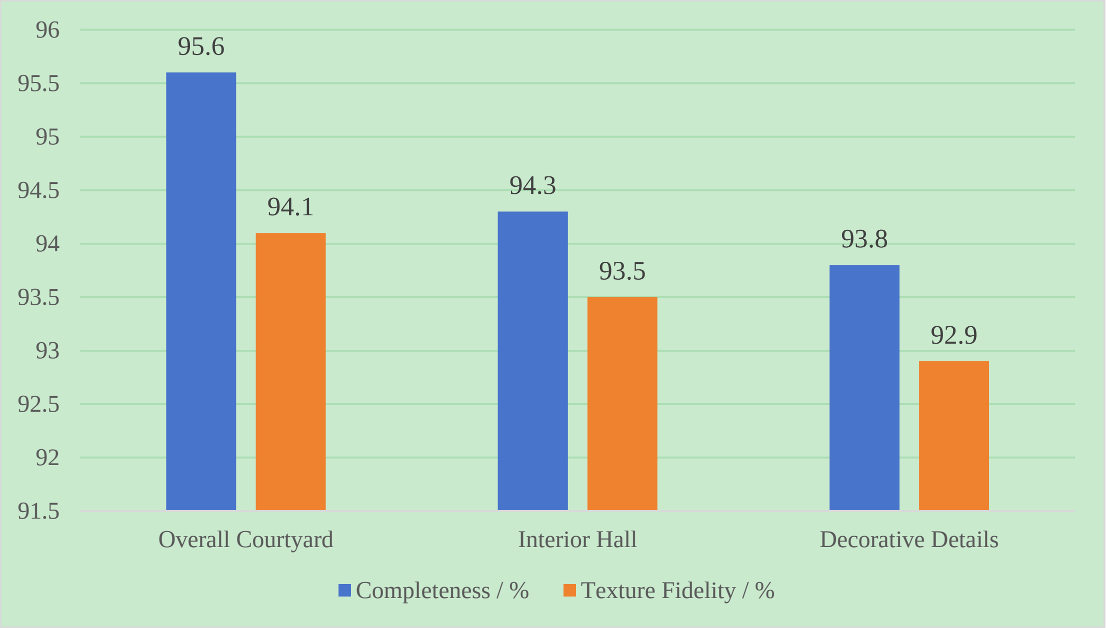

Figure 5: Completeness versus texture fidelity for different scene types.

Table 2 further presents the main results of the virtual scene construction. The overall accuracy of semantic node binding reaches 92.6%, the success rate of scene loading reaches 97.1%, and the node coverage rate reaches 93.4%, which indicates that a stable mapping relationship has been formed between geometric models and cultural content. The node coverage of the hall interior scene is slightly lower, and the compact indoor space and partial occlusion will affect the component boundary extraction and label projection. Although the scale of decorative detail scene is small, the node binding accuracy is still 91.8% under high-resolution sampling and local compensation.

Table 2: Results of virtual scene construction.

Scene Type	Completeness / %	Fidelity / %	Binding Accuracy / %	Coverage / %	Loading Success Rate/%
Overall Courtyard	95.6	94.1	93.4	94.6	97.8
Interior Hall	94.3	93.5	92.7	92.8	97.0
Decorative Details	93.8	92.9	91.8	92.7	96.5
Overall	94.8	93.5	92.6	93.4	97.1

As shown in Figure 6, the three types of scenarios have obvious differences in the distribution of the number of knowledge nodes and the effective reach rate. The overall courtyard scene contains 48 knowledge nodes, and the effective reach rate is 91.7%. The internal scene of the hall contains 39 knowledge nodes, and the effective reach rate reaches 93.2%. The decorative detail scene contains 26 knowledge nodes, and the effective reach rate is 88.5%.

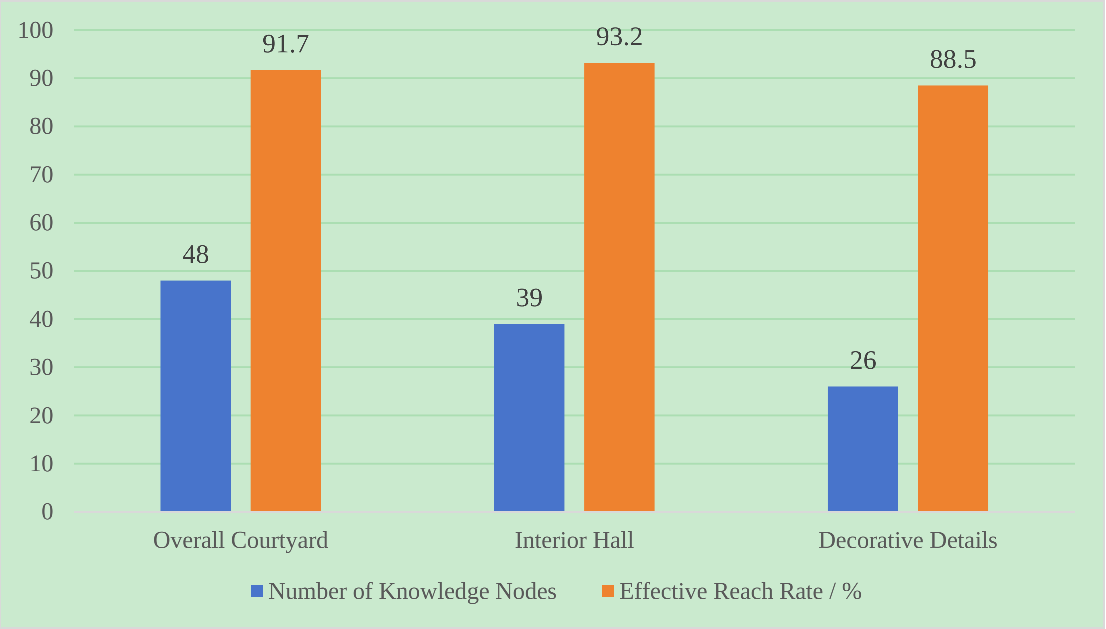

Figure 6: Number of nodes versus effective reach rate for different scenario types.

According to the data performance, the number of nodes in the overall scene of the courtyard is the largest, which is suitable for undertaking spatial guide and overall cognitive tasks. Although the total number of nodes in the interior scene of the hall is slightly less, the reach rate is the highest, indicating that the structure description, etiquette information and spatial function content are easier to be stably called in this level. The number of nodes and the reach rate of decorative detail scenes are relatively low, because the scale of local components is small and the visual focus is scattered, and users are more likely to jump in short browsing. Combined with Figure 5 and Table 3, it can be seen that the virtual inheritance system has formed a good spatial reduction ability, semantic binding ability and content organization ability in the scene of traditional residential architectural cultural heritage, which can provide a stable scene foundation for subsequent interactive application analysis.

3.2 Interactive application effect analysis of virtual inheritance system

Based on the above virtual scene construction results, this paper further analyzes the interactive application performance of the virtual heritage system in the traditional residential architectural cultural heritage scene. In the experiment, three tasks were selected: courtyard guide, hall identification and decoration review, and node access, path response, content backtracking and task completion were recorded. The results show that in 4120 interaction samples, the average task completion rate of the system is 89.7%, the knowledge point recognition consensus rate is 91.4%, the cultural content recall score is increased by 15.3% and the spatial identification score is increased by 17.9% compared with the baseline, which indicates that the task semantic matching, node scheduling and feedback update mechanisms have formed a stable application support. Taking the tasks of beam identification and ritual node location in the hall scene as an example, the access sequence recorded by the system is close to the manual statistical results, and the node trigger rhythm and look-back behavior maintain strong consistency, which indicates that the multimodal content mapping can play a stable role in the real browsing path.

Table 3 presents the error analysis of the interactive application results. The performance of the system is close to the manual statistical values in task completion rate, node hit rate, content backtracking rate and average response delay. The error of task completion rate is 1.1%, node hit rate error is 1.6%, content backtracking rate error is 1.5%, and the average response delay error is controlled within 3ms. The deviation for the path-guiding task is slightly higher, mainly related to node switching in local high-density texture regions, but it still maintains good interaction continuity and content carrying capacity as a whole.

Table 3: Error analysis of interactive application results of virtual inheritance system.

Metric	Manual Statistical Value	System Recorded Value	Error	Relative Error
Task Completion Rate / %	90.8	89.7	1.1	1.2%
Node Hit Rate / %	93.0	91.4	1.6	1.7%
Content Review Rate / %	87.5	86.0	1.5	1.7%
Average Response Latency / ms	79	76	3	3.8%

As shown in Figure 7, the response time delay of different interaction stages shows a relatively obvious two-line variation feature. The response delays of the head-mounted terminal in the stages of entrance guide, node trigger, hall turning, beam identification, decoration expansion, content review and task confirmation were 54ms, 68ms, 129ms, 61ms, 141ms, 39ms and 40ms, respectively, and the overall average was 76ms. The response delays of the corresponding phases of the desktop terminal are 41ms, 55ms, 101ms, 49ms, 109ms, 35ms and 51ms, respectively, and the overall average is 63ms.

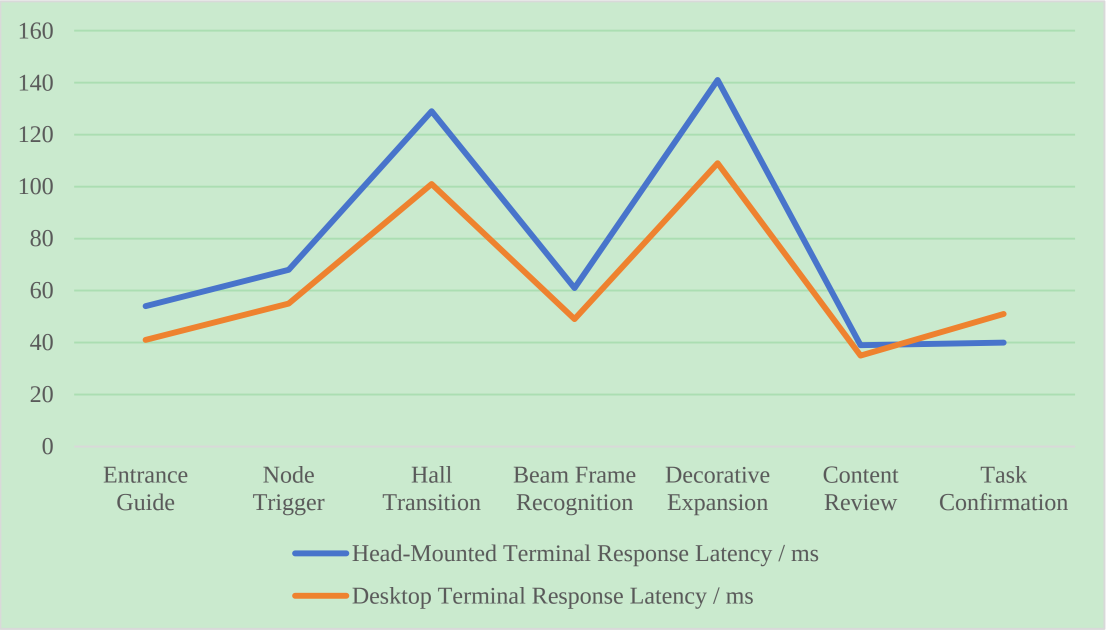

Figure 7: Response time delay variation for different interaction phases.

The two sets of data show a significant increase in the hall turning and decoration expansion stages, indicating that high-resolution tile loading, multi-modal content synchronous expansion and local view switching will significantly increase the system overhead. In the stage of content review and task confirmation, it fell back quickly, indicating that the system has better recovery ability in the later stage of interaction. On the whole, both types of terminals maintain a stable task advancement rhythm. The desktop terminal has a smaller fluctuation range, and the head-mounted terminal shows higher computational load characteristics in immersive interaction scenarios. Combined with Table 4, it can be seen that the virtual inheritance system has good interaction stability, content scheduling ability and task support ability in the scene of traditional residential architectural cultural heritage, which can provide reliable results for subsequent method comparison and ablation experiments.

3.3 Comparison of different methods

In order to verify the effectiveness of the proposed virtual inheritance system in the digital inheritance education scene of traditional residential architectural cultural heritage, this paper compared it with five methods: HBIM-Viewer, VR-Link, XR-Guide, SceneGraph-VR and ManualRoute. Each method uses the same sample of 286 traditional dwellings, 4120 groups of interaction data and 13,840 semantic annotations. The training rounds, input scale and task script are consistent. The comparison indicators include scene integrity, mapping accuracy, task completion rate and average response delay, the results are shown in Table 5.

Table 4 shows that the proposed method outperforms the comparison methods in all four core indicators. The completeness of the scene reaches 94.8%, which is higher than 92.7% of SceneGraph-VR and 91.5% of HBIM-Viewer, indicating that the courtyard outline, hall structure and decoration details can be continuously expressed after the joint modeling of multi-view reconstruction and semantic constraints. The mapping accuracy reaches 92.6%, which is 3.8 percentage points higher than that of XR-Guide. The task completion rate reaches 89.7%, which is significantly higher than 81.6% of ManualRoute. The average response delay is 76ms, which is lower than 88ms for VR-Link and 84ms for SceneGraph-VR.

Table 4: Comparison of results between the virtual inheritance system and different methods.

Method	Completeness / %	Mapping Accuracy / %	Completion Rate / %	Response Latency / ms
HBIM-Viewer	91.5	86.9	83.8	93
VR-Link	90.8	87.6	84.5	88
XR-Guide	92.1	88.8	86.2	81
SceneGraph-VR	92.7	90.4	87.5	84
ManualRoute	89.6	84.2	81.6	97
Proposed Method	94.8	92.6	89.7	76

As shown in Figure 8, the distribution differences of different methods in the two dimensions of task completion rate and response delay are obvious. The task completion rate of HBIM-Viewer is 83.8%, and the response delay is 93ms. The task completion rate of VR-Link is 84.5%, and the response delay is 88ms. XR-Guide achieves 86.2% and 81ms, respectively; SceneGraph-VR 87.5% and 84ms, respectively; ManualRoute has the lowest task completion rate (81.6%) and response delay (97ms). In contrast, the task completion rate of the proposed method reaches 89.7%, which is 2.2 percentage points higher than that of SceneGraph-VR, 5.9 percentage points higher than that of HBIM Viewer, and the response delay is reduced to 76ms, 5ms shorter than that of XR-Guide, and 21ms shorter than that of ManualRoute.

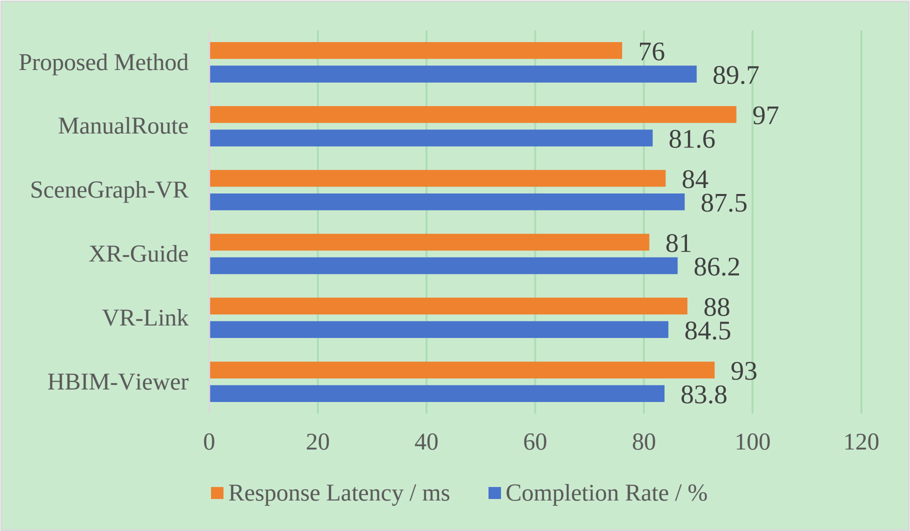

Figure 8: Comparison of different methods in completion rate versus response delay.

The data show that the static path method is difficult to maintain stable task advancement in complex traditional residential scenes. Although the system relying solely on scene display or fixed node links can complete basic browsing, there is still an obvious gap in real-time semantic scheduling and interactive feedback. The proposed method maintains a good balance between completion rate and delay, indicating that the cooperative mechanism of geometric reconstruction, cultural mapping and interactive feedback has formed a stable computational advantage.

3.4 Ablation experiments

In order to verify the contribution of the key links of the virtual inheritance system to the inheritance effect of traditional residential heritage, this paper carries out ablation experiments around the four modules of semantic mapping, path guidance, feedback update and multi-modal fusion. All experiments are conducted on the same sample of 286 houses, 4120 sets of interaction data and 13840 semantic annotations, and the training rounds and input scale are consistent with the task script. The results are shown in Table 5, and the full model is optimal in three metrics: mapping accuracy, task completion rate, and response delay. After removing the semantic mapping, the mapping accuracy decreased most obviously. After removing the path guidance, the task completion rate decreased more. After canceling the feedback update, the delay fluctuation increases. After removing the multimodal fusion, the backtracking performance is weakened. The results show that each module has a synergistic effect on spatial expression, knowledge triggering and interaction stability, and the complete structure is more suitable for virtual inheritance scenarios.

Table 5: Results of ablation experiments.

Configuration	Mapping Accuracy / %	Task Completion Rate / %	Average Response Latency / ms
Full Model	92.6	89.7	76
Without Semantic Mapping	88.3	84.9	79
Without Path Guidance	89.4	82.6	78
Without Feedback Update	90.1	85.8	84
Without Multimodal Fusion	90.7	86.2	81

From a comprehensive point of view, semantic mapping, path guidance, feedback update and multi-modal fusion are not isolated additional modules, but the core components to maintain the stable operation of the virtual inheritance system. When any link is removed, the efficiency of knowledge organization, the coherence of task advancement and the stability of interaction response in the scene will decrease to varying degrees. It can be seen that the complete model can better maintain the coordination relationship between spatial expression, content invocation and behavior feedback in the digital inheritance education scene of traditional residential architectural cultural heritage, and also provides a more stable experimental basis for the subsequent analysis of system performance and equipment adaptation.

3.5 Analysis of system operation performance and device adaptation

In order to evaluate the performance of the virtual inheritance system in actual deployment, this paper carries out loading and interaction tests on desktop terminals, head-mounted terminals and mobile terminals respectively, and synchronously records the average delay, task completion rate, loading success rate and resource occupancy. The test platform uses RTX4090 graphics card and i9 processor, the desktop runs the complete scene in Unity environment, the head mounted terminal uses Pico4 streaming mode, and the mobile terminal uses hierarchical loading and texture compression strategy. The results show that the average delay of desktop terminal is 63ms, head-mounted terminal is 76ms, and mobile terminal is 118ms. The task completion rates of the three types of devices reach 90.4%, 89.7% and 82.6%, respectively, indicating that the system still has good task support ability in the high-load immersion environment. This is shown in Figure 9.

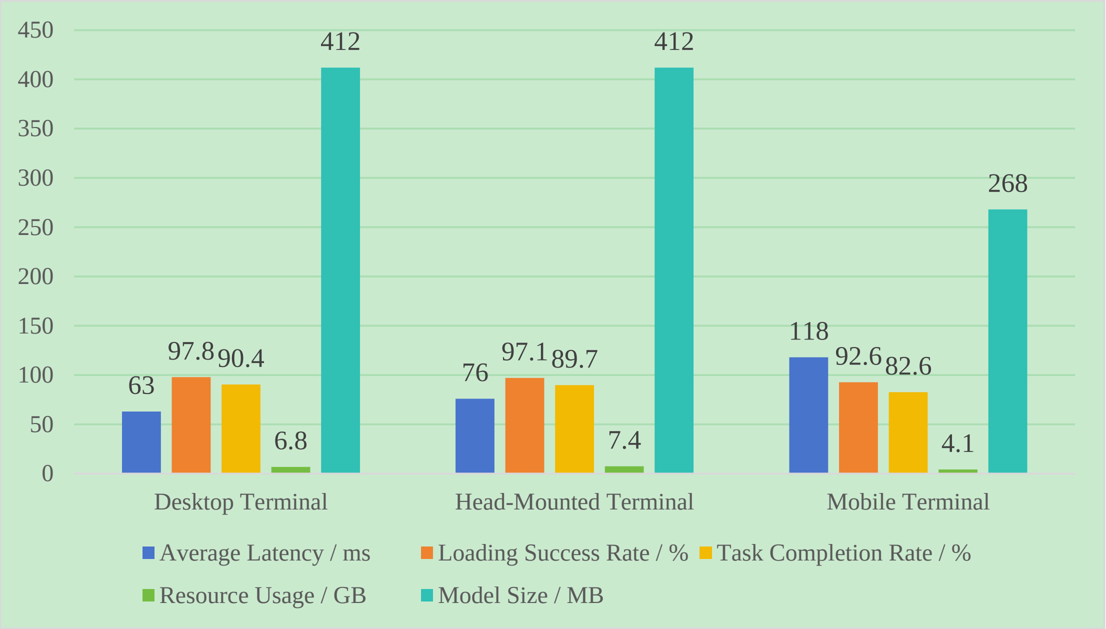

Figure 9: System operation and device adaptation results.

Figure 9 shows that the desktop has the most stable performance in terms of delay and completion rate, which is suitable for high-precision display and complex interaction tasks. Although the head-mounted terminal has higher computational overhead, the loading success rate and completion rate are close to the desktop terminal, which can meet the needs of immersive navigation and teaching. The mobile terminal is limited by computing power and memory, and the delay increases significantly, but it can still complete the basic inheritance task with the help of node pruning and asynchronous scheduling. On the whole, the system maintains good deployment elasticity on different devices. The system switch is stable, the operation process is continuous, the access link is complete, and the deployment adaptation is more reliable.

3.6 Discussion

Based on the above results, the virtual inheritance system constructed in this paper shows a good comprehensive ability in the digital inheritance education scene of traditional residential architectural cultural heritage. Whether it is the results of virtual scene construction, interactive application, method comparison and deployment analysis, the complete model maintains a relatively stable advantage. The scene integrity reaches 94.8%, the cultural content mapping accuracy reaches 92.6%, the task completion rate reaches 89.7%, and the average response time of the head-mounted terminal is controlled at 76ms, which shows that the geometric reconstruction, semantic organization and behavior feedback have formed an effective collaboration. Compared with the methods that only emphasize model display or fixed path jump, the proposed method incorporates component-level modeling, knowledge node mapping and task-driven interaction into a unified computing framework, so that the spatial form, structural relationship and cultural information in traditional dwellings can be continuously expanded in the same virtual environment. This processing not only enhances the interpretability of content expression, but also improves the user's coherent experience during browsing, recognition, and review. At the same time, the system still has local loading fluctuations in high-density decorative areas and mobile terminal conditions, which indicates that there is still room for further improvement in detailed granularity texture scheduling and lightweight deployment. In the future, graph structure coding, dynamic resource clipping and adaptive rendering mechanism can be combined to further enhance the stability and cross-platform adaptation ability in complex scenes. In addition, the current system has the engineering implementation foundation, which can provide a unified interface for digital archiving, immersive guide, knowledge learning and content update of traditional residential resources, making it closer to the sustainable operation of virtual heritage platform system form.

4 Conclusion

The cultural heritage of traditional residential buildings not only preserves the regional construction experience and spatial order, but also carries the material skills, life rituals and cultural memories. This paper focuses on the digital inheritance education in the immersive virtual environment, and constructs a virtual inheritance system composed of multi-source data acquisition, 3D reconstruction, semantic organization, interactive feedback and result evaluation, so that the courtyard pattern, component relationship and cultural content can be continuously expressed in a unified computing framework, forming a complete technical link for the digital inheritance of traditional dwellings.

In the system implementation, we use multi-view images and depth information to complete geometric reconstruction, and organize cultural content through component-level semantic annotation and knowledge node binding. Experimental results show that the integrity of 3D reconstruction reaches 94.8%, the accuracy of cultural content mapping reaches 92.6%, and the task completion rate reaches 89.7%. The average response delay of the head-mounted terminal is controlled at 76ms, and the response delay of the desktop terminal is 63ms. The comparison of different methods and ablation experiments further show that semantic mapping, path guidance, feedback update and multi-modal fusion jointly support spatial expression, content invocation and interaction stability in the process of virtual inheritance.

At the same time, the current system still has some limitations. The component difference between cross-regional samples will affect the semantic transfer effect. The coverage of complex residential types by existing data scale still needs to be extended. In the future, technologies such as lightweight reconstruction, graph structure learning, adaptive scheduling, cross-terminal rendering collaboration and dynamic resource cutting can be further promoted, so that the digital inheritance education of traditional residential architectural cultural heritage can achieve a more coordinated performance between authenticity, stability and engineering adaptability, and provide more reliable support for resource update, system expansion and platform deployment. At the same time, the application boundary and deployment conditions of the system are further clarified.

References

[1] Boboc R G, Băutu E, Gîrbacia F, et al. Augmented reality in cultural heritage: an overview of the last decade of applications[J]. Applied Sciences, 2022, 12(19): 9859.https://doi.org/10.3390/app12199859

[2] De Paolis L T, Chiarello S, Gatto C, et al. Virtual reality for the enhancement of cultural tangible and intangible heritage: The case study of the Castle of Corsano[J]. Digital Applications in Archaeology and Cultural Heritage, 2022, 27: e00238. https://doi.org/10.1016/j.daach.2022.e00238

[3] Hulusic V, Gusia L, Luci N, et al. Tangible user interfaces for enhancing user experience of virtual reality cultural heritage applications for utilization in educational environment[J]. ACM Journal on Computing and Cultural Heritage, 2023, 16(2): 1-24. https://doi.org/10.1145/3593429

[4] Süvari A, Okuyucu Ş E, Çoban G, et al. Virtual reconstruction with the augmented reality technology of the cultural heritage components that have disappeared: The Ayazini Virgin Mary Church[J]. ACM Journal on Computing and Cultural Heritage, 2023, 16(1): 1-16.https://doi.org/10.1145/3579361

[5] Lucifora C, Schembri M, Poggi F, et al. Virtual reality supports perspective taking in cultural heritage interpretation[J]. Computers in Human Behavior, 2023, 148: 107911.https://doi.org/10.1016/j.chb.2023.107911

[6] Alatrash S, Arnab S, Antlej K. Communicating engineering heritage through immersive technology: A VR framework for enhancing users' interpretation process in virtual immersive environments[J]. Computers & Education: X Reality, 2023, 3: 100040. https://doi.org/10.1016/j.cexr.2023.100040

[7] Innocente C, Ulrich L, Moos S, et al. A framework study on the use of immersive XR technologies in the cultural heritage domain[J]. Journal of Cultural Heritage, 2023, 62: 268-283.https://doi.org/10.1016/j.culher.2023.06.001

[8] Lucchi E. Digital twins for the automation of the heritage construction sector[J]. Automation in Construction, 2023, 156: 105073. https://doi.org/10.1016/j.autcon.2023.105073

[9] Izaguirre J I, Ferrari A A, Acuto F A. Exploring cultural heritage and archaeological research from a VR-based approach[J]. Digital Applications in Archaeology and Cultural Heritage, 2024, 32: e00318.https://doi.org/10.1016/j.daach.2024.e00318

[10] Zhang J, Wan Yahaya W A J, Sanmugam M. The impact of immersive technologies on cultural heritage: A bibliometric study of VR, AR, and MR applications[J]. Sustainability, 2024, 16(15): 6446.https://doi.org/10.3390/su16156446

[11] Rafeiro J, Tomé A, Nazário M. Immersive learning for lost architectural heritage: interweaving the past and present, physical and digital in the monastery of Madre de Deus[J]. Sustainability, 2024, 16(3): 1156.https://doi.org/10.3390/su16031156

[12] Rodriguez-Garcia B, Guillen-Sanz H, Checa D, et al. A systematic review of virtual 3D reconstructions of Cultural Heritage in immersive Virtual Reality[J]. Multimedia Tools and Applications, 2024, 83(42): 89743-89793. https://doi.org/10.1007/s11042-024-18700-3

[13] Wang Y, Wang S, Pan Y, et al. Immersive virtual reality and computer vision for heritage: visual evaluation and perception of the industrial heritage sites along the Yunnan–Vietnam railway (Yunnan section)[J]. Heritage Science, 2024, 12(1): 1-15. https://doi.org/10.1186/s40494-024-01145-0

[14] Tini M A, Forte A, Girelli V A, et al. Scan-to-HBIM-to-VR: An integrated approach for the documentation of an industrial archaeology building[J]. Remote Sensing, 2024, 16(15): 2859.https://doi.org/10.3390/rs16152859

[15] Penjor T, Banihashemi S, Hajirasouli A, et al. Heritage building information modeling (HBIM) for heritage conservation: Framework of challenges, gaps, and existing limitations of HBIM[J]. Digital Applications in Archaeology and Cultural Heritage, 2024, 35: e00366.https://doi.org/10.1016/j.daach.2024.e00366

[16] Ramtohul A, Khedo K K. Augmented reality systems in the cultural heritage domains: A systematic review[J]. Digital Applications in Archaeology and Cultural Heritage, 2024, 32: e00317.https://doi.org/10.1016/j.daach.2024.e00317

[17] Buragohain D, Meng Y, Deng C, et al. Digitalizing cultural heritage through metaverse applications: challenges, opportunities, and strategies[J]. Heritage Science, 2024, 12(1): 1-16.https://doi.org/10.1186/s40494-024-01403-1

[18] Corrales-Serrano M, Merchán P, Merchán M J, et al. Virtual reality applied to heritage in higher education—validation of a questionnaire to evaluate usability, learning, and emotions[J]. Heritage, 2024, 7(6): 2792-2810.https://doi.org/10.3390/heritage7060132

[19] Zheng H, Chen L, Hu H, et al. Research on the digital preservation of architectural heritage based on virtual reality technology[J]. Buildings, 2024, 14(5): 1436. https://doi.org/10.3390/buildings14051436

[20] Iacono S, Scaramuzzino M, Martini L, et al. Virtual reality in cultural heritage: a setup for balzi rossi museum[J]. Applied sciences, 2024, 14(9): 3562. https://doi.org/10.3390/app14093562