Sensors and Materials, Vol. 37, No. 10 (2025) 4657–4675
MYU Tokyo

S & M 4205

Artificial Intelligence-based Immersive Virtual Reality Technology in Digital Dissemination of Intangible Cultural Heritage

Xuehong Zhao,[1] Mingyu Zhao,[2*] Hailing Wang,[3] and Yanshan Zhou[4]

[1]Harbin Finance University; Harbin 150030, China
[2]Heilongjiang Academy of Sciences; Harbin 150001, China
[3]Heilongjiang University of Science and Technology; Harbin 150022, China
[4]Harbin University of Science and Technology; Harbin 150040, China

(Received June 5, 2025; accepted October 20, 2025)

Keywords: artificial intelligence, virtual reality, 3D, intangible cultural heritage, immersive experience

Intangible cultural heritage (ICH) carries rich human wisdom and cultural value through its "living" form of craftsmanship and performances. However, it is difficult to preserve and disseminate ICH and its values through traditional means. Therefore, we developed an AI-based immersive virtual reality (VR) digital dissemination system (AI+VR system). In the system, we restored the damaged parts and patterns of ICH images by using adversarial training, objective function, and convergence conditions of generative adversarial networks (GANs), a neural radiance field (NeRF) mathematical model, and a six-degree-of-freedom (6-DoF) kinematic model. The NeRF model was constructed for 3D view synthesis using the volume rendering equation based on the image's color density, while the 6-DoF kinematic model was established using visual-inertial odometry (VIO) to track viewer perspectives in VR scenes. The outcomes of the AI+VR system were compared with those of traditional ICH display methods (2D videos, images, and text). The developed AI+VR system enhanced real-time interaction frame rates, reduced latency, and automated modeling speed. The system's hardware adaptability and resource consumption were validated. The AI+VR system can be used for content creation and the experience economy of the cultural industry. The integration of AI and VR in content creation serves as a new means for the dissemination of ICH, proving that technological empowerment can be used for preserving cultural inheritance.

1. Introduction

Intangible cultural heritage (ICH) embodies the ethnic traditions and artistry passed down through generations, preserved in oral expressions and craftsmanship. Defined by its dynamic and fleeting nature, ICH evolves while maintaining its cultural essence. In contrast, tangible cultural heritage is preserved by physical objects and written records, which endure for millennia. However, ICH often lacks intuitive carriers, making its preservation and dissemination highly challenging. Two-dimensional video recordings, photographs, and image-text descriptions have primarily been used for disseminating ICH. Although these tools are still

*Corresponding author: e-mail: zhaomingyu2794@dingtalk.com
https://doi.org/10.18494/SAM5804

ISSN 0914-4935

widely used, they lack the immersiveness and interactivity to provide viewers with a detailed experience, reducing their effectiveness in disseminating ICH. Additionally, the inheritance of ICH becomes difficult as inheritors are aging, and younger generations are not interested in its inheritance. To better preserve and disseminate ICH, public interest must be enhanced to maintain cultural identity.

Recently, advanced technologies, such as virtual reality (VR), have offered solutions for ICH dissemination. VR allows viewers to immerse themselves in digitally reconstructed ICH and experience traditional techniques or performances from a first-person perspective. Such an immersive experience transcends temporal and spatial constraints and is useful in preserving cultural memories in a digital form. Through VR, viewers can immerse themselves in history, experiencing the traditions and artistry of ICH. However, constructing a virtual ICH has technical challenges. First, high-quality 3D content is difficult to create. Second, VR applications require high frame rates and low latency to prevent viewers' motion sickness. Third, users' free movement in VR requires precise 6-DoF positioning and tracking.

However, AI can be used to address such challenges. Deep learning, a subset of AI, enables convenient image generation and restoration, 3D reconstruction, and sensor data fusion, enabling the digitization of ICH content and optimization of interactive performance. In particular, sensor data is used in image restoration as it provides depth and texture details, allowing AI models to reconstruct missing or degraded parts of an image. Multisensor fusion techniques enhance image quality and generate realistic textures by integrating data from different sources [e.g., RGB cameras, depth sensors, and light detection and ranging (LiDAR)] and creating highly detailed 3D models using data from sensors such as the global navigation satellite system (GNSS), inertial measurement unit (IMU), LiDAR, and unmanned aerial vehicles (UAVs) to create highly detailed 3D models. Deep learning models such as local feature transformer (LoFTR) and NeuralRecon enable image matching and point cloud generation for accurate 3D reconstruction.[1]

In this study, we applied AI-based VR technology for the dissemination of ICH. First, we restored damaged ICH images and patterns using generative adversarial networks (GANs) along with loss functions and training convergence mechanisms.[2] We also introduced the mathematical equations of neural radiance fields (NeRF) to reconstruct ICH scenes. Using the volume rendering equation, high-fidelity scenes were generated. For the real-time positioning and tracking of VR objects, a 6-DoF kinematic model was constructed using visual-inertial odometry (VIO). The performance of the AI-based immersive VR digital dissemination system (AI+VR system), consisting of GANs, NeRF, and VIO, was evaluated in terms of latency, frame rate, and modeling speed. The system demonstrated superior performance to traditional 2D videos and static image-text methods. Through an industry–academia–research collaboration, we developed "Virtual ICH Experience Halls" and mobile VR ICH applications using the developed AI+VR system.

The system facilitates content co-creation, copyright protection, and commercialization, advancing the experience economy. Additionally, it highlights AI's role in preserving and transmitting ICH, emphasizing the importance of technology integration in its dissemination.

2. Technology Review

Through a literature review, we analyzed the characteristics of ICH digital dissemination, immersive experience theory, embodied communication theory, cultural representation theory, and digital twin to design the AI+VR system for the effective dissemination of ICH.

2.1 Digital dissemination of ICH and immersive experience

The inheritance of ICH relies on practice and interaction. However, traditional dissemination methods fail to provide on-site experience, making it difficult to foster emotional and physical engagement in ICH and to be broadly disseminated.[3] Recently, VR has been introduced as a solution, offering an immersive approach to ICH preservation and accessibility.[4] VR enables viewers to immerse themselves in ICH and its related cultural environments, overcoming temporal and spatial limitations. Even without the physical presence of ICH, VR provides deep engagement and accessibility. Immersion in ICH facilitates emotional resonance and cognitive engagement with ICH, enabling viewers to understand its spiritual and cultural significance.[5] Through a high degree of realistic interaction, the immersive experience improves the effect of digital dissemination of ICH as viewers feel as if they are present. To provide such experience, a highly interactive design and appropriate cultural scenarios must be created to deliver multisensory stimulation (visual, auditory, and tactile feedback). Such digital means effectively provide immersive experiences in ICH, which fosters the effective dissemination of ICH.

2.2 Embodied communication theory and immersive interaction

Embodied communication theory emphasizes the role of the human body in communication, positing that knowledge transmission relies on the physical presence and participation of the communicator. In ICH inheritance, craftsmanship techniques and cultural connotations constitute "tacit knowledge" that is difficult to articulate and is inherently tied to the bodily practices of inheritors.[6] Schreurs proposed the concept of tacit knowledge, with which he claimed that ICH craftsmanship techniques can be mastered through physical demonstration and practice.[7] Embodiment is critical in ICH dissemination and knowledge transmission as it enables the physical engagement of the one that inherits through motion imitation and posture interaction. Written or visual archives alone cannot be used to reconstruct the situational context of ICH inheritance.[8] In accordance with the embodied communication theory, body-interactive elements were included in the virtual environments to simulate the inheritor's presence in the developed AI+VR system. For example, motion capture or gesture recognition enables users to embody virtual inheritor roles and practice ICH craftsmanship techniques through an immersive "learning-by-doing" effect. This embodied and immersive interactive experience enables inheritees to perceive and comprehend the essence of ICH craftsmanship through virtual body movements. This aligns with the embodied cognition theory, in which the integration of bodily perception and movement is emphasized as a cognitive process. Appropriately designed interactions in virtual environments enhance users' comprehension and retention of the learned

content.[9] Therefore, the embodied communication theory provides a human-centered theoretical foundation for VR ICH experiences and is the basis of immersive system design that prioritizes natural interaction. By enabling users to "learn through doing", the AI+VR system fosters a deeper appreciation of ICH's intrinsic beauty.

2.3 Cultural representation theory and digital media

Cultural representation or cultural reproduction theory focuses on how cultural content is reconstructed and interpreted through different media.[10] In the digitalization of ICH, it is challenging to maintain the authenticity and integrity of cultural expressions. Traditional recording methods statically preserve surface-level information, while the deeper meanings and values of ICH need to be manifested separately. When reconstructing ICH through digital media, it is essential to preserve its original context and interactive elements, enabling viewers to experience a culturally authentic encounter. Therefore, ICH in the virtual environment must have resemblance and spiritual likeness, presenting craftsmanship and conveying its cultural essence.[11] For example, when digitally reconstructing a dance performance, the original performance's ambiance (music, costumes, props, spatial arrangement, etc.) and the interactive dynamics between audience and dancers must be delivered in the virtual environment. Only then can audiences genuinely grasp the cultural significance of the dance. Digital technologies provide powerful tools for such cultural representation. Through 3D modeling, performance venues and props are reconstructed, and the performers' movements are reproduced through motion capture and animation. AI is used to generate accurate scenes based on historical data.[12] However, simple replication is not enough to present artistic details without distortion or oversimplification. Applying cultural reproduction theory, we introduced the authenticity and integrity of ICH in constructing the VR immersive environment, with the help of ICH experts, to present the "form" and the "spirit" of ICH. On the basis of the theory, the AI+VR system was developed to be evidence-based to enhance the credibility and cultural richness of ICH.

2.4 Digital twin in VR

Digital twin refers to the real-time mapping and simulation of the state and behavior of physical entities in the virtual space of digital models.[13] In the preservation of ICH, the concept of digital twin is extended to the digital replication of cultural assets and their associated knowledge systems.[14] The digital twin model emphasizes dynamic correlation and high-fidelity representation between virtual and real-world objects. Through multisource data fusion and real-time updates, virtual digital twin models evolve synchronously with their physical counterparts.[15] As ICH's cultural form is characterized by "living" attributes, it requires a digital twin model. In this study, we established a heritage digital twin (HDT) to include the 3D models of tools and venues, the process representation of craftsmanship techniques, and historical context and meanings.[16] The digital twin provides virtual "avatars" of ICH in the virtual environment for the digital reproduction and monitoring of ICH.[17] When integrated with IoT and AI technologies, digital twins acquire the real-time data of ICH inheritance

activities in the real world (such as movements captured by sensors and environmental parameters) and map them into the virtual environment to maintain consistency and interactivity.[18] We incorporated a digital twin into the system architecture to map performers, interactive objects, and environmental elements. For example, wearable sensors were employed to capture the movements and physiological responses of inheritors, enabling avatars to synchronize their performances with precision. Environmental sensors were used to collect data, including temperature, humidity, and sound, to adjust the virtual environment to enhance the viewer's immersion. The digital twin model enables the integration of virtuality and reality and ensures the reliability of the virtual environment and capacity for prompt updates. In the AI+VR system, users participate in ICH activities virtually through interaction[19] and interact with digital inheritors through dialogue to learn craftsmanship techniques or collaborate to complete artworks.[20] This demonstrates the value of digital twins in immersive ICH dissemination using the AI+VR system. The system provides ICH practices and innovative inheritance and experiential means in the virtual environment .[21]

2.5 Integration of AI and VR

AI technology enables content generation and intelligent interaction in VR. The integration of AI and VR for digital ICH dissemination addresses inherent challenges in traditional VR content production and interaction. GANs are used for image restoration, content generation, and reconstructing deteriorated ICH images by automatically generating virtual scene textures. NeRF supports 3D reconstruction by learning radiance properties from photos or videos to produce high-fidelity 3D virtual scenes. VIO combines computer vision and inertial sensing for autonomous positioning and tracking, enabling VR systems to track an object's 6-DoF pose. These technologies help users move freely and be precisely positioned in virtual environments. AI algorithms significantly enhance VR systems' capabilities of reconstructing complex ICH scenes and real-time interaction. Compared with traditional algorithms, deep-learning-based image processing algorithms markedly improve the quality and speed of scenes and interactions in them. NeRF recovers 3D scene information from unstructured data and eliminates the need for cumbersome laser or structured-light scanning. VIO enables VR or augmented reality (AR) devices to provide precise self-localization without additional positioning devices, which provides immersive experiences. The integration of such AI technologies in VR platforms is the technical backbone of the AI+VR system. In this study, we employed GAN-based image restoration, NeRF-based 3D reconstruction, and VIO-based positioning and tracking for immersive ICH dissemination.

3. Methods

We designed the system architecture and module composition to incorporate GAN, NeRF, and VIO. Moreover, the system performance was compared with that of traditional methods in terms of accuracy, speed, and hardware requirements.

3.1 System architecture

The AI+VR system comprised the image restoration module, 3D reconstruction module, positioning and interaction module, and rendering and display module. In the mage restoration module, GANs were used to restore and enhance ICH images.[22] GANs removed the noise and damage in an image and recovered a high-definition image. In the 3D reconstruction module, NeRF converted multiview photos into high-fidelity 3D virtual scenes.[23] In the positioning interaction module, a VIO algorithm was employed for the autonomous positioning and trajectory tracking of users.[24] In the rendering and display module, virtual scenes were created in real time, responding to user interaction detected by the head-mounted VR device (Fig. 1).

The AI+VR system captured the inheritor's physical movements and multimedia data (historical photographs, videos, sensor data, etc.) and processed them using GANs for NeRF reconstruction. The processed images and constructed 3D models were imported into the VR engine to construct immersive scenes. The VIO of the positioning interaction module tracked and mapped the 6-DoF movements of the head and hands to update the corresponding viewpoint of avatars. The system rendered virtual environments at high frame rates, enabling users to freely view restored historical scenes, examine digitally reconstructed ICH artifacts, or learn movements by following an avatar's demonstrations. In the system, AI technologies ensured

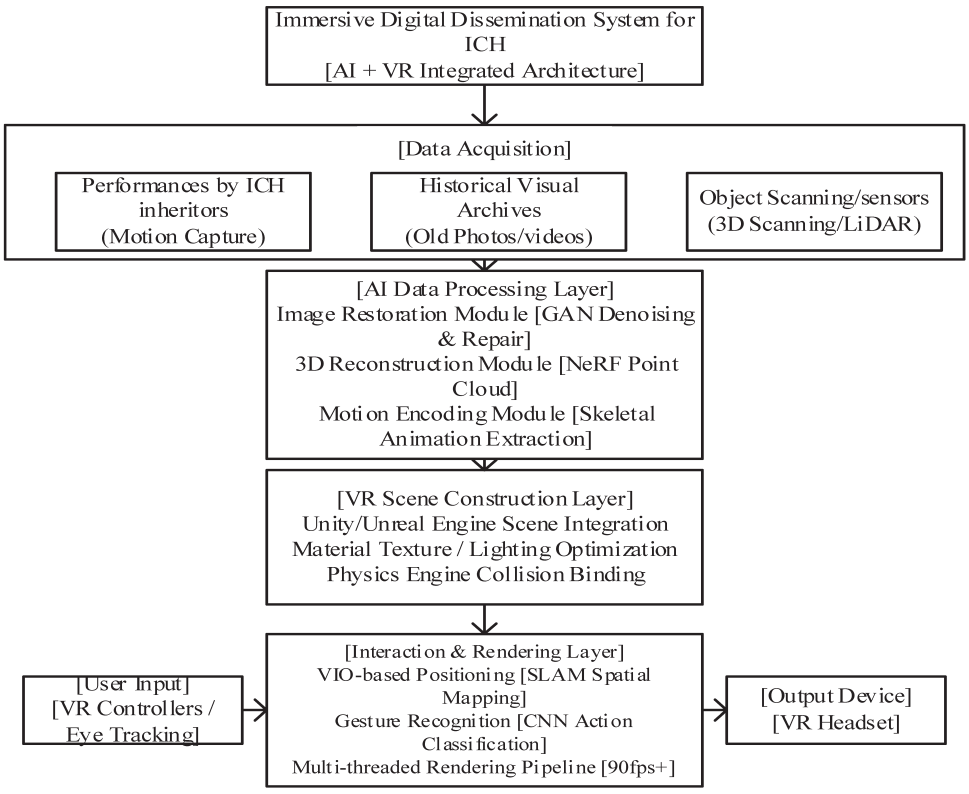

Fig. 1. Architecture of developed AI+VR system.

content authenticity and high quality, and immersion and engagement through interactions in the virtual environment.

3.2 Restoration of image and pattern

The preservation of ICH involves the restoration and reconstruction of images and patterns in images. Blemishes and damaged areas in old black-and-white photographs and traditional patterns need to be colored. These tasks are usually performed using GANs. GAN consists of a generator and a discriminator for image generation through adversarial training. The generator takes random noise or incomplete images as input and outputs restored images, while the discriminator distinguishes whether the input is a real image or a generated one. Their objective function is formulated as a min-max optimization problem. The generator minimizes the discriminator's recognition ability, while the generator maximizes its discrimination accuracy. The value function of GAN is expressed as

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}\left(\log D(x)\right) + E_{z\sim p_z(z)}\left(\log\left(1-D\left(G(z)\right)\right)\right), \tag{1}$$

where $V(D, G)$ is the value function that both networks optimize, $E_{x\sim p_{data}(x)}$ is the expected value over all real data samples (x) drawn from the real data distribution (p_{data}), $\log D(x)$ is the discriminator's judgment on real data, $D(x)$ is the probability that D classifies the real image x as real, $E_{z\sim p_z(z)}$ is the expected value over all noise samples (z) drawn from the noise distribution (p_z), which is a uniform or Gaussian distribution, $D(G(z))$ is the discriminator's judgment on fake data, $G(z)$ is the image generated by the generator from the noise vector z, and $\min_G \max_D V(D,G)$ represents the minimum-maximum game.

The discriminator (D) tries to maximize $V(D, G)$ for better classification, while the generator (G) tries to minimize $V(D, G)$ to fool the discriminator This makes the term $\log(1 - D(G(z)))$ close to $\log(1) = 0$, thus maximizing the overall value function $V(D, G)$. In training, the generator's and discriminator's loss functions (L_G and L_D) are reformulated as follows.

$$L_G = -E_{z\sim p_z}\left(\log D\left(G(z)\right)\right) \tag{2}$$

$$L_D = -E_{x\sim p_{data}}\left(\log D(x)\right) - E_{z\sim p_z}\left(\log\left(1-D\left(G(z)\right)\right)\right) \tag{3}$$

Through alternating optimization of L_G and L_D, the pixel distribution of generated images converges toward that of real images. Eventually, the discriminator output approaches 0.5, indicating that the generated images become indistinguishable from real ones, that is, Nash equilibrium is achieved. In image restoration, the GAN architecture is enhanced by incorporating a convolutional neural network (CNN) structure (e.g., U-Net) with self-attention mechanisms (e.g., transformer modules). Such enhancement strengthens the model's capacity to capture complex textures and global stylistic patterns. In the loss function design, the reconstruction loss

(L1/L2 norm) is introduced to ensure the consistency of the known regions and minimize the fundamental adversarial loss. The perceptual and style losses are used to optimize the subjective quality of images. The total loss function (L_{total}) for a mural restoration model is

$$L_{total} = \alpha L_{rec} + \beta L_{adv}, \tag{4}$$

where L_{rec} represents the loss function of reconstruction error, L_{adv} denotes the loss function of the adversarial loss, and α and β are weighting coefficients.

The employed model in the developed AI+VR system outperformed traditional algorithms in mural sample restoration. The peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) were significantly improved, ensuring image clarity and texture coherence (Table 1). The restored and real images are presented in Fig. 2.

GAN demonstrated superior restoration quality over the traditional method (PatchMatch). When handling complex textures or large-area defects, GAN reconstructed coherent and reasonable content by leveraging learned global semantic information, whereas PatchMatch

Table 1
GAN-based image restoration and algorithm performances.

Method	Accuracy (PSNR/SSIM)	Processing speed (for 512 images)	Hardware demands
GAN restoration	30.5 dB/0.92	0.5 s (GPU accelerated)	Graphics processing unit (GPU) [4 GB video random access memory (VRAM)]
Traditional restoration	28.1 dB/0.88	2.0 s (CPU single-thread)	Central processing unit (CPU) only (sufficient RAM)

For PSNR, higher values are better, while for SSIM, lower values are better. The data represent average values from the test set.

(a) (b)

Fig. 2. (Color online) Images restored using (a) traditional method and (b) GAN.

produced distorted patchwork results. While GAN-based inference was successfully performed in near real time with GPU acceleration, its implementation required extensive training on large sample datasets and significant computational resources. However, for relatively static images, trained GANs repeatedly conduct more efficient batch restoration. In processing 4-4-K-resolution images, PatchMatch's memory usage increased but remained manageable, while GAN required image tiling, thereby increasing implementation complexity. GAN-based restoration was efficient for digital content preservation and high-quality restoration. GAN also reconstructed immersive virtual scenes effectively.

3.3 NeRF

For the immersive presentation of ICH scenes in the virtual environment, high-fidelity 3D digital scenes are required. Traditional 3D modeling methods, such as panoramic photography, geometric modeling, and photogrammetry, require high production costs, long development cycles, or limited viewing angles. Therefore, NeRF is widely used as an emerging implicit 3D modeling technology for high-precision scene reconstruction. NeRF uses a multilayer perceptron (MLP) to construct a function F_Θ. For any spatial point $\boldsymbol{x} = (x, y, z)$ and viewing direction d, the function outputs a color $\boldsymbol{c} = (r, g, b)$ and volume density σ, thereby modeling the scene through volume rendering. For any camera ray $r(t) = \boldsymbol{o} + z\boldsymbol{d}$, the corresponding color in the image is computed using the following equations.

$$C(\boldsymbol{r}) = \int_{t_n}^{t_f} T(t)\sigma(\boldsymbol{r}(t))c(\boldsymbol{r}(t), \boldsymbol{d})dt \tag{5}$$

$$T(t) = \exp\left(-\int_{t_n}^{t} ds\right) \tag{6}$$

Here, t_f and t_n represent the near and far bounds of the ray's intersection with the scene, respectively, $T(t)$ denotes transmittance, indicating the probability that the ray remains unabsorbed while traversing the scene, σ is the volume density function, and $\boldsymbol{c}$ is the color radiance function. All sample points are included to compute the final pixel color, and the integration process of the points is approximated using discrete sampling.

$$\hat{C}(\boldsymbol{r}) = \sum_{i=1}^{N} T_i \left(1 - e^{-\sigma_i \delta_i}\right) c_i \tag{7}$$

$$T_i = \prod_{j=1}^{i-1} \exp\left(-\sigma_j \delta_j\right) \tag{8}$$

Monte Carlo integration is used through stochastic sampling for an optimal balance between efficiency and accuracy. NeRF is trained by high-fidelity 3D models using multiview images, and outperforms traditional methods in detail reproduction and lighting reconstruction. For

example, in the traditional image, NeRF reconstructs the patterns and artifacts, renders them from any angle in VR, and effectively enhances the users' immersion experience. Compared with conventional photogrammetry, NeRF eliminates the need for explicit mesh modeling while maintaining geometric precision, offering superior flexibility and texture continuity. Meanwhile, NeRF significantly improves the training speed. For the 3D digitization of images and objects, structured light scanning or LiDAR is used in traditional methods to acquire high-precision point clouds and reconstruct mesh models. While achieving millimeter-level accuracy of 1–2 mm, the traditional methods suffer from high hardware costs, complex operation procedures, and stringent environmental requirements. In contrast, NeRF utilizes multiple photographs taken by ordinary cameras to establish neural network mappings from coordinates to color/density, enabling high-fidelity 3D reconstruction and novel view synthesis without explicit mesh generation. In this study, NeRF reconstruction showed an average geometric error of 18.6 mm and superior visual realism even under less-structured lights. In structured light scanning, tens of minutes are typically needed for data acquisition and point cloud registration. In contrast, NeRF with accelerated algorithms such as instant neural graphics primitives (NGP) completes training in 5 min to render novel views at ≥30 frames per second (FPS), showing significant modeling efficiency. For structured light scanning, specialized projectors, high-resolution cameras, and stable environments are necessary, along with substantial CPU/RAM for data processing. However, NeRF only requires GPUs (RTX 3090 with approximately 6 GB of VRAM in this study) and modest storage of tens of megabytes. Therefore, NeRF considerably lowers the hardware demands and operation complexity while maintaining high visual fidelity, making it appropriate for large-scale image digitization. Table 2 presents the comparison of NeRF and traditional methods, and Fig. 3 shows the images processed by these methods. Although traditional structured light scanning showed better accuracy, which is appropriate for precise quantification, NeRF excelled in texture and lighting reconstruction, delivering highly realistic rendering effects with a lower geometric accuracy. NeRF enables the simplicity of operation and low cost, as it needs only ordinary cameras and GPUs. This makes NeRF ideal for on-site constraints or situations where artifacts cannot be touched. Structured light scanning requires specialized equipment and longer processing times, which is appropriate for small-scale, high-precision data acquisition. Considering the advantages, we employed NeRF for constructing immersive scenes, while structured light scanning was used for artifact modeling. In post-processing refinements to optimize details of the images, an effective balance between efficiency and accuracy was achieved.

Table 2
Performances of NeRF and traditional method.

Method	Reconstruction accuracy (geometric error)	Processing speed	Hardware demands
NeRF	Average error of 18.6 mm	5 min training, >30 FPS rendering	Ordinary camera + high-performance GPU
Structured light scanning	Average error of 1–2 mm	Tens of minutes (scanning + reconstruction)	Structured light projector + professional camera/compute cluster

(a) (b)

Fig. 3. (Color online) Images restored by (a) NeRF and (b) structured light scanning.

3.4 VIO positioning and tracking

The interactive experience in immersive VR relies on the free motion modeling and real-time positioning and tracking of a user's head and body. To achieve high-precision 6-DoF positioning, VIO was used as it enables stable positioning without external devices, simply by fusing data from cameras and IMUs. The state vector of a VIO system typically comprises a position $\boldsymbol{p}(t) \in R^3$, velocity $\boldsymbol{v}(t) \in R^3$, and orientation rotation matrix $\dot{\boldsymbol{R}}(t) = SO(3)$ [or equivalently represented by the quaternion $\overline{q}(t)$ to denote the system's 3D spatial orientation], along with gyroscope bias $\boldsymbol{b}_g$ and accelerometer bias $\boldsymbol{b}_a$. The calculation model is defined as follows.

- Orientation update

$$\dot{\boldsymbol{R}}(t) = \boldsymbol{R}(t)\left(\boldsymbol{\omega}(t)\right)_{\times} \tag{9}$$

$$\boldsymbol{R}_{k+1} = \boldsymbol{R}_k \exp\left(\left(\boldsymbol{\omega}_k - \boldsymbol{b}_g\right)\Delta t\right) \tag{10}$$

- Velocity update

$$\dot{\boldsymbol{v}}(t) = \boldsymbol{R}(t)\left(\boldsymbol{a}(t) - \boldsymbol{b}_a\right) + \boldsymbol{g} \tag{11}$$

$$\boldsymbol{v}_{k+1} = \boldsymbol{v}_k + \left(\boldsymbol{R}_k\left(\boldsymbol{a}_k - \boldsymbol{b}_a\right) + \boldsymbol{g}\right)\Delta t \tag{12}$$

- Position update

$$\dot{\boldsymbol{p}}(t) = \boldsymbol{v}(t) \tag{13}$$

$$\boldsymbol{p}_{k+1} = \boldsymbol{p}_k + \boldsymbol{v}_k \Delta t + \frac{1}{2}\left(\boldsymbol{R}_k\left(\boldsymbol{a}_k - \boldsymbol{b}_a\right) + \boldsymbol{g}\right)\Delta t^2 \tag{14}$$

The model estimates short-term motion states through IMU integration, but suffers from error accumulation due to sensor noise and biases. To mitigate this, VIO incorporates visual feature data for state correction. The visual component is used to compute relative pose either with feature point tracking + PnP algorithms or direct photometric error minimization, and then fuses the IMU prediction results through extended Kalman filter (EKF) or sliding-window optimization for high-precision trajectory reconstruction.

In immersive VR systems, accurate user position and orientation tracking are critical. Therefore, we employed VIO to fuse IMU data with visual feature tracking data for accurate 6-DoF positioning when using head and handheld devices. The results were compared with those using external base station positioning (e.g., HTC Vive Lighthouse). The external devices showed submillimeter-level precision [a standard deviation (SD) of 0.5 mm], while VIO demonstrated a positional error of 3−5 mm in indoor scenes. Despite larger errors, the precision level of VIO meets the requirements for immersive experiences. Systems with external devices and VIO showed a position update rate of 100 Hz, with VIO's edge-computing architecture enabling tracking latency of below 10 ms. In terms of environmental adaptability, external positioning is limited by base station coverage, while VIO depends on scene textures and demonstrates enhanced robustness using closed-loop simultaneous localization and mapping (SLAM) algorithms. VIO's peripheral-free design significantly enhances portability and adaptability, making it particularly appropriate for mobile applications, such as museum exhibitions and outdoor displays. Table 3 shows the performances of the systems with external devices and VIO.

VIO and external devices exhibited distinct advantages and limitations. Using the external devices enabled accurate but complicated methods to deploy, while VIO enabled flexible applications. In this study, millimeter-level tracking discrepancies were negligible to user perception. The deviations of a few millimeters cannot be felt when viewing virtual artifacts. Conversely, the unconstrained mobility of VIO enhances immersive quality and practical

Table 3
Performances of VIO and external devices.

Method	Positioning accuracy	Tracking frequency/ latency	Hardware and environmental demands
VIO	3−5 mm	Approximately 100 Hz; up to 10 ms latency	HMD-integrated stereo cameras + IMU; Requires textured environments
External devices	<1 mm	Typically ≥100 Hz; 5−15 ms latency	External positioning devices (laser base stations/cameras); Requires installation and calibration

usability, which makes VIO a more appropriate solution. Therefore, we adopted VIO in ICH dissemination in this study.

4. Results and Discussion

We evaluated the performance of the AI+VR system for system latency, frame rate, interaction responsiveness, model loading speed, and GPU memory utilization.

4.1 Experiment and parameters

To evaluate the system performance, we used a computer configured with an Intel i7-11700KF CPU, 32 GB of RAM, and an NVIDIA RTX 3080 GPU (10 GB of VRAM) running on Windows 10. The VR device was an all-in-one head-mounted display with a refresh rate of 90 Hz and a per-eye resolution of 1832 × 1920. The device was equipped with built-in stereo cameras and IMU sensors for VIO-based 6-DoF real-time positioning and tracking. The platform was developed using Unity Engine 2021 to construct immersive content. GAN-restored images and a NeRF-generated 3D model were loaded into the Unity Engine to render images, scenes, and interaction. The VIO positioning data built into the VR device controlled the movement of virtual cameras in real time.

The selected scenes in the experiment were Dunhuang Murals in panoramic views and localized details, as illustrated in Figs. 4 and 5. Figure 4 shows a digitally restored Dunhuang Mural, for which GAN was used to provide texture details and colors in high resolution. NeRF was used for 3D restoration, enabling viewers to examine the murals from multiple angles in the virtual environment. Figure 5 shows the details of mural figures in the restored image (the right image). The original damaged image is also presented in the figure. GAN was used to restore the

Fig. 4. (Color online) Digital restoration of Dunhuang murals.

Fig. 5. (Color online) Detailed close-up of Dunhuang mural figures.

details and stylistic consistency in the virtual environment. The virtual scene contained more than 500000 vertices in a 3D model in a nm image size of about 200 MB. A dynamic loading module enables seamless transitions between multiple scenes in real time.

To ensure the validity and reliability of test results, the images were restored using the average data values of 10 measurements. System latency was measured as the time difference between physical actions captured by a high-speed camera at 240 FPS and corresponding updates in the virtual environment. Frame rates for each frame's rendering duration were measured in real time. Interaction response time was measured as the delay between user inputs (e.g., button clicks) and the response of the virtual scene. Model loading time and tracking time were measured as the time from the scene-transition command to the new-scene rendering. GPU memory utilization was measured as a peak value recorded by the NVIDIA Nsight developer in real-time monitoring. The metrics were measured automatically to avoid interference from human factors.

4.2 Results

The AI+VR system met the performance criteria required for immersive applications (Table 4). The system showed an average end-to-end latency of 20 ms with a maximum latency of not exceeding 25 ms, ensuring near real-time synchronization between user movements and corresponding responses in the virtual environment. This performance falls within the industry-recognized immersion threshold (under 20 ms to prevent motion sickness). The frame rate was consistent at 90 FPS, matching the device's refresh rate. In the simultaneous rendering of multiple images, minor fluctuations were observed, with the lowest observed frame rate of 85 FPS, which ensured a comfortable experience. The test results are presented in Fig. 6. The dashed line indicates the headset's maximum refresh rate of 90 FPS. The system maintained 90

Table 4
Performance test results of developed system.

Performance metric	Test result (average value)	Description
End-to-end system latency	20 ms (maximum 25 ms)	User movements to screen update, below perceptible threshold
Rendering frame rate (FPS)	90 FPS (minimum 85 FPS)	Smooth screen, meeting visual continuity requirements
Interaction response time	50 ms	Near-instant feedback for user inputs
Scene loading time	2.3 s	Smooth multiscene transitions with acceptable wait time
GPU memory usage	3.8 GB	Efficient resource allocation with no performance bottlenecks

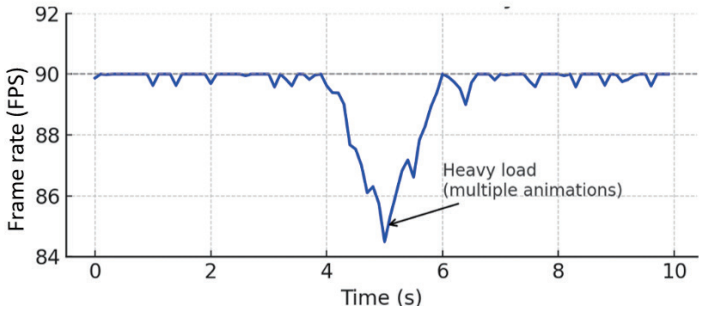

Fig. 6. (Color online) System frame rate over time in stability test.

FPS during operation, with a brief decrease to 85 FPS between 4 and 6 seconds after the initiation of the operation. However, after that, a stable frame rate was recovered. The arrow in the figure indicates the concurrent rendering of multiple images, which increased the computational load.

The average latency (interaction time) was 50 ms between user input and response, which enhanced immersion and operational naturalness. Scene loading times averaged 2.3 s, during which the system provided transitional animations. User feedback confirmed that the loading time was acceptable and did not cause a disruption of immersive continuity. In the most complex scene, GPU memory usage was 3.8 GB (38% of the RTX 3080's total memory capacity). This demonstrated substantial performance without a resource bottleneck.

The system showed superior performance in delay control, rendering fluency, interactive response, loading delay, and resource management, meeting the engineering criteria for immersive experience. The ultralow latency of 20 ms ensured precise synchronization between user movements and visual feedback, particularly appropriate for delicate scenes in ICH craftsmanship training. A high frame rate (90 FPS) ensured visual coherence and stability, effectively minimizing VR-induced motion sickness and fatigue for users. The interactive response time (50 ms) led to the natural feeling and operation of the system. The scene switching time (2.3 s), combined with the optimization of transition processing, provided excellent immersion consistency. The GPU memory utilization remained at an optimal level, supporting larger virtual scenes for the synchronous experiences of multiple users.

The AI+VR system demonstrated outstanding performance metrics and technically viable implementation potential. The system can be used for the preservation and dissemination of

ICH. The system needs to be tested by multiple users in a complex environment to optimize system stability and resource efficiency in diverse applications.

5. Application of AI+VR System in ICH Dissemination

The developed AI+VR digital dissemination system, integrating GANs, NeRF, and VIO, is a superior and effective solution for the preservation and interactive dissemination of ICH. The system can be applied to create high-fidelity, immersive, and interactive ICH experiences, specifically in virtual opera and craftsmanship transmission, highlighting the difference in resulting engagement and knowledge transfer compared with traditional methods.

5.1 Virtual opera: from passive viewing to immersive presence

Opera, encompassing genres such as Peking Opera, is an ICH involving complex performances, intricate costumes, and historical venues. Traditional dissemination methods, such as 2D video recordings, only offer a passive experience with a fixed perspective, failing to capture the full spatial and cultural richness. The system developed in this study can transform this experience. GANs are leveraged to restore damaged or faded historical images of opera set designs and costumes, automatically generating high-quality, authentic textures and colors to enrich the virtual environment. Simultaneously, NeRF reconstructs performance venues and stages into photorealistic 3D scenes from simple multiview photographs, effectively capturing realistic lighting and eliminating the need for complex, time-consuming laser scanning. This allows viewers to observe the performance ambiance and details from any angle in the virtual space, unlike the single, fixed viewpoint of video recordings. The experience can be made comfortable through VIO technology, which ensures ultralow latency tracking of the user's head and hand movements with a response time as low as 20 ms. This low latency is crucial for mitigating motion sickness, allowing users to move freely in the virtual opera house and examine the details of a performer's gestures or costume from a first-person perspective without the constraints of external trackers. This is different from 2D videos as the system offers a superior, personalized, and interactive experience that enhances cultural engagement. Such a technical fusion enables users to become active participants in the heritage, beyond documenting the performance.

5.2 Virtual craftsmanship transmission: embodied learning of tacit knowledge

ICH craftsmanship techniques, such as traditional pottery or embroidery, are characterized by knowledge and skills acquired through physical demonstration and repetitive practice. Traditional methods, reliant on text, images, or standard videos, often fail to transfer this embodied knowledge effectively.

The AI+VR system creates HDT that focuses on process and embodiment. NeRF generates high-realism 3D models of the workspace, tools, and artifacts with exceptional texture, allowing the user to inspect them in detail. In addition, VIO facilitates embodied, interactive learning.

The system accurately tracks the user's 6-DoF hand and head movements and maps them onto the virtual inheritor role.

The interaction response time of the system developed, set at a rapid 50 ms, is optimized for simulating the delicate, precise movements required in craftsmanship. This allows the user to practice the techniques, for instance, virtually manipulating clay on a digital wheel by mimicking the instructor's motions. By providing realistic, natural, and near-instant feedback to the user's physical input, the system enables the perception and comprehension of the inherent physical essence of the craft through virtual body movements, effectively transferring tacit knowledge. This approach transforms the learning process from passive observation to active participation, enabling a depth of skill transmission. This aligns with embodied communication, which emphasizes that physical action enhances learning and cultural understanding.[25]

6. Conclusion

We developed the AI+VR system for the dissemination of ICH in this study. Adopted technologies were validated, and case studies were conducted to propose appropriate collaboration models. VR integrated with AI empowers the preservation, dissemination, and innovation of ICH. GAN can be used for the automatic restoration and reproduction of ICH images and patterns to provide high-quality images for digital preservation. NeRF can solve the bottlenecks of traditional 3D reconstruction, efficiently creating digital twins of ICH scenes and presenting them in the virtual environment. VIO-based positioning and real-time rendering ensure the reliability of immersive experiences. The AI+VR system ensures interactivity and an immersive experience compared with traditional 2D display methods. Although its implementation requires substantial computing power, technological advancements are continuously lowering the barriers to the adoption of the system. The AI+VR system evolves ICH dissemination and cultural engagement from passive "seeing" and "hearing" to active "doing" and "experiencing". Viewers become "participants" in the virtual environment, where they can play traditional instruments, immerse themselves in opera roles, or collaborate with digital craftsmen. The system increases cultural consumption and enhances public interest and participation in ICH, paving new pathways for integrating ICH in diverse applications. Case studies illustrate how various VR-based ICH projects have been integrated into tourism, education, and cultural and creative industries, successfully bridging technology and arts. However, cultural essence is the foundation of the technology integration. To preserve the authenticity of ICH, excessive commercialization or distorted representations must be avoided as they might undermine its cultural integrity. Legal frameworks must be implemented to safeguard intellectual property and the rights of inheritors, ensuring that cultural commodification does not lead to alienation or conflict.

Acknowledgments

This research was funded by the 2024 Heilongjiang Province Philosophy and Social Sciences General Project "Research on Innovative Strategies for Digital Dissemination of Intangible Cultural Heritage in Heilongjiang Province" (Project No. 24YSB014).

References

1 L. Zhao, H. Zhang, and J. Mbachu: Remote Sens. **15** (2023) 1264. https://doi.org/10.3390/rs15051264
2 I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio: ArXiv. https://doi.org/10.48550/arXiv.1406.2661
3 C. Qin: J. Commun. Univ. China **6** (2019) 105. https://doi.org/10.3969/j.issn.1007-8770.2019.08.020
4 E. Selmanović, J. Bužinko, C. Harvey, and D. Dusnaka: J. Comput. Cult. Herit. **13** (2020) 1. http://dx.doi.org/10.1145/3377143
5 T. Pistola, S. C. Stentoumis, E. A. Stathopoulos, G. Loupas, and T. Mandilaras: Proc. 2021 IEEE Int. Conf. Intelligent Reality (ICIR, 2021) 17. https://doi.org/10.1109/ICIR51845.2021.00012
6 X. Chen: https://doi.org/10.13965/j.cnki.gzmzyj10026959.2020.11.015 (Accessed May 2025).
7 E. Schreurs: The Tacit Dimension: Architecture Knowledge and Scientific Research, Ed. L. Schrijver (Leuven University Press, Leuven, 2021) Chap. 5. https://doi.org/10.1353/book.83868
8 H. Xie: ARS (2018) 97. https://link.oversea.cnki.net/doi/10.16065/j.cnki.issn1002-1620.2018.04.016
9 M. Moussaid, V. R. Schinazi, M. Kapadia, and T. Thrash: Front. Robot. AI **5** (2018). https://doi.org/10.3389/frobt.2018.00082
10 S. Hall: Representation: Cultural Representations and Signifying Practices (Sage Publications, Thousand Oaks, 1997) Chap. 1. https://fotografiaeteoria.wordpress.com/wp-content/uploads/2015/05/the_work_of_representation__stuart_hall.pdf.
11 M. Jie and Z. Hou: Ethn. Art Stud. **34** (2021) 139 (in Chinese). https://link.oversea.cnki.net/doi/10.14003/j.cnki.mzysyj.2021.06.16
12 Y. Gao and Z. Xu: Decoration **5** (2022) 142. https://link.oversea.cnki.net/doi/10.16272/j.cnki.cn11-1392/j.2022.05.011
13 A. Fuller, A. Fan, C. Day, and C. Barlow: IEEE Access **8** (2020) 108952. https://doi.org/10.1109/ACCESS.2020.2998358
14 L. T. De Paolis, S. Chiarello, C. Gatto, S. Liaci, and V. De Luca: DAACH **26** (2022) e00238. https://doi.org/10.1016/j.daach.2022.e00238
15 Y. Chen, Y. Chen, Z. Pei, and C. Wang: Chin. Mech. Eng. **31** (2020) 797. https://doi.org/10.3969/j.issn.1004-132X.2020.07.005
16 R. Yang, Y. Li, Y. Wang, Q. Zhu, N. Wang, Y. Song, F. Tian, and H. Xu: Sustain. **16** (2024) 5281. https://doi.org/10.3390/su16135281
17 Y. Chen, Y. Chen, Z. Peim, and C. Wang: Intell. Data Work **11** (2020) 797. https://doi.org/10.3969/j.issn.1002-0314.2018.02.016
18 L. F. R. Correia, R. Bartholo, A. Brufato, and E. C. T. Sanchez: Advances in Tourism, Technology and Systems (2023) p. 419. http://dx.doi.org/10.1007/978-981-99-0337-5_35
19 S. Yan: Packag. Eng. **44** (2023) 1. https://link.oversea.cnki.net/doi/10.19554/j.cnki.1001-3563.2023.20.001
20 J. Lu and A. Wang: Mod. Ancient Cult. Creat. **47** (2021) 63. https://link.oversea.cnki.net/doi/10.20024/j.cnki.cn42-1911/i.2021.47.028
21 A. H. Y. Hon and E. Garnot: Int. J. Tour. Res. **24** (2022) 216. https://doi.org/10.1002/jtr.2495
22 B. Mildenhall, P. P. B., Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng: Proc. 2020 European Conf. Computer Vision (ECCV, 2020) 405. https://doi.org/10.1007/978-3-030-58452-8_24
23 C. Liu, S. Yu, and M. Yim: Proc. 2020 IEEE Int. Conf. Robotics and Automation (ICRA, 2020) 826. https://doi.org/10.1109/ICRA40945.2020.9196880
24 Z. Wan, B. Zhang, D. Chen, P. Zhang, D. Chen, and F. Wen: IEEE PAMI **45** (2023) 2071. https://doi.org/10.1109/TPAMI.2022.3163183
25 P. Dourish: Where the Action Is: The Foundations of Embodied Interaction, P. Dourish, Ed. (The MIT Press, Cambridge, 2001) Chap. 4. https://doi.org/10.7551/mitpress/7221.003.0005

About the Authors

Xuehong Zhao received her M.S. degree in industrial economics from Harbin University of Science and Technology in 2009. Since 2021, she has been an associate professor at Harbin Finance University, China. Her research interests span intangible cultural heritage (ICH) communication, brand management, and artificial intelligence. (2009029@hrbfu.edu.cn)

Mingyu Zhao received his M.S. degree in software engineering from Northeastern University in 2015. Since 2011, he has served as an assistant researcher at the Heilongjiang Academy of Science, focusing on deep learning algorithms and sensor technology. (zhaomingyu2794@dingtalk.com)

Hailing Wang received her M.S. degree in computer application technology from Harbin Engineering University in 2009, followed by a Ph.D. degree in computer application technology from Harbin Engineering University, P.R. China, in 2013. Since then, she has been a lecturer at Heilongjiang University of Science and Technology, specializing in knowledge graph, virtual reality, and software engineering. (wanghailing@usth.edu.cn)

Yanshan Zhou received his M.S. degree in technology economics and management from the School of Economics and Management at Harbin University of Science and Technology in 2005. He earned his Ph.D. degree in management science and engineering from the same institution in 2010. Since then, he has been an associate professor of the School of Economics and Management at Harbin University of Science and Technology. His research interests include operation and supply chain management, and management decision-making and optimization. (yszhou76@hrbust.edu.cn)