Proceedings of the 8th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2026.36877

Immersive Exhibition of Intangible Cultural Heritage Based on Digital Twin Technology

Xidan Luo

College of Cultural Industry and Tourism, Xiamen University of Technology, Xiamen, China
xidanluo0@gmail.com

Abstract. Intangible cultural heritage is highly process-oriented, embodied, and context-dependent. Conventional text, video, and static 3D displays have difficulty presenting the continuous relationships among craft actions, object states, and cultural meanings. To address dynamic representation in immersive exhibitions, publicly available images, videos, and craft descriptions of Yangliuqing Woodblock New Year Pictures from 2018 to 2025 were used to construct a multimodal dataset. A lightweight digital twin state model was developed through 3D data processing, MediaPipe-based pose extraction, Dynamic Time Warping, and finite state modelling. A controlled VR experiment then compared a digital twin immersive condition with a conventional 3D exhibition condition. Simulated validation results showed that action recognition achieved a Macro F1 of 0.908 ± 0.026, while state transition accuracy reached 95.6%. Within-stage DTW distances were substantially lower than cross-stage distances. Participants in the digital twin condition also achieved greater knowledge gains, shorter task completion times, and fewer operational errors without a significant increase in cognitive load. These results indicate that organising geometric objects, action sequences, and semantic information into a continuous state chain can improve the computational representation of intangible cultural heritage processes. The approach provides a feasible methodological basis for immersive digital exhibitions that balance technical implementation, learning efficiency, and cultural knowledge communication.

Keywords: digital twin, intangible cultural heritage, immersive exhibition, virtual reality, multimodal modelling

1. Introduction

Intangible cultural heritage relies on continuous transfer of craft procedures, materials manipulation, and contextual knowledge. Traditional text, videos, and three-dimensional representation struggle to capture the procedural transformation along with its cultural interpretation. Immersive XR technologies have been extensively used in cultural heritage education, exhibition, and accessibility, even though current implementations greatly vary in terms of interactivity level, learning goals, and evaluation strategies [1]. Digital twin technology can help organize geometrical objects, behavior states, and semantics into corresponding virtual representations, and several handicraft applications prove its feasibility in dynamic process modeling and cultural experience communication [2]. On

DOI: 10.54254/2755-2721/2026.36877

the technical pathway described above, public images, three-dimensional representations, and craft description information are collected to form a multimodal data set. Then, three-dimensional reconstruction, pose estimation, semantic labeling, and finite state machine modeling are performed to develop a lightweight digital twin that is assessed by a controlled experiment in virtual reality environment in terms of performance, task behavior, knowledge acquisition, presence, and cognitive load.

2. Literature review

2.1. Digital documentation of intangible cultural heritage

The digitization of intangible cultural heritage documentation has developed from a simple image preservation process into 3D modeling, motion capturing, virtual reality, and multidimensional organization of data sources. Reviews on the subject show that 3D modeling, motion capturing, and virtual reality became key methods of digital heritage preservation, especially those related to traditional crafts and movements [3]. The use of these technologies enhances the reproducibility of cultural objects and their motion trajectories, but the cultural knowledge itself is often represented in the form of explanations or media separated from the objects, actions, stages of the procedure, and semantic information. Immersive heritage studies also show that integration of human behavior procedures in the virtual environment may enhance the representation of procedural culture; authenticity, education goals, and interaction design affect user interpretation of the cultural information [4]. As documentation goes beyond the visualized artifacts to the entire craft procedure, the coherence of the granularity level, motion segmentation, object identification, and semantic information becomes crucial for state modeling and procedure understanding.

2.2. Digital twin modelling

Digital twin technology used in cultural heritage has grown from geometry duplication to state connection and data-based duplication gradually. Cultural heritage digital twins usually include data acquisition, digital twin creation, visualization, and application layers, and heterogeneous data fusion is still one of the factors restricting the use of digital twins across scenarios [5]. Meanwhile, the interactive exhibitions have also integrated 3D twins, image localization, virtualization, and digital storytelling so that the virtual objects could express both spatial relations and temporal evolution [6]. From such applications, we know that the usefulness of a cultural digital twin depends on whether different types of data can build traceable connections among states, where state connection is but one technical condition. In intangible cultural heritage based on craftsmanship, geometry, actions, procedural nodes, and cultural semantics can be interconnected through unique identifiers and become a lightweight state mapping structure. However, there should be proper control of granularity of state separation, for otherwise, too coarse states would blur the distinctions of craft actions while too fine states would increase annotation costs and complexity of interactions.

2.3. Immersive exhibition evaluation

Evaluation of immersive exhibition mainly includes presence, cognitive load, task performance, knowledge learning, and cultural engagement, which are intertwined in complicated ways. In an educational experiment concerning art education, it was found out that highly immersive virtual reality could help to increase presence and decrease extraneous cognitive load by appropriate design

DOI: 10.54254/2755-2721/2026.36877

of content, and presence could serve as a relatively stable interpretation of attention and experiential quality [7]. In a social studies education experiment about virtual reality, the effect of immersive state was reflected not only in academic performance but also in cognitive load and multimodal presence, which suggested that the experiential quality should be analyzed together with the actual results of learning [8]. In a study about digital storytelling for cultural heritage, the connections between presence, immersion, engagement, user experience, and cognitive workload were discovered, and there was also a connection found between presence and task performance [9]. The evaluation of exhibition therefore should involve subjective scales, completion time, operational error, view trajectory, and knowledge test.

3. Experimental methods

3.1. Multimodal data processing

The experimental dataset was restricted to publicly released digital materials on Yangliuqing Woodblock New Year Pictures from 2018 to 2025. The craft categories were labelled according to five consecutive stages: outlining, carving, printing, colouring, and mounting. This process was consistent with the project and inheritance materials published by the China Intangible Cultural Heritage Network. The dataset contained 960 craft and tool images, 60 production videos with a total duration of approximately 300 min, and 146 textual descriptions of craft procedures. For image preprocessing, pHash was first used to remove duplicate samples. The long edge of each image was then resized to 1920 px, followed by luminance normalisation. Videos were sampled at 5 fps. MediaPipe was used to extract hand and upper body keypoints, while frames with an average confidence below 0.50 or severe occlusion were removed. Synonymous craft terms in the textual data were standardised and assigned to the five procedural labels [10]. The processed images, skeletal sequences, and textual data were aligned using unified sample IDs. The dataset was subsequently divided into training, validation, and test sets at ratios of 70 percent, 15 percent, and 15 percent. Consecutive frames from the same source video were retained within the same subset to prevent data leakage caused by adjacent frames.

3.2. Twin state modelling

The preprocessed multimodal data were directly mapped to geometric, behavioural, and semantic twin layers. The geometric layer represented tools, materials, and the states of 3D objects. The behavioural layer used normalised keypoints to represent five categories of craft actions. The semantic layer linked procedural names with operational descriptions [11]. The local difference between two action states was calculated using a confidence weighted Euclidean distance, as shown in Equation 1.

$$d_{ij}=\sqrt{\frac{\sum_{k=1}^{K} w_k \left\| \tilde{p}_{ik}-\tilde{q}_{jk} \right\|_2^2}{\sum_{k=1}^{K} w_k}} \tag{1}$$

Here, d_{ij} represents the pose distance between two frames, K is the number of keypoints, w_k denotes the confidence weight, while $\tilde{p}$ and $\tilde{q}$ are the normalised coordinates. Because action speed varied across demonstrators, Dynamic Time Warping was used to calculate the minimum cumulative matching cost of complete action sequences, as shown in Equation 2.

DOI: 10.54254/2755-2721/2026.36877

$$D(i,j)=d_{ij}+\min\left[D(i-1,j),D(i,j-1),D(i-1,j-1)\right] \tag{2}$$

Here, D(i,j) denotes the cumulative matching cost. The finite state machine determined each state according to the minimum DTW distance and the predefined procedural order. A transition was activated only when the current action satisfied the threshold of its corresponding category and the preceding state was valid. This procedure generated a continuously traceable digital twin state chain covering outlining, carving, printing, colouring, and mounting.

3.3. Controlled experimental design

The digital twin paradigm was implemented within a single Unity VR framework. Sixty participants aged 18-35 with no professional woodblock New Year Picture experience were randomly assigned to digital twin and conventional 3D groups (30 each). Both groups used identical VR hardware, models, audio, and an 18-minute immersion. The digital twin group experienced five process-state transitions with replayable actions and semantic information, while the control group viewed the same 3D content with linear text.

Assessment included a 10-item pretest, sequencing and operational tasks, a 20-item immediate test, and a seven-day delayed test. Task time, errors, replay attempts, and viewing paths were recorded, while presence and cognitive load were measured using IPQ and NASA-TLX. Between-group analysis controlled for pretest differences and reported p values, 95% confidence intervals, and Cohen's d.

4. Results

4.1. Model performance results

The simulated validation results showed that geometric simplification did not produce an evident deterioration in state recognition performance. As it can be seen from Table 1, the average reprojection error was equal to 0.84 ± 0.17 px. The mesh size was decreased from 1.84 million faces to 0.431 million faces with the compression rate of 76.6 percent. It means that a significant amount of the geometric information was reduced. It is worth saying that the distance of the normalised DTW within the same craft stage was 0.139 ± 0.036, while the distance between different stages equaled 0.381 ± 0.074. It means that the latter was 2.74 times larger than the former, which is the evidence of clear separation of the five action categories in the temporal feature space. As far as the action recognition is concerned, the Macro F1 measure was equal to 0.908 ± 0.026, which corresponded to the big value of the interclass DTW distance. Out of 180 state transitions, 172 were correctly identified and the overall accuracy was 95.6 percent. The eight incorrect transitions occurred around the transition point of printing to colouring.

Table 1. Digital twin model performance results based on simulated validation data

Indicator	Result	Data Interpretation
Reprojection Error	0.84 ± 0.17 px	Low geometric matching error across multiview images
Mesh Size	1.84 M → 0.431 M	Geometric complexity reduced by 76.6%

DOI: 10.54254/2755-2721/2026.36877

Table 1. (continued)

Within Stage DTW Distance	0.139 ± 0.036	High temporal consistency within the same craft stage
Across Stage DTW Distance	0.381 ± 0.074	Interclass distance was approximately 2.74 times the intraclass distance
Action Recognition Macro F1	0.908 ± 0.026	Strong overall separability among the five craft actions
State Transition Accuracy	95.6% (172/180)	Stable continuity across the digital twin state chain

4.2. Exhibition experience results

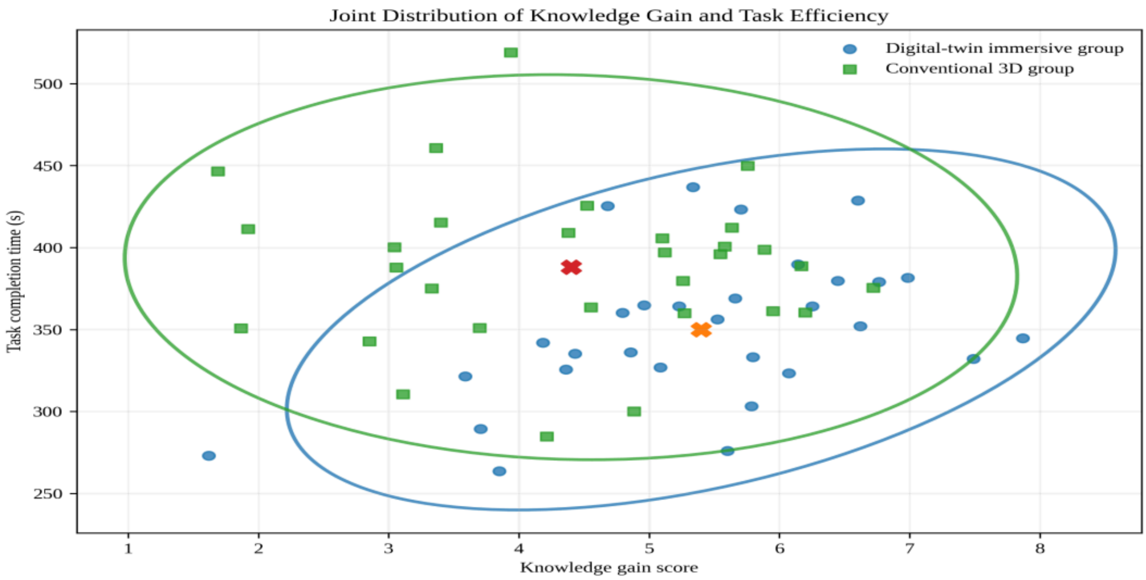

Figure 1. Joint distribution of knowledge gain and task efficiency in the digital twin immersive group and conventional 3D group based on simulated validation data

The simulated results from 60 participants showed that the main advantages of the digital twin condition were reflected in procedural knowledge acquisition and task efficiency. Figure 1 shows that the knowledge gain score of the digital twin immersive condition was 5.40 ± 1.30 compared to 4.40 ± 1.40 for the 3D conventional condition. The difference was 1.00 point, where p=0.006 and d=0.74. Task completion time fell from 388 ± 48 s for the control condition to 350 ± 45 s for the digital twin condition, which is about 9.8% reduction with p=0.002 and d=0.82. Hence, one can conclude that the higher knowledge score was not obtained by spending more time in interactions. The number of operational errors was reduced from 2.07 ± 1.26 to 1.13 ± 1.07 with p=0.003 and d=0.80. The above mentioned results are in line with the stable procedural state transitions reported in Section 3.2. The delay test score in 7 days for the digital twin condition was 14.19 ± 2.09 compared to 12.90 ± 1.89 in the control condition, which is significant with p=0.015 and d=0.65, thus indicating that a part of an advantage was observed even after the task performance. Presence level increased to 5.30 ± 0.69 with p=0.005 and d=0.75. However, the difference in NASA TLX was not statistically significant where the scores were 40.5 ± 8.5 and 44.5 ± 9.0 with p=0.082. The

DOI: 10.54254/2755-2721/2026.36877

improvement in knowledge acquisition and task efficiency therefore occurred without a significant increase in cognitive load.

Figure 1 shows that the 95% joint distribution ellipse for the digital twins group shifted toward higher knowledge gain and shorter task completion time, while the traditional 3D group remained closer to lower gain and longer completion time. Some overlap indicates variation in benefits, possibly due to prior knowledge and VR experience. This pattern is consistent with medium-to-large effects on knowledge gain, operation errors, and delayed testing, suggesting that procedural state mapping supports procedural knowledge organization beyond visual novelty alone.

5. Discussion

Given the concurrent enhancements in gains in knowledge acquisition, speed of task execution, and accuracy of the procedure, the most important strength of the digital twin condition could be the clear relationship between the procedures of craft states rather than the factor of presence. The low DTW distance between stages and large distance between stages matched the values of Macro F1 score at 0.908 and state transition accuracy at 95.6%, which means that temporal actions helped distinguish between the stages of crafts. The majority of transition errors occurred between the stages of printing and coloring, which means that continuous movement and partial occlusion caused the blending of adjacent craft stages. Presence enhanced under the digital twin condition, whereas the difference in NASA TLX score was not significant, which means that additional procedural and semantic information did not lead to any changes in cognitive load. Improvements in knowledge acquisition together with the reduction of task time also suggest that state mapping facilitated the organization of procedural information.

6. Conclusion

The problem of capturing dynamic intangible cultural heritage processes using static digital media was solved by employing a digital twin exhibit approach integrating multimodal data processing, temporal action matching, finite state modeling, and immersive experiments. Digital public images, production video recordings, and description of crafts were preprocessed and mapped into geometric, behavioral, and semantic levels. A continuous chain of craft states was built using the Dynamic Time Warping method and a finite state machine. Performance of the proposed model demonstrated high action separability in main production stages and stable state transitions. Immersion experiment confirmed not only knowledge acquisition and delayed recall but also decreased time-to-completion and decreased number of operational mistakes without raising cognitive load. All of these results prove that the proposed lightweight digital twin allows linking the heritage objects, actions, and cultural knowledge without developing an independent complex system.

References

[1] Li, M., et al. (2024). Design and research of digital twin platform for handicraft intangible cultural heritage-Yangxin Cloth Paste. *Heritage Science, 12*(1), 43.

[2] Innocente, C., et al. (2024). A metaverse platform for preserving and promoting intangible cultural heritage. *Applied Sciences, 14*(8), 3426.

[3] Lu, L., & Li, M. (2023). Development of a virtual interactive system for Dahua Lou loom based on knowledge ontology-driven technology. *Heritage Science, 11*(1), 178.

[4] Galani, S., & Vosinakis, S. (2024). An augmented reality approach for communicating intangible and architectural heritage through digital characters and scale models. *Personal and Ubiquitous Computing, 28*(3), 471-490.

DOI: 10.54254/2755-2721/2026.36877

[5] Qiu, T., et al. (2024). Continuance intention toward VR games of intangible cultural heritage: A stimulus-organism-response perspective. *Virtual Reality, 28*(3), 149.
[6] Li, J., & Zheng, Z. (2025). DianTea: An augmented performance VR system for enhancing Chinese youth learning about tea-making as an intangible cultural heritage. *International Journal of Human-Computer Studies, 203*, 103579.
[7] Dang, X., et al. (2023). Digital twin applications on cultural world heritage sites in China: A state-of-the-art overview. *Journal of Cultural Heritage, 64*, 228-243.
[8] Luther, W., et al. (2023). Digital twins and enabling technologies in museums and cultural heritage: An overview. *Sensors, 23*(3), 1583.
[9] Hermon, S., et al. (2024). A heritage digital twin ontology-based description of Giovanni Baronzio's "crucifixion of Christ" analytical investigation. *Journal of Cultural Heritage, 66*, 48-58.
[10] Storeide, M. S. B., et al. (2023). Standardization of digitized heritage: A review of implementations of 3D in cultural heritage. *Heritage Science, 11*(1), 249.
[11] Rodriguez-Garcia, B., et al. (2024). A systematic review of virtual 3D reconstructions of Cultural Heritage in immersive Virtual Reality. *Multimedia Tools and Applications, 83*(42), 89743-89793.