ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

# The Application of Digital Twin Technology in the Revitalization of Jilin Traditional Craft Cultural Heritage and Aesthetic Education Practice in Colleges and Universities

*Wenjian Yan*[1,*]

[1] Jilin Animation Institute, Changchun, Jilin 130012 China

**SUMMARY:** *Aiming at the problems of insufficient pattern restoration, lack of process semantics and weak interactive verification in the digital modeling of Jilin traditional craft cultural heritage, a digital twin system for aesthetic education practice in colleges and universities was constructed. The system collects object images, point clouds, production videos, process texts and interaction logs, and completes multi-source data coding, image point cloud registration, dense reconstruction, mesh optimization, texture mapping, action semantic coding and state synchronization. The Jilin-CraftDT dataset is constructed, which contains 50 craft objects, 2500 multi-view images, 50 sets of point clouds, 800 local images of patterns, 120 production videos, and 4800 interaction logs. Comparative experiments show that the average reconstruction error of the proposed system is reduced to 1.16 mm, and the texture similarity is 94.0%. The frame rate of the five technology scenarios was maintained at 58.4-66.7fps, and the video memory occupation was 3.1-4.6GB. The interactive task completion rate reaches 91.0%, the average response delay is 91 ms, and the state synchronization accuracy is 95.6%. The results show that the system can provide modelable, interactive and verifiable technical support for the revitalization of Jilin traditional craft cultural heritage and aesthetic education practice in colleges and universities.*

***KEYWORDS:*** *Digital Twin; Traditional Craft of Jilin; Revitalization of Cultural Heritage; College Art Education Practice*

# 1 Introduction

The application of digital twin technology in the field of cultural heritage digitization has shifted from static 3D display to interactive, updatable and verifiable system modeling. Cultural heritage objects are no longer just scanned, archived and displayed, but form continuous mapping relationships through images, point clouds, videos, sensor data and interactive records, which support the digital expression of heritage morphology, structural semantics and use process. Existing studies have explored cultural heritage digital twin, museum digitization and heritage scene interaction, and consider that multi-source perception, 3D reconstruction, semantic modeling and visualization engine are the key technical foundations for the construction of such systems [1, 2].

The cultural heritage of Jilin traditional handicrafts has distinctive characteristics of regional materials, patterns and handcrafted processes. Its digital processing cannot stop at the level of image acquisition, video recording or text description, but also needs to structurally model the shape, texture details, production actions, tool use and process nodes. The research

*yanwenjian82@163.com
https://doi.org/10.65102/is2026962
Published: 30 April, 2026

on digital twin platform of handiccraft intangible heritage, XR cultural heritage interaction and semantic point cloud modeling provides technical reference for 3D reconstruction, process expression and virtual-real interaction of traditional craft objects [3-5]. From the perspective of system implementation, the traditional process digital twin model should contain geometric model, texture model, semantic label and state data at the same time, so that process resources can be directly invoked by subsequent interactive modules.

The practice of aesthetic education in colleges and universities provides a more clear verification scene for the digital twin system of Jilin traditional technology. The system can bind craft objects, pattern areas, production steps and interactive tasks into observable, manipulable and recordable digital objects, and collect student behavior data in model observation, local magnification, process selection and virtual operation. Relevant studies show that digital twin of cultural heritage not only needs to present visual form, but also should evaluate the operation effect of the system through interactive logic, real-time feedback and learning behavior data [6-8]. Research on VR immersive experience and virtual enhancement of tangible and intangible cultural heritage also shows that the digital activation of heritage resources needs to connect form display, process reproduction and user operation [9, 10].

Focusing on the above problems, this paper takes Jilin traditional craft cultural heritage as the object to construct a digital twin system for aesthetic education practice in colleges and universities. The research focuses on multi-source data acquisition, 3D reconstruction, process semantic coding, state synchronization and interactive verification. The second chapter establishes the system architecture and data modeling method, and specifies the data organization and operation link of the twin. The third chapter designs the interactive driving method of traditional process twin model construction and aesthetic education, completes image point cloud registration, texture mapping, action coding and interactive task association. Chapter 4 verifies the system performance in the scene of cultural heritage activation and aesthetic education in colleges and universities, and analyzes the reconstruction accuracy, response delay, rendering efficiency and interactive data results.

# 2 Digital Twin system architecture and data modeling of Jilin traditional craft cultural heritage

## 2.1 Multi-source data acquisition and preprocessing in traditional technology

The digital twin modeling of Jilin traditional craft cultural heritage relies on the stable collection and unified organization of multi-source data. The traditional craft object includes the outline of the object, the details of the pattern, the material texture, the production action, the process flow and the interactive behavior. When taking Jilin paper-cut, Manchu embroidery, pine and flower stone carving, straw weaving, wood carving and other samples as objects, a single picture archive or video record is difficult to support three-dimensional reconstruction, process reasoning and aesthetic education interaction in colleges and universities. The system needs to connect images, point clouds, videos, text annotations and interaction logs to the same link, so that the process resources can be converted into computable and callable digital twin input [11].

Multi-source collected data can be defined as heterogeneous datasets with timestamps, spatial calibration, and semantic labels:

$$D=\{(x_i^m,t_i^m,c_i^m,l_i^m) \mid m\in\{I,P,V,T,L\},i=1,2,\ldots,n_m\} \quad (1)$$

where, m represents the data modality, and I, P, V, T and L correspond to image, point cloud, video, text and log data respectively. $x_i^m$ represents the ith sample of the m-th class. $t_i^m$represents the acquisition timestamp. $c_i^m$denotes space calibration or equipment parameters; $l_i^m$ represents semantic labels generated manually or automatically; nm represents the number of samples. The structure can index data from different sources uniformly, and support image pattern localization, point cloud registration, action segmentation and interactive tracking.

During the acquisition process, the object images cover the front view, side view, top view, back view, local patterns and edge details, and the point cloud acquisition retains the surface undulation, concave and convex boundaries, bottom structure and scale information. The production video is sliced according to the process stage, and the shooting distance, light intensity and operating perspective are consistent. The process text adopts the field structure of "process number -- action name -- tool material -- input state -- output state". The interaction log records the user number, task number, operation type, trigger time, response delay, completion status and error number by the system background.

The multi-source data have differences in scale, sampling frequency, noise distribution and field dimension. The preprocessing stage needs to complete denoising, calibration, alignment and standardization. For the MTH type of data features, standardized fusion with weight constraints can be used:

$$\tilde{x}_i^m=\omega_m\cdot\frac{x_i^m-\mu_m}{\sqrt{\sigma_m^2+\varepsilon}}+\eta_m\cdot\phi(x_i^m,c_i^m,l_i^m) \quad (2)$$

where, $\tilde{x}_i^m$represents the preprocessed features; μm and $\sigma_m^2$ denote the mean and variance of the m-th class, respectively. ε is a stable term; ωm denotes the modal weights; ηm represents the calibration and label constraint coefficients; Let $\phi(\cdot)$ denote the correction function formed by the device parameters and semantic labels. This processing can reduce the bias caused by device differences, data scales, and noise levels, so that images, point clouds, video frames, process text, and log fields can enter the unified modeling link.

To ensure the consistency of system calls, the collected data are merged according to the source, processing method and modeling purpose. Table 1 lists the multi-source data configuration.

*Table 1: Traditional Craft Multi-source Data Collection and Preprocessing Configuration*

| Data Type | Collection Method | Preprocessing Method | Model Purpose |
|---|---|---|---|
| Overall Artifact Image | Multi-angle High-definition Shooting | Cropping, Denoising, Color Correction, Scale Calibration | Extract Artifact Contour and Appearance Features |
| Local Pattern Image | Macro Shooting and Local Magnification Collection | Edge Enhancement, Pattern Segmentation, Feature Point Extraction | Support Texture Mapping and Pattern Restoration |
| 3D Point Cloud Data | Structured Light Scanning or Laser Scanning | Outlier Removal, Coordinate Registration, Point Cloud Downsampling | Construct Artifact 3D Geometric Model |
| Production Process Video | Fixed Camera Position Phased Recording | Frame Extraction, Action Slicing, Key Frame Annotation | Generate Craft Action Sequence |
| Craft Process Text | Manual Sorting and Expert Annotation | Process Splitting, Label Encoding, Node Numbering | Construct Process Relationship and Semantic Nodes |
| Interaction Log Data | Automatic Recording by System Backend | Outlier Cleaning, Timestamp Alignment, Field Encoding | Verify Interaction Response and Task Completion Status |

The preprocessed image features were transformed into pattern region indexes, point cloud data were transformed into geometric reconstruction inputs, video key frames were transformed into action clips, process texts were transformed into process nodes, and interaction logs were transformed into user behavior sequences. See Figure 1 for the above processing links.

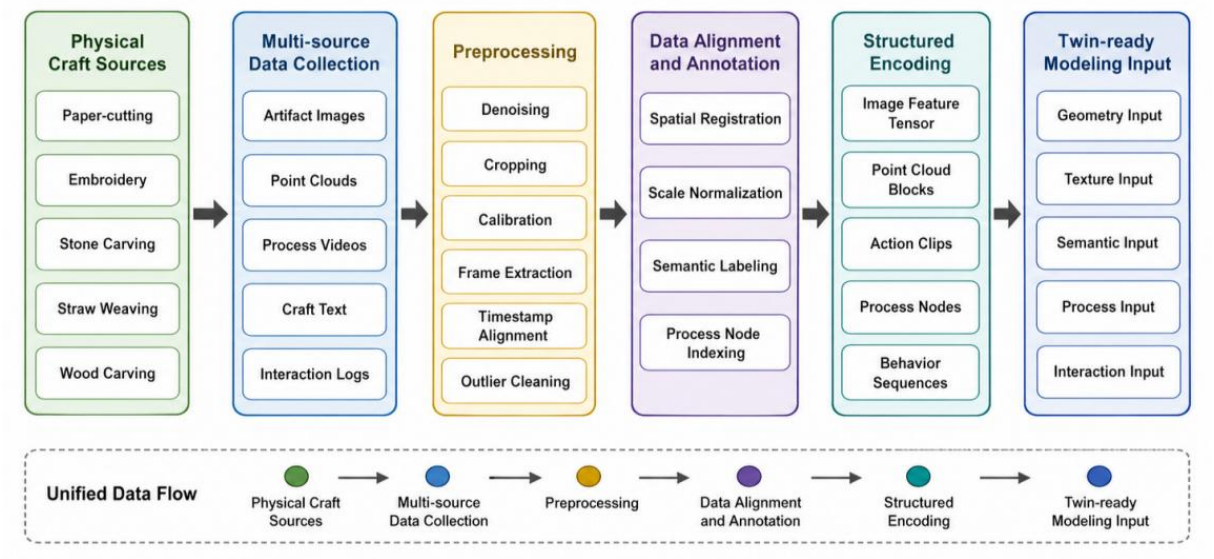


*Figure 1: Flowchart of Multi-source Data Acquisition and Structured Coding in Traditional Processes*

Through this link, the physical form, pattern information, production actions, and interaction records of the traditional craft cultural heritage of Jilin Province are transformed into structured data units, providing a data foundation for subsequent twin body division, three-dimensional model construction, process action coding, and university aesthetic education interaction verification.

## 2.2 Data Structure and Module Division of Craft Object Twins

The twin body of the craft object needs to describe the physical form, pattern material, process flow, action state, and interaction feedback at the data level [12]. After multiple-source collection of Jilin traditional craft samples, they cannot be directly input into the system as loose materials, but should be encapsulated as hierarchical twin objects. This paper divides the twin body into geometric layer, texture layer, semantic layer, process layer, state layer, and interaction layer. The geometric layer is responsible for storing the grid of the object, point cloud coordinates, and spatial scale; the texture layer is responsible for recording the pattern image, material parameters, and texture index; the semantic layer is used to label the parts of the object, the craft tools, and the pattern type; the process layer stores the production steps, dependencies before and after, and task nodes; the state layer records the current posture, process stage, and rendering state of the model; the interaction layer saves user operations, response delay, and feedback results [13]. The data of each layer together constitutes a queryable, updatable, and drivable digital twin object.

The twin body of the craft object can be expressed as:

$$DT_o=\{G_o,T_o,S_o,P_o,A_o,U_o\} \tag{3}$$

where, DTo represents the digital twin of the o-th traditional process object; Go represents geometric structure data, including point clouds, meshes, and scale parameters. To represents texture and material data, including pattern maps, color distributions, and surface properties. So represents semantic annotation data, including instrument location, tool category and process label. Po represents process relational data, including process nodes and their dependencies. Ao represents state data, including model pose, process progress, and rendering state. Uo represents interaction data, including user actions, task recordings, and feedback metrics. The structure can extend the traditional process entity from a single 3D model to a running system object.

The mapping relationship between entity objects and twin modules can be further expressed as follows.

$$M_o = \arg\min_{\theta} \left( \alpha \left\| G_o - F_g(X_o;\theta_g) \right\|_2^2 + \beta \left\| T_o - F_t(X_o;\theta_t) \right\|_1 + \gamma L_s(S_o,\hat{S}_o) + \delta L_p(P_o,\hat{P}_o) \right) \quad (4)$$

where, Mo represents the optimal mapping result from entity object to twin module; Xo represents the acquired multi-source input; $Fg(\cdot)$ and $Ft(\cdot)$ represent the geometric reconstruction function and texture mapping function, respectively. θg and θt denote the corresponding model parameters. Ls stands for semantic annotation loss; Lp represents the loss of process relationship; α, β, γ, δ are the weight coefficients. The mapping model integrates geometric error, texture error, semantic deviation and process deviation into the optimization goal, avoiding that the twin only completes the appearance reproduction without the technical process expression. The input-output relationship of each module of the twin is shown in Table 2.

*Table 2: Craft Object Twin Module Division*

| Module Name | Input Data | Output Result |
|---|---|---|
| Geometric Reconstruction Module | Artifact Image, 3D Point Cloud, Scale Calibration Parameters | Point Cloud Model, Mesh Model, Spatial Coordinates |
| Texture Mapping Module | Local Pattern Image, Material Photo, Mesh Surface Parameters | Texture Map, Material Parameters, Pattern Region Index |
| Semantic Annotation Module | Craft Text, Part Annotation, Tool Tag | Artifact Semantic Node, Craft Tag, Tool Category |
| Process Encoding Module | Production Video, Process Text, Action Slicing | Process Node, Action Sequence, Process Relationship |
| State Management Module | Model Pose, Craft Stage, Rendering Parameters | Current State, Process Progress, Update Record |
| Interaction Synchronization Module | User Operation Log, Task Data, Feedback Record | Operation Trajectory, Response Latency, Task Completion Status |

Through the above module division, Jilin traditional craft objects are no longer processed as independent materials, but are organized as composite twins with geometric structure, texture details, process semantics and interaction states. This structure provides a unified data interface for subsequent image point cloud registration, pattern reconstruction, action coding and interactive driving of aesthetic education in colleges and universities.

## 2.3 System Operation link and deployment architecture for aesthetic education scene in Colleges and universities

The digital twin system for aesthetic education in colleges and universities needs to support process object loading, pattern area recognition, production process switching, virtual operation feedback and behavior data recording. The system should not be designed as a simple 3D display page, but should form a running link composed of data end, server end, rendering end, interactive end and log end [14]. After the twin encapsulation of Jilin traditional process objects, the back-end database stored geometric models, texture maps, semantic labels, process scripts and interaction rules. The server is responsible for model indexing, resource scheduling, status update and interface return. The front end is responsible for 3D scene display, local pattern enlargement, process node triggering and interactive feedback. The log side records user actions, response latency, task status, and error counts [15]. The deployment method can transform traditional process resources into callable digital objects, and also provide an operational basis for subsequent performance verification and data analysis of aesthetic education practice in colleges and universities.

The system operating link can be expressed as follows.

$$Y_t=R\left(S\left(M\left(DT_o,Q_t;\theta_m\right),E_t;\theta_s\right),C_t;\theta_r\right) \tag{5}$$

where, Yt represents the front-end output at time t, including rendering screen, process status and interactive feedback; DTo denotes the invoked process object twin; Qt represents the user's current task request. $M(\cdot)$ represents the model scheduling function for object retrieval, model loading, and texture calling; Et represents an interaction event. $S(\cdot)$ represents the state synchronization function, which is used to update the process node, model pose and task state. Ct represents the client rendering environment. $R(\cdot)$ denotes the render output function; θm, θs, θr denote the scheduling, synchronization, and rendering control parameters, respectively. This link integrates twin call, event processing and rendering output into the same calculation process, so that model observation, pattern viewing, process selection and virtual operation form a continuous response.

In the deployment implementation, the process object model is stored separately according to lightweight grid, texture map, semantic node and process script. The geometric model is loaded by LOD grading, high-precision grid is called for close observation, and low-facet model is called for normal browsing state [16]. When students enlarge the pattern area, the system only loads the corresponding local map to reduce the repeated transmission of the whole high-resolution texture. The process script is bound to the process node through the interface. After the front-end triggers the tasks of "process view", "pattern recognition" and "virtual operation", the server returns the node status and interaction rules. After the user rotates the object, enlarges the pattern, selects the process node or completes the virtual operation, the interaction event enters the state queue, and the middle layer updates the state of the twin and synchronizes it to the rendering end and the log end [17].

System task scheduling needs to take into account model loading time, rendering output time, state synchronization delay and interactive task failure rate. The task queue scheduling objective can be expressed as follows.

$$J=\min_{\pi}\sum_{t=1}^{T}\left(\lambda_1 L_t^{load}+\lambda_2 L_t^{render}+\lambda_3 L_t^{sync}+\lambda_4 R_t^{fail}\right) \tag{6}$$

where, J represents the total scheduling cost of the system; Let π denote the task scheduling

policy; T represents the running time window; $L_t^{load}$ represents model and texture loading time. $L_t^{render}$indicates the rendering output time; $L_t^{sync}$is the front-end state synchronization delay. $R_t^{fail}$represents the interactive task failure rate; λ1, λ2, λ3, and λ4 are the weight coefficients. The objective function is used to constrain the balance between resource calls and interactive responses, so that the crafts, pattern areas and process nodes required by the current task are loaded first, and the occupation of irrelevant resources and front-end waiting time are reduced.

The system deployment structure is based on the twin database, and the model scheduling service completes the on-demand loading of geometric models, texture resources and process scripts. Then the interactive events, task states and rendering results are connected to the front-end scene by the state synchronization queue, forming a closed loop that can be recorded, replayed and evaluated.

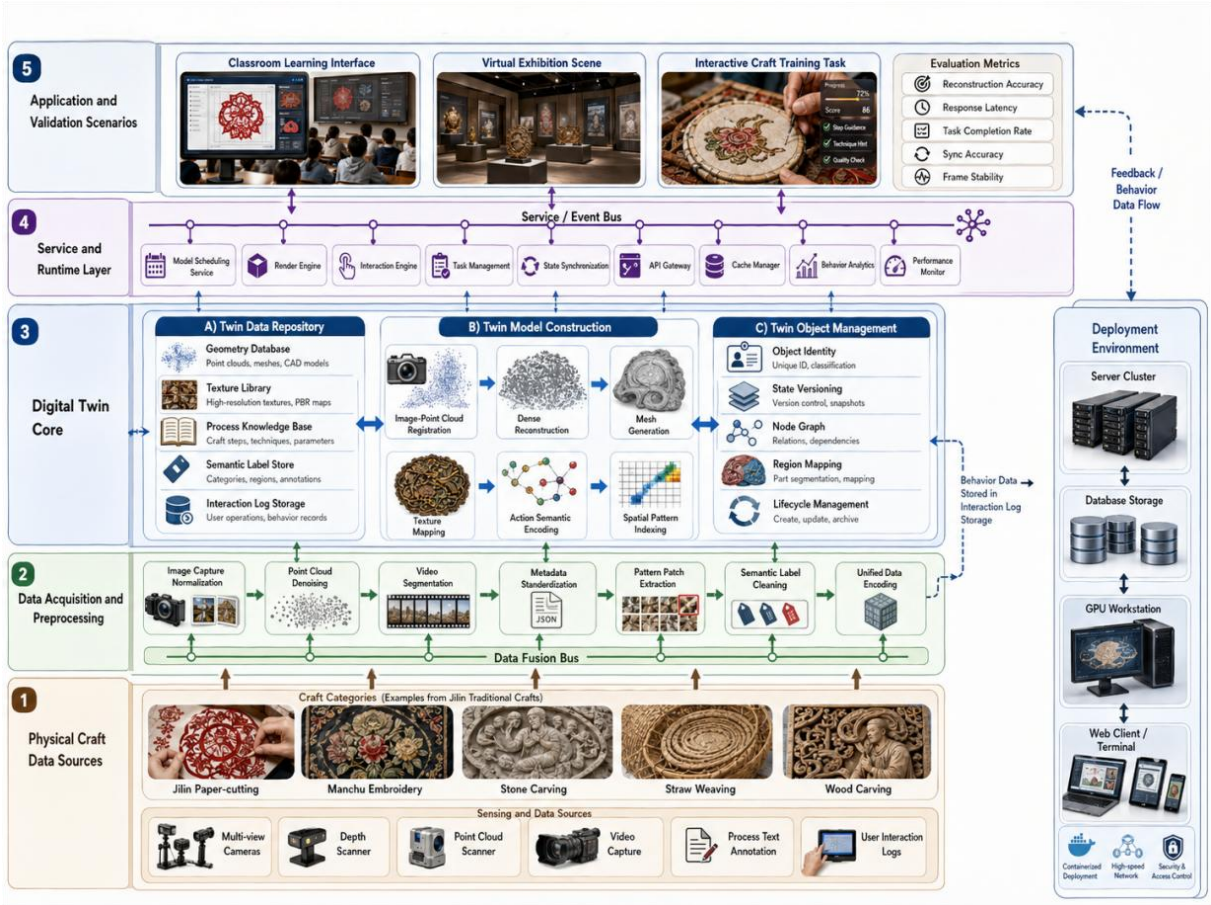


*Figure 2: Overall architecture diagram of digital twin system of Jilin traditional craft cultural heritage*

Under this architecture, Jilin traditional craft cultural heritage resources are organized as schedulable twin objects. The object observation, pattern recognition, process operation and task feedback in the aesthetic education scene of colleges and universities are executed through a unified link, and the behavior log and performance data are generated synchronously in the background of the system, which provides a running environment for subsequent image point cloud registration, pattern reconstruction, action semantic coding and system verification.

# 3 Traditional Craft Twin Model Construction and Aesthetic Education Interactive Driving Method

## 3.1 Image Registration Method for Craft Objects and Point Clouds

The registration of the images of craft objects and point clouds is the basic step in the construction of traditional craft twin models. Craft objects such as Jilin paper-cutting, Manchu embroidery, pine stone carving, straw weaving, and wood carving have significant differences in their morphological structure and pattern distribution. Solely relying on multi-view images can obtain relatively complete texture information, but the spatial scale and local concave-convex structure are prone to errors; solely relying on point cloud scanning can retain geometric contours and depth information, but it is insufficient in expressing the pattern boundaries, color gradations, and material details [18]. The system needs to uniformly register the image feature points, camera poses, scanned point clouds, and object coordinate systems to establish the same spatial reference for subsequent dense reconstruction, mesh generation, and texture mapping.

After inputting multi-view images, the system extracts features from the edges of the objects, pattern turning points, local holes, and high-contrast areas. To improve the matching stability of the craft pattern area, feature point matching not only considers the descriptor distance but also introduces spatial neighborhood consistency constraints. Let $u_i^a$ be the feature points in the a-th image and $u_j^b$ be the feature points in the b-th image. The matching score can be expressed as:

$$S_{ij}^{a,b}=\exp\left(-\frac{\left\|d_i^a-d_j^b\right\|_2^2}{\tau_d}\right)\cdot\exp\left(-\frac{\left\|\Delta u_i^a-\Delta u_j^b\right\|_2^2}{\tau_s}\right)\cdot\rho\left(r_i^a,r_j^b\right) \tag{7}$$

where, $S_{ij}^{a,b}$ represents the matching scores of feature points in the two images; $d_i^a$ and $d_j^b$ represent feature descriptors; $\Delta u_i^a$ and $\Delta u_j^b$ represent the spatial distribution of feature points in the neighborhood. $\tau d$ and $\tau s$ are the control parameters of descriptor distance and spatial distance, respectively. Let$\rho\left(r_i^a,r_j^b\right)$ denote the pattern region consistency weight. The formula can reduce the mismatching caused by repeated patterns, similar edges and partial occlusion, and make the stable feature points in the craft object image enter the subsequent space for solving.

After the feature matching is completed, the projection relationship between the 2D image coordinates and the 3D space points needs to be established. Suppose the 3D points are Xw=[X,Y,Z,1]T, the image pixels are u=[u,v,1]T, the camera internal parameter matrix is K, and the external parameters are composed of the rotation matrix R and the translation vector t, then the projection model can be written as follows.

$$s\begin{bmatrix}u\\v\\1\end{bmatrix}=K[R\mid t]\begin{bmatrix}X\\Y\\Z\\1\end{bmatrix},\quad K=\begin{bmatrix}f_x&0&c_x\\0&f_y&c_y\\0&0&1\end{bmatrix} \tag{8}$$

where, s represents the scale factor; fx and fy denote the focal lengths of the camera in both directions; cx and cy denote the coordinates of the principal points. R and t represent the pose parameters between the camera coordinate system and the world coordinate system. The

projection relation is used to back-project the pattern boundaries, object contours and feature points in multi-view images into three-dimensional space, which provides the coordinate basis for the registration between image reconstruction point clouds and scanning point clouds.

The point cloud registration stage requires aligning the image reconstruction point cloud $P=\{p_i\}_{i=1}^{N}$ with the scanned point cloud $Q=\{q_i\}_{i=1}^{N}$. Due to the surface hollowedness, relief, fabric wrinkles and local wear on the traditional crafts, the rigid closest point distance is easy to be affected by local noise. In this paper, normal consistency and boundary constraints are added to the registration objective, and the error optimization model is as follows.

$$(R^*,t^*)=\arg\min_{R,t}\ \sum_{i=1}^{N} w_i \left\|p_i-(Rq_i+t)\right\|_2^2+\lambda_n \sum_{i=1}^{N}\left(1-n_{p_i}^{T} R n_{q_i}\right)+\lambda_b \sum_{k=1}^{K}\left\|b_k^p-(Rb_k^q+t)\right\|_2^2 \quad (9)$$

where R∗ and t∗ represent the optimal rotation matrix and translation vector. wi denotes the weights of the point pairs. npi and nqi denote the normal vector of the corresponding point; $b_k^p$ and $b_k^q$ represent the boundary control points in the two types of point clouds. λn is the normal consistency weight. Let λb denote the boundary constraint weights. The first term constrict the overall spatial position, the second term reduces the surface orientation bias, and the third term ensures that the object edges, hole contours, and pattern boundaries remain consistent after registration. For the objects with obvious surface fluctuation, such as the pine and flower stone carving, the normal constraint can reduce the dislocation of local concave and convex regions. For objects with dense patterns, such as paper-cutting and embroidery, boundary constraints can improve the accuracy of fitting the contour of patterns to the 3D surface.

In the implementation, the system first generates sparse reconstruction points from multi-view images, and then uses the scanned point cloud to provide global scale constraints. In the initial registration stage, the coarse pose is estimated through feature point pairs and boundary point pairs. In the iterative stage, the weighted nearest neighbor search is used to update the point pair relationship, and the points with abnormal distance and excessive normal deviation are eliminated [19]. The average spatial error, boundary offset and normal consistency were calculated after each iteration, and the update was terminated when the error decrease was less than the set threshold. After registration, the coordinates of image pattern, scanned point cloud and local semantic label of the object are unified into the same 3D coordinate system, which provides a coordinate reference for subsequent texture mapping, mesh surface binding and process action space positioning.

In order to observe the spatial relationship between the image reconstruction point cloud and the scanning point cloud before and after registration, this paper selected typical process samples to visualize the point cloud distribution results of the three stages of coarse registration, iterative optimization and final alignment, and the relevant results are shown in Figure 3.

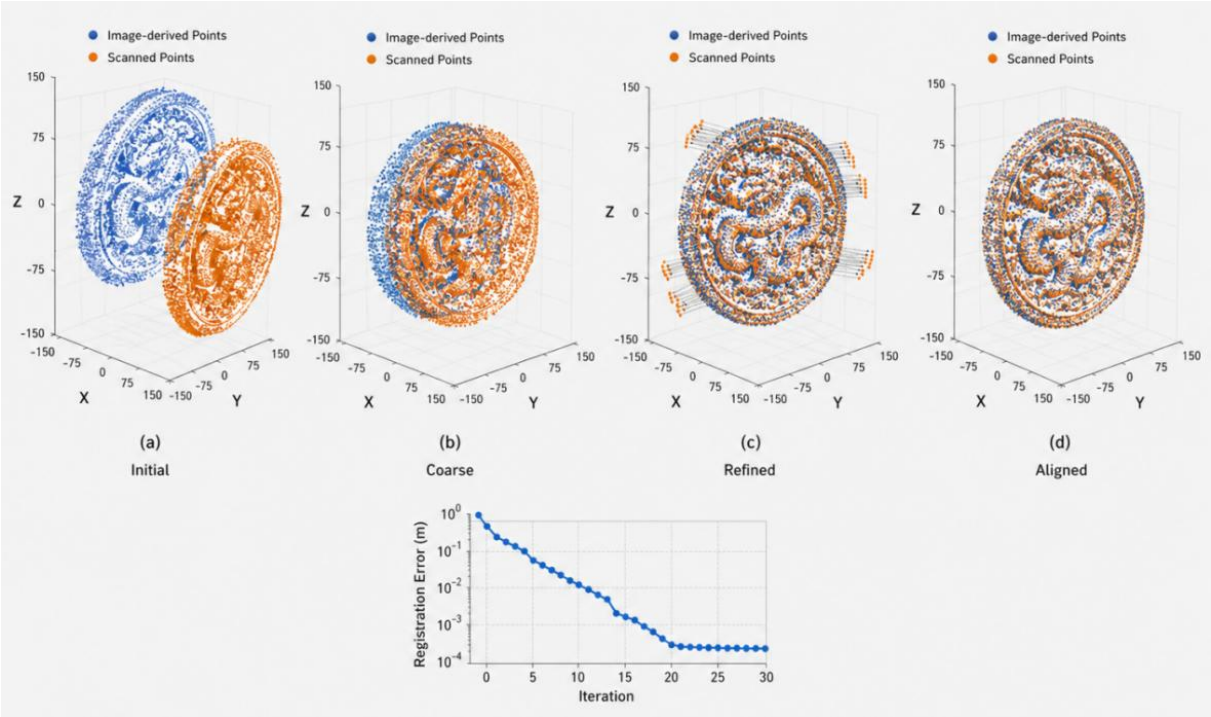


*Figure 3: Process object image-point cloud registration 3D spatial distribution map*

Before registration, the reconstructed points and the scanned points are obviously offset in space position, and the overlap degree of local boundary and surface area is low. After coarse pose estimation, the main contours of the two types of point clouds began to approach. After entering the iterative optimization stage, the corresponding point relationship is gradually stable, the number of outliers is significantly reduced, and the registered point cloud shows a high degree of overlap in the edge of the object, the undulating area of the pattern and the local turning area. The results show that the constructed registration method can realize the unified alignment of image point clouds and scanning point clouds in 3D space, and provide a stable geometric foundation for subsequent texture mapping and process action binding.

## 3.2 Dense reconstruction of technological patterns, mesh generation and texture mapping

After the image registration of the craft object and the point cloud is completed, it is still necessary to expand the sparse spatial points into a continuous surface and accurately bind the pattern details to the grid structure [20]. In the sample of Jilin traditional crafts, there are small hollows at the edge of paper-cut, the surface of Manchu embroidery has undulating lines, the pine and flower stone carving contains bas-relief texture, and the straw weaving and wood carving have more curved corners. If the model is directly generated from the sparse point cloud, the boundary of the object will be broken, and the pattern area is difficult to correspond to the 3D surface stably. In order to improve the observability and interactivity of the twin model, this section uses the processing link of "depth estimation -- dense point cloud fusion -- mesh optimization -- UV mapping" to map the process pattern from the 2D image space to the 3D mesh surface.

In the dense reconstruction stage, the depth value of the pixels in the multi-view image needs to be recovered according to the disparity relationship. Let the pixel in the RTH reference view be u, the candidate depth be z, and the set of adjacent views be V. The depth estimation target can be expressed as follows.

$$z^*(u)=\arg\min_z \sum_{v\in V}\psi\left(I_r(u)-I_v\left(\Pi_v\left(\Pi_r^{-1}(u,z)\right)\right)\right)+\lambda_d\|\nabla z(u)\|_1 \quad (10)$$

where z*(u) represents the optimal depth of pixel uu; Ir denotes the reference image, and Iv denotes the adjacent view image. Π r - 1 (.). Denotes the backprojection from 2D pixels to 3D space, and $\Pi_r^{-1}$(.)denotes the projection of 3D points to adjacent views. Let ψ( · ) denote the photometric consistency loss; Let λd denote the depth smoothness constraint coefficient. The proposed model can suppress the local noise while preserving the boundary variation of the pattern, so that the embroidery lines, stone lines and wood carving edges can obtain more stable depth estimation results.

After the dense point cloud is generated, it is necessary to build a continuous surface through a triangular mesh. Let the mesh vertex set be $V=\{v_i\}_{i=1}^N$, the edge set be E, and the point cloud observation point be pi. The mesh optimization objective can be written as follows.

$$V^*=\arg\min_V \left[\sum_{i=1}^N \omega_i \left\|v_i-p_i\right\|_2^2+\lambda_s\sum_{(i,j)\in E}\left\|v_i-v_j\right\|_2^2+\lambda_c\sum_{i=1}^N\|n_i-\bar{n}_i\|_2^2\right] \quad (11)$$

In the formula, V* represents the optimized grid vertices; ωi represents the credibility weight of the observation point; λs represents the coefficient of grid smoothing constraint; λc represents the coefficient of surface normal consistency constraint; ni represents the current vertex normal, and $\bar{n}_i$ represents the average normal of the neighborhood. The first term ensures that the mesh fits the dense point cloud, the second term controls the local surface jittering, and the third term constrains the continuity of the surface direction. For paper-cut samples with hollow edges, the weight of boundary vertices needs to be appropriately increased; for surface samples such as stone carvings and wood carvings, the normal consistency constraint can reduce local mesh folding.

During the texture mapping stage, the two-dimensional pattern image is bound to the surface of the three-dimensional mesh. Let the three-dimensional surface points be X, the vertices of the triangular face patch where it is located be v1, v2, v3, the corresponding two-dimensional texture coordinates be τ1, τ2, τ3, the centroid coordinates be α1, α2, α3, and the texture coordinate mapping relationship can be expressed as:

$$\tau(X)=\sum_{k=1}^3\alpha_k\tau_k,\quad X=\sum_{k=1}^3\alpha_k v_k,\quad \sum_{k=1}^3\alpha_k=1,\ \alpha_k\geq 0 \quad (12)$$

where τ(X) represents the 2D texture coordinates corresponding to the 3D surface point X; Let τk denote the UV coordinates of the vertices of the triangular facets; Let αk denote the barycenter weight. The mapping method can ensure the continuous expansion of the pattern along with the mesh surface, and avoid the obvious stretch during local enlargement. For Manchu embroidery and straw weave textures, the system adds local clarity detection after texture expansion, and invokes high-resolution pattern mapping for fuzzy areas. For the surface of the turquoise stone carving, the roughness and normal map should be preserved in the material parameters, so that the bas-relief details maintain a three-dimensional sense in real-time rendering.

3D reconstruction and texture mapping parameters directly affect model accuracy and front-end operation efficiency, Table 3 shows the main configurations adopted in this study.

*Table 3: 3D Reconstruction and Texture Mapping Parameter Configuration*

| Parameter Name | Parameter Setting | Function Description |
|---|---|---|
| Image Resolution | 4096 × 3072 | Preserve Artifact Contour and Pattern Details |
| Feature Threshold | 0.72 | Screen Stable Image Feature Points |
| Point Cloud Density | 1.5 mm | Control Dense Point Cloud Spatial Interval |
| Mesh Simplification Ratio | 35% | Reduce Front-end Rendering Load |
| Texture Size | 4096 × 4096 | Support Local Magnified Display of Patterns |
| Normal Map Level | 8 bit | Preserve Shallow Relief and Line Trace Undulation |
| LOD Levels | 3 | Realize Hierarchical Loading at Far and Near Distances |

The binding effect of pattern texture on 3D surface depends not only on the modeling process, but also on UV unwrapping, patch density and texture residual. The quality of the model after texture mapping needs to be checked from three levels: mesh continuity, UV unrolling stability, and local texture residual; Figure 4 shows the quality analysis results after dense reconstruction and texture mapping.

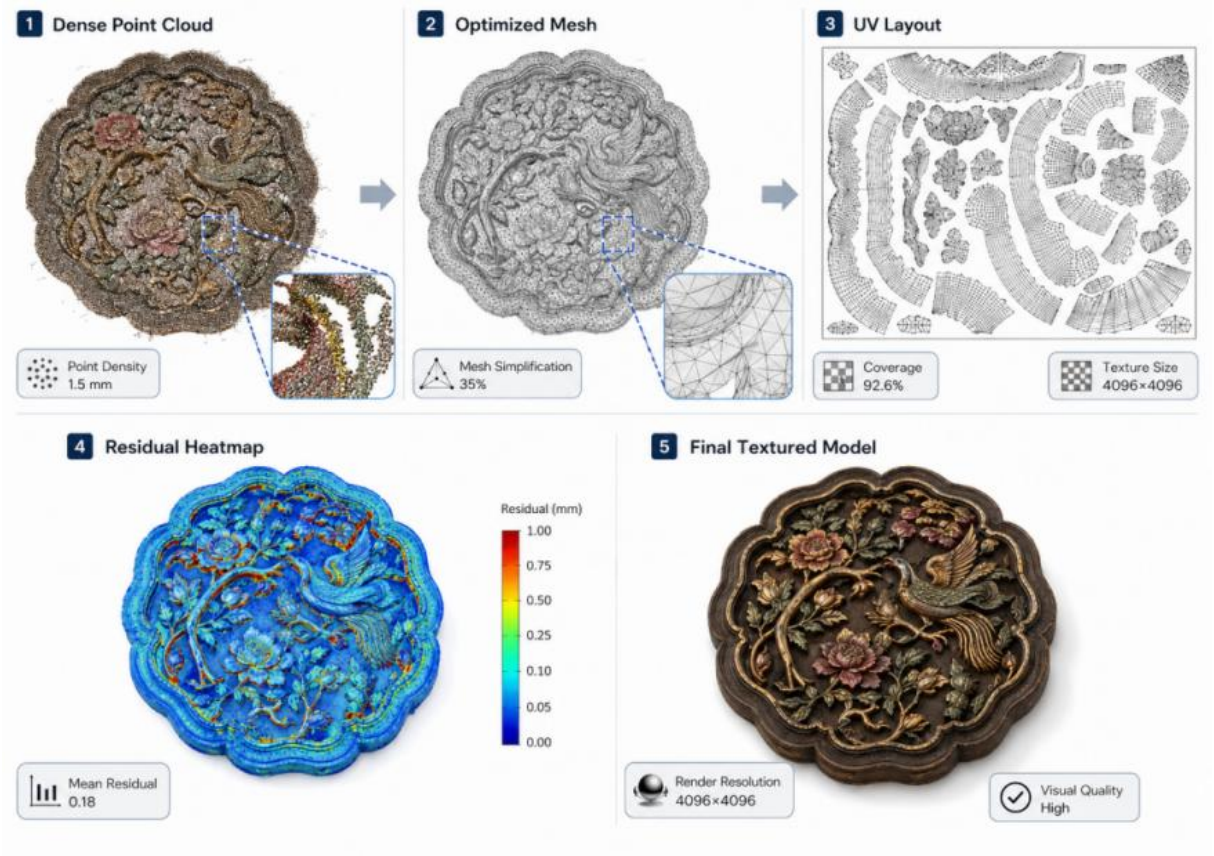


*Figure 4: Quality analysis of process pattern dense reconstruction and texture mapping*

The surface of the model maintains good continuity in the main pattern area, and the local residual is mainly concentrated in the hollow-out edge, the overlap area of the embroidery lines and the turning point of the stone relief. It can be corrected by improving the resolution of the local map and adding the boundary control points.

## 3.3 Semantic coding of process actions and mapping of interactive tasks

After the 3D reconstruction and texture mapping of the craft objects, the twin model still mainly stays at the observable level, and has not yet formed an operable and feed-back interactive object. The production process of Jilin traditional technology includes clear action sequence and tool dependence, such as folding, line tracing, cutting and unfolding in paper cutting, positioning, lead, stitch switching and edge closing of pattern in Manchu embroidery, and surface selection, rough carving, fine trimming and polishing in pine and flower stone carving. If these processes are only treated as video clips, the system cannot identify the process stage corresponding to the user's current operation, and it is difficult to judge whether the interactive task is completed. In order to make the digital twin model enter the practice scene of aesthetic education in colleges and universities, it is necessary to decompose the production action into encoded semantic units, and then map them to task nodes such as observation, recognition, sorting, simulation operation and feedback evaluation.

The process action sequence can be jointly described by video key frames, tool states, hand trajectories and process labels. Let the video sequence of the RTH technological process be Vr, the action time window be [ts,te], the key frame feature be ft, the tool state be gt, the hand trajectory be ht, and the process label be pt, then the action semantic coding can be expressed as follows.

$$a_r=\mathrm{Enc}\left(\sum_{t=t_s}^{t_e}\left(\alpha_f f_t+\alpha_g g_t+\alpha_h h_t+\alpha_p p_t\right)\cdot\kappa(t-t_s,t_e-t_s)\right) \tag{13}$$

where, ar represents the semantic encoding result of the RTH process action; Enc(·) denotes the action-encoding function. αf, αg, αh and αp represent the weights of key frames, tool states, hand trajectories and process labels, respectively. κ(·) represents the time window weighting function, which is used to highlight the action onset, transition, and completion phases. The coding method can transform the continuous production process into discrete action unit, so that the system can locate the corresponding process step and model area when the user triggers the task.

The production video is divided into several action segments according to the process stage, and each segment is bound with a process node number. The system extracts the object region, tool position and hand motion direction from the key frames, and then aliges them with the action description in the process text. The folding, tracing, cutting and unfolding of paper-cut samples have obvious action boundaries, and the contour changes and edge generation states are mainly retained during encoding. Stitch switching, stitch direction and pattern edge collection of embroidery samples are more critical. Tool trajectories and local pattern areas need to be recorded when coding. The rough carving, fine trimming and polishing stages of stone and wood carving samples are reflected as surface deformation, and the mesh surface area and normal changes should be bound when coding. After the above processing, process actions are no longer just video instructions, but computational units capable of driving changes in the state of the twin model.

Action semantic coding needs to be further mapped to aesthetic education interaction tasks in colleges and universities. Interaction tasks should not only include "viewing model" or "playing video", but also cover pattern area identification, process sequence judgment, local operation simulation, process difference comparison and task feedback recording. Suppose the action semantic encoding is ar, the interaction task is qj, the model region required by the task is zj, and the current state of the user is ut, then the matching score between the action and the task can be expressed as follows.

$$\Gamma(a_r,q_j)=\sigma\left(\beta_1 a_r^T e_j+\beta_2 IoU(z_r,z_j)+\beta_3 Sim(p_r,p_j)-\beta_4 C(u_t,q_j)\right) \quad (14)$$

where, Γ(ar,qj) represents the matching score between action unit ar and interaction task qj. ej represents the task vector; zr and zj represent the action association region and the task target region, respectively. IoU(·) denotes the degree of region overlap; Sim(·) represents the semantic similarity of the process; C(ut,qj) is the cost of executing the task in the user's current state. β1, β2, β3 and β4 are the weight coefficients. Let σ(·) denote the normalization function. In this model, action semantics, spatial region, process relationship and user state are brought into the same matching process to avoid the disconnection between interaction tasks and process steps.

After the process action is semantically encoded, it needs to establish a correspondence with the specific interaction task. Different actions have differences in data sources, semantic labels and task goals, and Table 4 lists the coding results and task mapping ways of typical process actions.

*Table 4: Craft Action Semantic Encoding and Interaction Task Mapping*

| Craft Action | Collected Data | Semantic Tag | Interaction Task |
|---|---|---|---|
| Folding Positioning | Video Key Frame, Edge Contour, Process Text | Fold-Position | Judge Folding Sequence and Observe Contour Change |
| Pattern Tracing | Local Image, Hand Trajectory, Tool Position | Pattern-Tracing | Mark Pattern Area and Match Corresponding Process |
| Cutting and Forming | Tool Trajectory, Boundary Change, Point Cloud Contour | Cut-Forming | Simulate Cutting Path and Detect Boundary Deviation |
| Threading Embroidery | Video Clip, Needle and Thread Trajectory, Pattern Texture Map | Stitch-Threading | Identify Stitch Direction and Complete Local Operation |
| Rough Carving and Shaping | Point Cloud Change, Normal Change, Tool Status | Rough-Carving | Select Processing Area and Observe Surface Change |
| Fine Polishing | Mesh Surface, Material Parameters, Action Clip | Polish-Finishing | Compare Texture and Roughness Before and After Polishing |

After the action coding and task mapping are completed, the system can call the corresponding process node according to the operation request of the student. When the user selects a pattern area, the front-end not only displays the local tile, but also returns the production action, tool type and process location associated with the area synchronously. When the user performs virtual cutting, embroidery or polishing tasks, the system determines whether the user enters the correct process according to the task matching score, and records the operation path, response delay and completion status. There is a many-to-many relationship between process action semantic coding and interaction tasks. An action may

correspond to multiple observation or simulation tasks, and a task may involve multiple continuous actions.

The relationship among action nodes, semantic labels, process regions and interaction tasks is not unidirectional. The same pattern region may be associated with multiple action segments, and the same interaction task may call multiple process nodes. To ensure that task triggering, state judgment, and feedback recording can be executed synchronously, Figure 5 constructs the relation index between semantic encoding of process actions and interactive tasks.

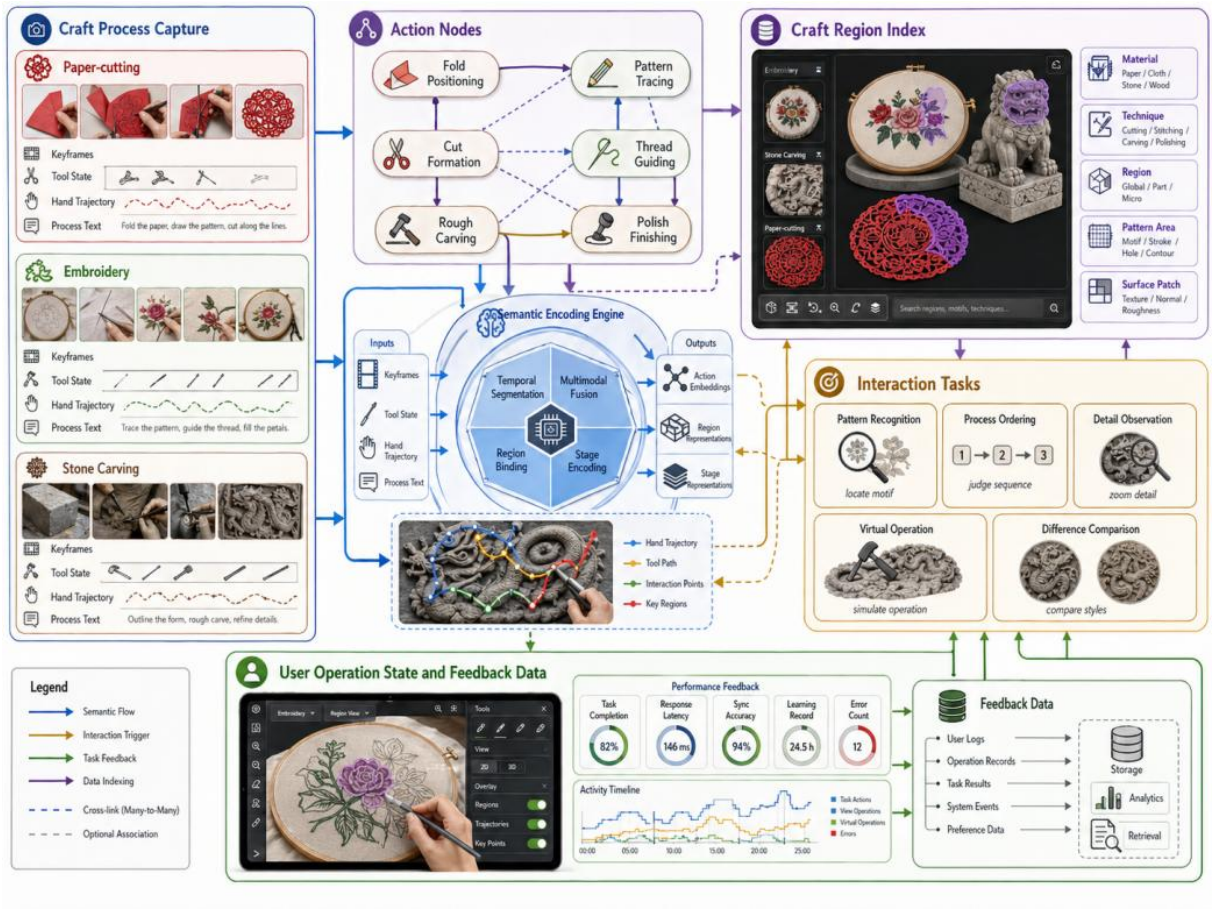


*Figure 5: Semantic coding of process actions and mapping diagram of interaction tasks*

The relationship structure binds process actions, 3D regions and user tasks in the same index. The system can quickly locate the corresponding model regions and process nodes after receiving the interaction request, and write the user operation results to the feedback data. This not only avoids the interaction tasks from being disconnected from the process flow, but also provides a clear data entry for subsequent status updates.

## 3.4 State update and interaction data synchronization mechanism of twin model

After the process action completes semantic encoding and task mapping, the twin model needs to continuously update its state according to user operations. State update does not simply change the display Angle of the model, but synchronously writes geometric pose, texture region, process node, interaction task and feedback data into the same state space. In the scene of aesthetic education in colleges and universities, students may continuously

perform rotation observation, pattern amplification, process selection, virtual cutting, local polishing and other operations. If the back-end twin state is not synchronized with the front-end rendering state, it will cause task judgment error, process node dislocation and interaction log distortion. In order to ensure the continuity of system operation, it is necessary to establish state recurrence, synchronization error and feedback correction mechanisms.

The state vector of the twin model at time t can be defined as follows.

$$S_t=\left[g_t,\tau_t,p_t,a_t,u_t,r_t\right] \tag{15}$$

where, St represents the current state of the twin model; gt represents geometric pose and spatial position; τt represents texture region and material display state; pt represents the process node; at represents the current action semantic unit. ut represents user interaction. rt represents the system feedback result. The state vector integrates model appearance, process and user behavior into a unified representation, so that each interaction can be corresponding to an explicit twin state.

When the user triggers a new interaction event, the system needs to combine the current state, action node, interaction input and time interval to complete the state recursion. The state update model can be expressed as follows.

$$S_{t+1}=F_s(S_t,A_t,U_t,\Delta t;\theta_s)=\sigma(W_sS_t+W_aA_t+W_uU_t+W_\Delta\Delta t+b_s) \tag{16}$$

where, St+1 represents the updated twin state; Fs(·) represents the state update function. At represents the current process action code; Ut represents user interaction input. Δt represents the time interval between adjacent states; Ws, Wa, Wu, WΔ are weight matrices; bs is the bias term; Let σ(·) denote the nonlinear mapping function. The model is able to describe the joint effects of virtual operations on process nodes, model poses and interactive feedback. For example, after a user finishes cutting a path in a paper-cutting task, the system not only updates the boundary state, but also synchronously updates the process progress, task completion mark and feedback record.

There needs to be consistency between the frontend rendered state and the backend twin state. Let the back-end state be $S_t^b$ and the front-end rendering state be $S_t^f$. The state synchronization error of both ends can be expressed as follows.

$$E_t^{sync}=\left\|S_t^b-S_t^f\right\|_2^2+\lambda_p\left\|p_t^b-p_t^f\right\|_1+\lambda_u\left|T_t^b-T_t^f\right| \tag{17}$$

where, $E_t^{sync}$ represents the synchronization error; $S_t^b$ and $S_t^f$ denote the back-end twin state and the front-end render state, respectively. $p_t^b$ and $p_t^f$ represent back-end process node and front-end display node respectively. $T_t^b$ and $T_t^f$ represent timestamps at both ends. λp and λu denote the process node bias versus time bias weights, respectively. When the synchronization error exceeds the threshold, the system gives priority to write back the back-end state, and repushes the current model pose, texture area and process node to avoid the inconsistency between the task interface seen by the user and the system judgment result.

Interaction logs are not only used for experimental statistics, but also for correcting twin model states. The user operation path, task completion, response delay and error number can reflect whether interactive tasks match with process nodes. The feedback correction model can be expressed as follows.

$$\hat{S}_{t+1}=S_{t+1}+\eta\left(\omega_1C_t+\omega_2e^{-L_t}-\omega_3E_t-\omega_4E_t^{sync}\right)\cdot H_t \tag{18}$$

where, $\hat{S}_{t+1}$ represents the corrected twin state; Let η denote the feedback correction step; Ct represents the task completion degree; Lt stands for interactive response delay; Et represents the number of user operation errors; $E_t^{sync}$ represents the synchronization error between front and back ends. ω1, ω2, ω3 and ω4 are the feedback weights; Ht represents the status mask associated with the current process node. This model makes the state correction only act on the relevant process area and task node, avoiding one wrong operation affecting the whole twin.

In the system implementation, interactive events enter the front-end event queue, and then are parsed by the server into task number, model area, action node and user state. The state update module calculates the next state according to Formula (16), the log module writes the user operation and response data, and the rendering end refreches the model display according to the update results. When the synchronization deviation is detected in Equation (17), the system recalieves the front-end display according to the back-end state. When Equation (18) obtains enough feedback data, the system locally corrects the task matching, node status and interaction prompt. This mechanism enables the Jilin traditional process twin model to maintain continuous state in the process of aesthetic education interaction in colleges and universities, and supports the analysis of task completion rate, response delay and state synchronization accuracy in subsequent system verification.

# 4 Systematic verification of cultural heritage activation and aesthetic education practice in colleges and universities

## 4.1 Virtual-Real Interaction and real-time rendering Implementation of craft twin scene

When the process twin scene was running, the system divided the Jilin traditional process object into five types of resources: geometric grid, texture map, semantic node, process script and interactive event. In the experimental scene, the number of mesh faces of a single process object is controlled from 82,000 to 216,000, the texture map is unified to 4096×4096, the LOD is set to 3 levels, the texture buffer space is set to 512 MB, and the state synchronization interval is 50 ms. After receiving user operations, the front-end rendering terminal writes rotation, scaling, pattern selection, process switching and virtual operations to the event queue. According to the task number, the server calls the corresponding twin object, LOD model and local texture cache. The state synchronization module writes the current process node, model pose, task completion status and log fields to the background in JSONL format. In the preliminary operation, the initial loading time of five types of process scenes is 1.42-2.85 s, and the frame rate of stable operation is 58.4-66.7 FPS.

The real-time rendering frame rate is used to measure the smoothness of the scene and is calculated as follows.

$$F_r=\frac{N_f}{\sum_{t=1}^{N_f}\left(T_t^{geo}+T_t^{tex}+T_t^{shade}+T_t^{event}+T_t^{sync}\right)} \tag{19}$$

where, Fr represents the average rendering frame rate; Nf represents the number of rendered frames in the statistical window; $T_t^{geo}$ represents the loading time of the geometric model in frame t. $T_t^{tex}$ is the texture reading time. $T_t^{shade}$ denotes the coloring computation time; $T_t^{event}$ represents the interaction event processing time. Ttsync denotes the state synchronization time. This index can reflect the comprehensive impact of model complexity,

texture invocation and event processing on scene real-time performance.

The process twin scenario involves model resource call, texture resource reading, interactive event response and state data writeback during operation. It is difficult to describe each module separately to reflect the collaborative relationship between the rendering end and the interaction end. In order to illustrate the data transmission process between real-time rendering and virtual-real interaction, this paper constructs the running pipeline as shown in Figure 6.

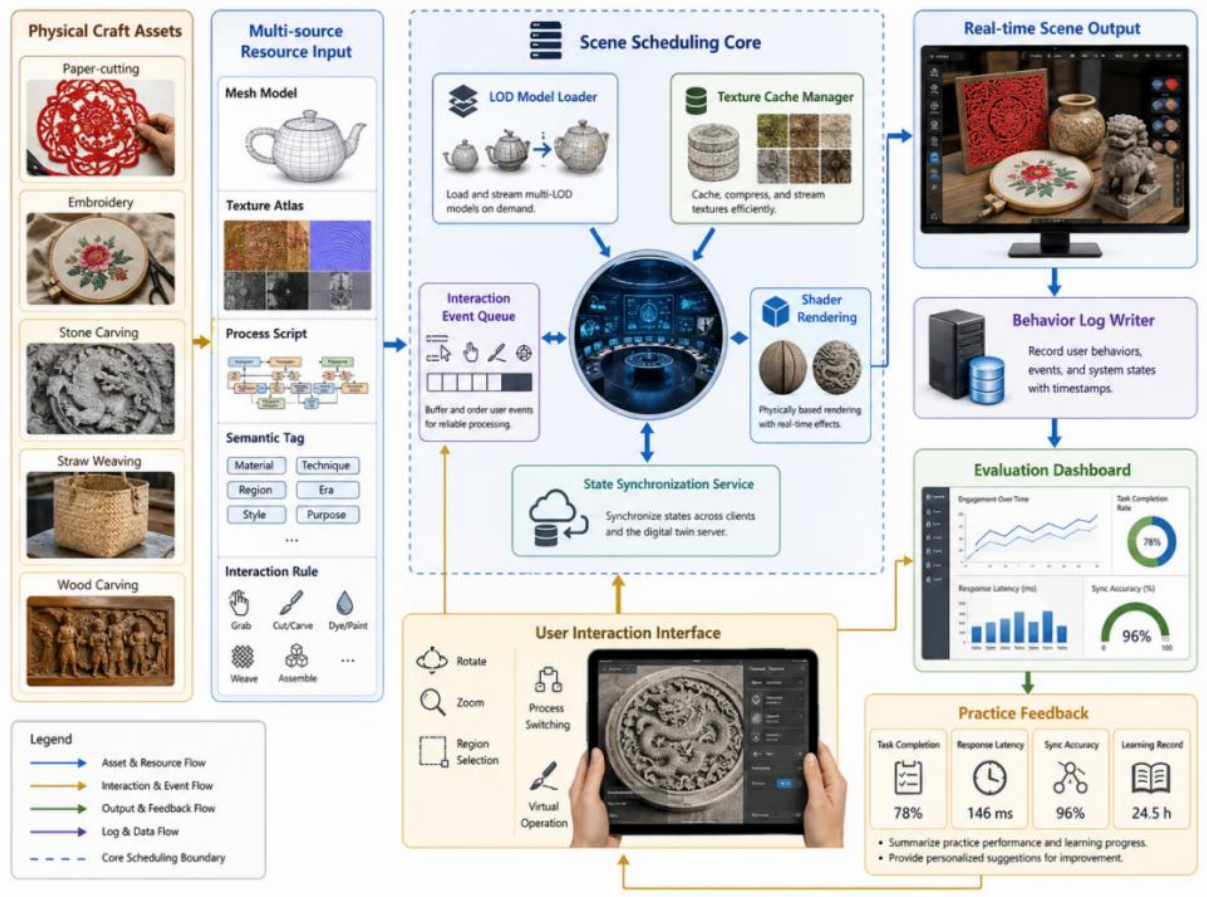


*Figure 6: Virtual-real interaction and real-time rendering pipeline diagram of process twin scene*

This pipeline organizes user input, event queuing, LOD model loading, texture caching, shader rendering, state synchronization, and behavior log writing into the same execution sequence. The high precision model is only loaded on demand in local observation and virtual operation tasks, the texture buffer preferentially retains the current pattern area, and the interactive events enter the state update module in time order, which provides a running basis for subsequent tests of response delay, frame rate stability and task completion rate.

## 4.2 Experimental environment, process data set and evaluation index design

The experiment uses the self-built Jilin-Craftdt Dataset as the data base, covering five types of traditional craft objects: Jilin paper cutting, Manchu embroidery, pine and flower stone carving, straw weaving and wood carving. The data set contains 50 craft samples, 10 pieces of each type, 2500 multi-view images, 50 groups of point cloud data, 800 partial images of

patterns, 120 production videos, 640 action clips, 1860 semantic tags and 4800 interaction logs. Sixty students were selected to participate in the interactive test of aesthetic education in colleges and universities. The types of tasks included pattern recognition, process sorting, local observation and virtual operation. See Table 5 for the experimental environment and dataset configuration.

*Table 5: Experimental Environment and Jilin-CraftDT Dataset Configuration*

| Category | Project | Configuration Value |
|---|---|---|
| Hardware | CPU | Intel Xeon Silver 4314 × 2 |
| | GPU | NVIDIA RTX 4090 24GB |
| | Memory | 128 GB |
| Software | Operating System | Ubuntu 22.04 |
| | Framework | PyTorch 2.2 |
| Rendering | Engine | Unity 2022 LTS |
| Dataset | Dataset Name | Jilin-CraftDT Dataset |
| | Craft Types | Paper-cutting, Embroidery, Stone Carving, Straw Weaving, Wood Carving |
| | Craft Objects | 50 |
| | Multi-view Images | 2500 |
| | Point Cloud Groups | 50 |
| | Pattern Patches | 800 |
| | Process Videos | 120 |
| | Action Clips | 640 |
| | Semantic Labels | 1860 |
| | Interaction Logs | 4800 |
| Evaluation | Participants | 60 students |
| | Tasks | Pattern Recognition, Process Sorting, Detail Observation, Virtual Operation |

In order to reflect the differences between the proposed method in 3D reconstruction and interactive verification, three kinds of comparison schemes are set up. Static 3D Display only supports static model browsing. VR Exhibition System supports virtual exhibition scene roaming; Point Cloud-only Modeling adopts Point Cloud scanning modeling without introducing image texture fusion and process semantic coding. The proposed method uses image point cloud fusion, pattern mapping, action coding and state synchronization mechanisms, and the experimental setup is shown in Table 6.

*Table 6: Comparative Methods and Experimental Settings*

| Comparative Method | Core Processing Method | Validation Content |
|---|---|---|
| Static 3D Display | Static 3D Model Display | Basic Observation and Model Browsing |
| VR Exhibition System | Virtual Exhibition Scene Loading | Scene Browsing and Basic Interaction |
| Point Cloud-only Modeling | Point Cloud Scanning Modeling | Geometric Reconstruction and Spatial Display |
| Proposed Digital Twin System | Image-point Cloud Fusion, Semantic Encoding, State Synchronization | Reconstruction Accuracy, Interaction Task and Feedback Validation |

The reconstruction error is used to measure the geometric deviation between the digital twin model and the reference point cloud, and is calculated as follows.

$$E_{rec}=\frac{1}{N}\sum_{i=1}^{N}\min_{q_j\in Q}\left\|p_i-q_j\right\|_2 \tag{20}$$

where, Erec represents the average reconstruction error; pi represents the sampling points of the reconstructed model; Q denotes the set of reference point clouds; qj denotes the neighboring points in the reference point cloud; N denotes the number of sampling points. A smaller value indicates that the 3D model is closer to the reference scan data.

Interactive response delay is used to measure the time cost between user input and system feedback, and is calculated as follows.

$$L_{res}=\frac{1}{M}\sum_{j=1}^{M}\left(t_j^{out}-t_j^{in}\right) \tag{21}$$

where, Lres represents the average response delay; $t_j^{in}$represents the JTH interaction input time; $t_j^{out}$represents the system feedback output time; M denotes the number of interactive operations. The subsequent result analysis focuses on reconstruction error, texture similarity, registration error, frame rate, video memory occupation, task completion rate, synchronization accuracy and the number of errors.

## 4.3 Reconstruction accuracy, interactive response and practical data verification analysis

Based on Jilin-CraftDT Dataset and the comparison schemes listed in Table 6, this paper carried out verification from three aspects: 3D reconstruction quality, rendering operation efficiency, and interactive performance of aesthetic education in colleges and universities. The interactive task was completed by 60 students, and the system automatically recorded the task completion status, response delay, synchronization accuracy, and number of errors, and the results are shown in Table 7.

*Table 7: Test Results of Interactive Tasks in College Aesthetic Education*

| Task Type | Completion Rate/% | Response Latency/ms | Synchronization Accuracy/% | Average Number of Errors |
|---|---|---|---|---|
| Pattern Recognition | 94.6 | 72 | 97.2 | 0.8 |
| Process Sorting | 89.4 | 96 | 94.8 | 1.6 |
| Detail Observation | 93.2 | 78 | 96.1 | 1.1 |
| Virtual Operation | 86.7 | 118 | 93.8 | 2.1 |

In order to compare the stability of different modeling methods in the 3D reconstruction of process objects, this paper counted the reconstruction error distribution of 50 process samples, and selected typical samples to observe the local texture residual. The results are shown in Figure 7.

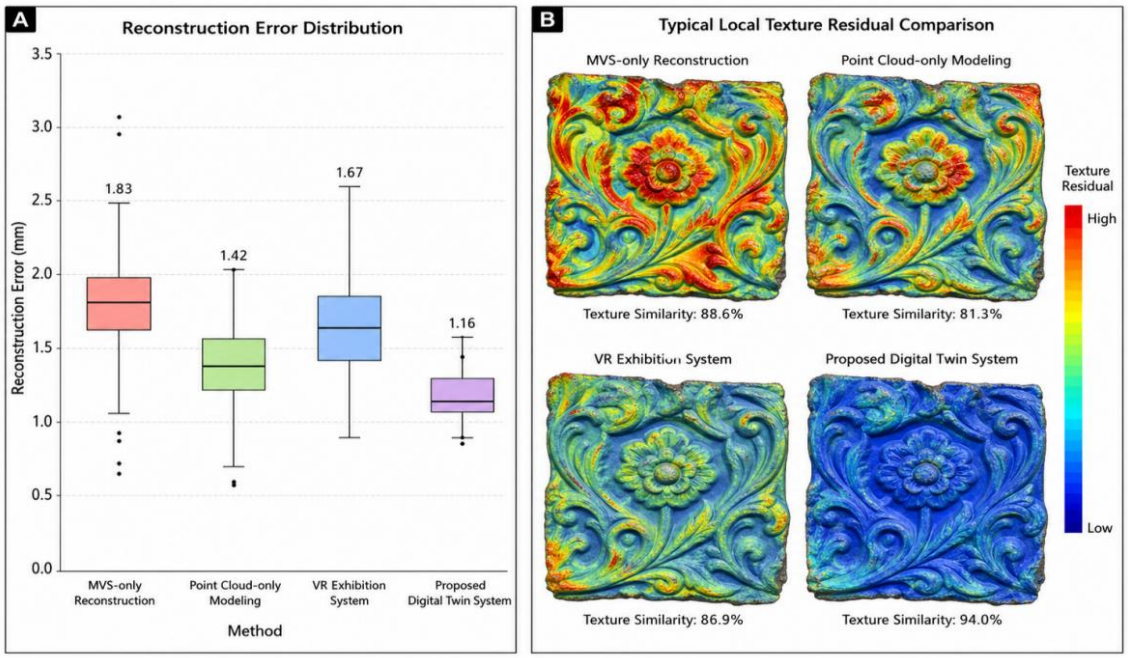


*Figure 7: Comparison of 3D reconstruction error distribution and texture residual under different methods*

The average Reconstruction error of MVS-only reconstruction is 1.83 mm, Point Cloud-only Modeling is 1.42 mm, VR Exhibition System is 1.67 mm, and the average reconstruction error of MVS-only reconstruction is 1.83 mm. The system in this paper is reduced to 1.16mm; The texture similarity is improved from 81.3%-88.6% to 94.0%. The texture residual is mainly concentrated in the hollow-out edge, the overlap area of embroidery lines and the turning point of stone relief, which indicates that the combination of image texture constraint and point cloud geometric constraint can improve the contour fit of the artifact and the reduction effect of the pattern.

In order to analyze the differences in resource occupancy of objects with different processes in real-time rendering, this paper records the average frame rate, video memory occupancy and model complexity of five types of objects, and the statistical results are shown in Figure 8.

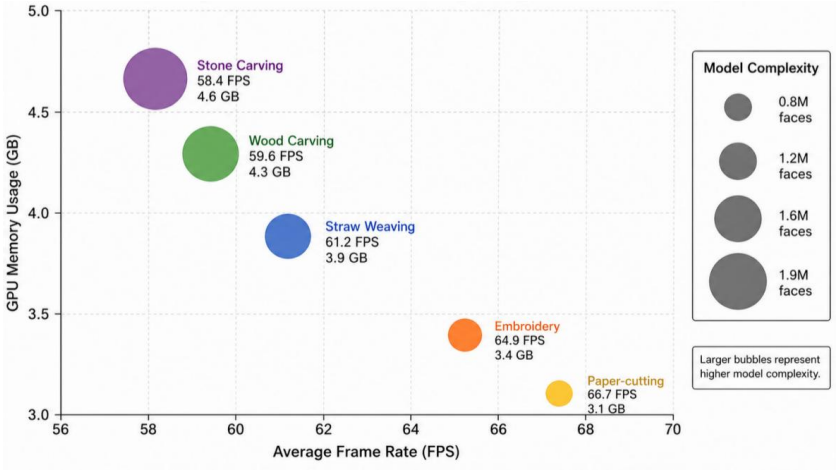


*Figure 8: Bubble chart of rendering frame rate - video memory occupancy for different process objects*

The average frame rate of Jilin paper-cut and Manchu embroidery is 66.7FPS and 64.9FPS, respectively, and the video memory occupation is 3.1GB and 3.4GB. The frame rates of stone carving, straw carving and wood carving are 58.4FPS, 61.2FPS and 59.6FPS respectively, and the video memory consumption is 3.8-4.6GB. The frame rate of five types of objects keeps above 58 FPS, which indicates that the LOD hierarchical loading and texture caching mechanism can support the stable operation of interactive scenes.

In order to verify the comprehensive interactive effect of the System in aesthetic education tasks in colleges and universities, this paper selects the Static 3D Display and VR Exhibition System to compare with the proposed system, and the results are shown in Figure 9.

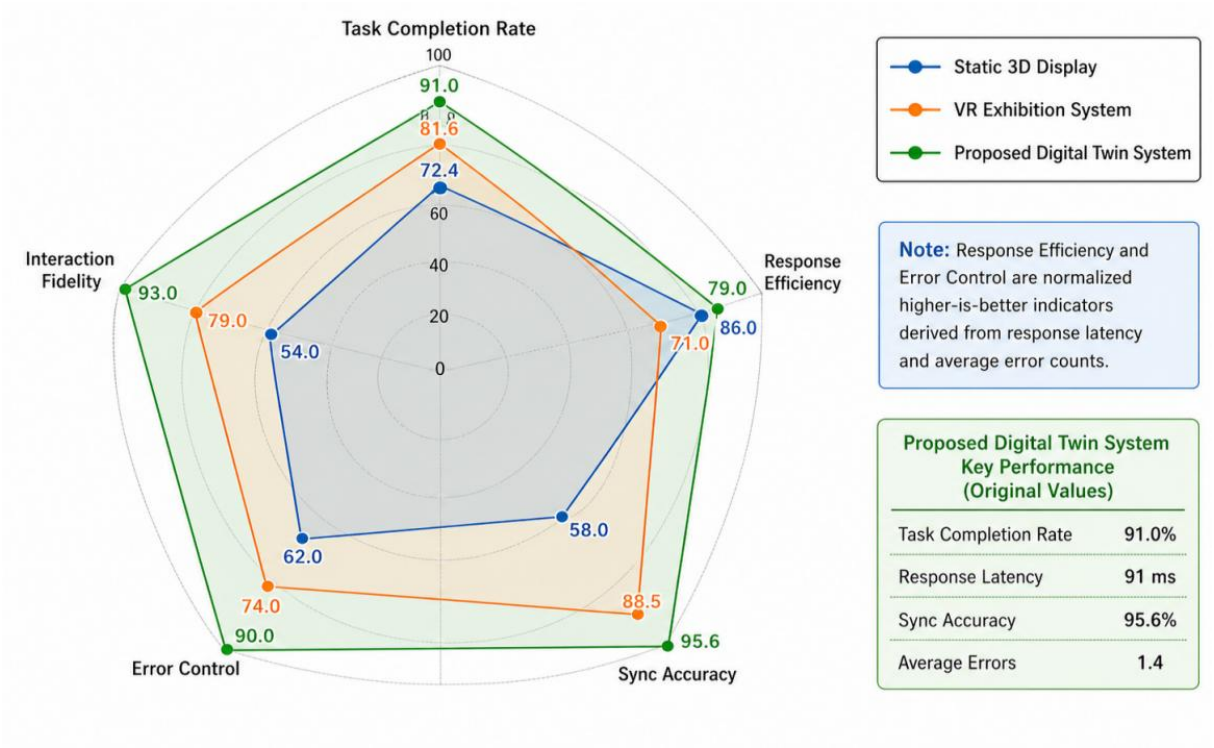


*Figure 9: Radar chart of interaction performance for different system schemes*

The task completion rate of Static 3D Display is 72.4%, that of VR Exhibition System is 81.6%, and that of the proposed system is 91.0%. The average response delay of the system is 91 ms, the synchronization accuracy is 95.6%, and the average number of errors is reduced to 1.4. Combined with Table 7, it can be seen that the completion rate of pattern recognition and local observation tasks is high, and the response delay of virtual operation tasks increases to 118 ms due to superposition of action matching, status update and front-end feedback. The overall results show that the proposed system has quantifiable verification results in terms of reconstruction accuracy, rendering efficiency and interactive task execution.

# 5 Conclusion

Oriented to the digital twin modeling of Jilin traditional craft cultural heritage and the interactive verification of aesthetic education in colleges and universities, this paper constructs a technical link of "multi-source acquisition, twin modeling, interaction driving-performance verification". The system encodes the image, point cloud, video, process text and interaction log in a unified way, and completes the image point cloud registration,

dense reconstruction, mesh optimization, texture mapping and action semantic coding of process objects. In the experimental verification, the reconstruction error, texture similarity, response time delay, task completion rate and synchronization accuracy are included in the same evaluation framework to test the accuracy of the model and the interactive operation effect. This method transforms the traditional craft object from static digital material into a twin unit that can be called, feedback and verified, and provides an executable engineering scheme for the revitalization of Jilin traditional craft cultural heritage and the practice of aesthetic education in colleges and universities. In the future, the local repair of complex patterns, multi-user concurrent synchronization and cross-process sample migration modeling can be further optimized.

# Author's Profile

Wenjian Yan, Gender: Female Date of Birth: June 3, 1982 Professional Title: Lecturer Highest Education: Master's Degree Highest Degree: Master of Arts Graduated from: Northeast Normal University Research Focus: Pre-production Animation and Arts and Crafts.

# Funding

This work was supported by.Jilin Provincial Education Science "14th Five-Year Plan" 2024 Annual Research Project: "Research on Implementation Pathways of Integrating Jilin's Folk Traditional Art into University Aesthetic Education under Digitalization Background"

# References

[1] DANG X, LIU W, HONG Q, et al. Digital twin applications on cultural world heritage sites in China: A state-of-the-art overview[J]. Journal of Cultural Heritage, 2023, 64: 228-243.

[2] LUTHER W, BALOIAN N, BIELLA D, et al. Digital twins and enabling technologies in museums and cultural heritage: An overview[J]. Sensors, 2023, 23(3): 1583.

[3] LI M, XU S, TANG J, et al. Design and research of digital twin platform for handicraft intangible cultural heritage-Yangxin Cloth Paste[J]. Heritage Science, 2024, 12(1): 43.

[4] GUO Y, LIU L, HUANG W, et al. Extending X-reality technologies to digital twin in cultural heritage risk management: A comparative evaluation from the perspective of situation awareness[J]. Heritage Science, 2024, 12(1): 245.

[5] PAN X, LIN Q, YE S, et al. Deep learning based approaches from semantic point clouds to semantic BIM models for heritage digital twin[J]. Heritage Science, 2024, 12(1): 65.

[6] AMELIO A, ZARRI G P. Cultural heritage digital twin: Modeling and representing the visual narrative in Leonardo Da Vinci’s Mona Lisa[J]. Neural Computing and Applications, 2024, 36(20): 11859-11876.

[7] NICCOLUCCI F, FELICETTI A. Digital twin sensors in cultural heritage applications[J]. Sensors, 2024, 24(12): 3978.

[8] PAOLANTI M, PUGGIONI M, FRONTONI E, et al. Evaluating learning outcomes of virtual reality applications in education: A proposal for digital cultural heritage[J]. ACM Journal on Computing and Cultural Heritage, 2023, 16(2): 1-25.

[9] GABELLONE F. Development of an immersive VR experience using integrated survey technologies and hybrid scenarios[J]. Heritage, 2023, 6(2): 1169-1186.

[10] 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.

[11] MURTIYOSO A, PELLIS E, GRUSSENMEYER P, et al. Towards semantic photogrammetry: Generating semantically rich point clouds from architectural close-range photogrammetry[J]. Sensors, 2022, 22(3): 966.

[12] PEPE M, ALFIO V S, COSTANTINO D, et al. Data for 3D reconstruction and point cloud classification using machine learning in cultural heritage environment[J]. Data in Brief, 2022, 42: 108250.

[13] PATRUCCO G, BAMBRIDGE P, GIULIO TONOLO F, et al. Digital replicas of British Museum artefacts[J]. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2023, XLVIII-M-2-2023: 1173-1180.

[14] LIN X, XUE B, WANG X. Digital 3D reconstruction of Ancient Chinese Great Wild Goose Pagoda by TLS point cloud hierarchical registration[J]. ACM Journal on Computing and Cultural Heritage, 2024, 17(2): 1-16.

[15] ZHOU L, WU G, ZUO Y, et al. A comprehensive review of vision-based 3D reconstruction methods[J]. Sensors, 2024, 24(7): 2314.

[16] XU L, XU Y, RAO Z, et al. Real-time 3D reconstruction for the conservation of the Great Wall's cultural heritage using depth cameras[J]. Sustainability, 2024, 16(16): 7024.

[17] STOEAN R, BACANIN N, STOEAN C, et al. Bridging the past and present: AI-driven 3D restoration of degraded artefacts for museum digital display[J]. Journal of Cultural Heritage, 2024, 69: 18-26.

[18] WANG Y, BI W, LIU X, et al. Overcoming single-technology limitations in digital heritage preservation: A study of the LiPhoScan 3D reconstruction model[J]. Alexandria Engineering Journal, 2025, 119: 518-530.

[19] HUANG S, HU Q, AI M, et al. Weakly supervised 3D point cloud semantic segmentation for architectural heritage using teacher-guided consistency and contrast learning[J]. Automation in Construction, 2024, 168: 105831.

[20] WANG H, SUN Y, WANG H, et al. Gamifying cultural heritage: Exploring the potential of immersive virtual exhibitions for cultural heritage preservation and dissemination[J]. Computers in Human Behavior: Artificial Humans, 2024, 2(2): 100046.