

PAPER

Mobile Augmented Reality–Driven Living Transmission of Intangible Cultural Heritage along the Grand Canal

Xi Zhao¹, Zhao Zhang¹(✉),
 Aimin Li²

¹Hebei Art & Design Academy,
 Baoding, China

²Tangshan Normal University,
 Tangshan, China

Zzn1112@163.com

ABSTRACT

New pathways for immersive presentation and community participation in the living transmission of intangible cultural heritage along the Grand Canal are provided by mobile augmented reality. However, severe challenges to spatial registration accuracy, robustness to network disconnection, and content update mechanisms are posed by weak network coverage in linear heritage environments, limited mobile-device computing power, and the need for intangible cultural heritage content to grow dynamically. A mobile augmented reality method centered on semantic anchors was proposed. Semantic anchors were generated through joint factor graph optimization of lightweight visual semantic features and global navigation satellite system/inertial measurement unit measurements, forming a geographically consistent and compact spatial index. Each anchor was stored compactly at the hundreds-of-bytes level, and decimeter-level spatial registration was achieved without prebuilt high-precision maps. On this basis, a device-edge-cloud collaborative architecture was constructed. Semantic anchors were simultaneously employed as persistent primitives for cross-session pose recovery and as spatial containers for user-generated intangible cultural heritage micro-narratives so that second-level relocation and continuous display were supported under fully disconnected conditions. Through edge-side lightweight quality assessment and context-aware recommendation, community-based accumulation and targeted distribution of living content were achieved. It is demonstrated by the experimental results that, for mobile augmented reality-driven living transmission of large-scale linear cultural heritage, a lightweight, robust, and engineering-feasible technical pathway is provided by the semantic-anchor method.

KEYWORDS

mobile augmented reality, semantic anchors, edge computing, intangible cultural heritage, weak-network adaptation, user-generated content

1 INTRODUCTION

The intangible cultural heritage along the Grand Canal is tightly coupled with real-world spatial environments, and its living transmission requires that physical-space

Zhao, X., Zhang, Z., Li, A. (2026). Mobile Augmented Reality–Driven Living Transmission of Intangible Cultural Heritage along the Grand Canal. *International Journal of Interactive Mobile Technologies (IJIM)*, 20(20), pp. 169–182. <https://doi.org/10.3991/ijim.v20i20.63555>

Article submitted 2026-06-26. Revision uploaded 2026-08-30. Final acceptance 2026-09-02.

© 2026 by the authors of this article. Published under CC-BY.

limitations be overcome so that cultural presentation can be made perceivable, participatory, and sustainably growing in real environments [1]. An effective pathway for this purpose has been provided by mobile augmented reality [2]. Recent systems engineering research has further shown that immersive cultural applications can be organized as closed-loop architectures integrating contextual data acquisition, adaptive content processing, constraint validation, immersive rendering, and user feedback [3]. However, the Grand Canal heritage belt is linearly distributed and spans a large spatial extent [4]; network coverage along the route is uneven; mobile devices are strictly constrained in computation, storage, and battery endurance [5]; and intangible cultural heritage content is also required to be dynamically accumulated through community participation [6]. Requirements for spatial registration accuracy, weak-network robustness, and content update mechanisms that are far more stringent than those in general scenarios are imposed by these characteristics [7]. Therefore, research on mobile augmented reality methods adapted to mobile devices and weak-network environments is of clear academic and engineering value for the living transmission of Grand Canal intangible cultural heritage.

Three main categories of limitations are associated with existing mobile augmented reality spatial registration methods. In some cloud-anchor methods, cloud-maintained visual maps or spatial features are relied upon [8], substantial storage and transmission overheads are incurred, and stable cross-session relocalization is difficult to guarantee under network interruption [9]. Severe drift is exhibited by pure visual odometry methods in canal scenes with weak or repetitive textures [10], and high computational cost is incurred [11]. Meter-level accuracy is typically provided by global navigation satellite system single-point positioning on commodity mobile devices, and fine-grained alignment of intangible cultural heritage content cannot be supported [12]. In addition, in existing cultural heritage augmented reality systems, content is largely treated as static resources [13]; spatial association of user-generated content [14] and mechanisms for quality control and conflict resolution [15] are lacking; and the community-participation closed loop required for living transmission is difficult to form. Practical deployment of mobile augmented reality for large-scale linear cultural heritage is constrained by these deficiencies.

Accordingly, a semantic-anchor-driven mobile augmented reality method for living transmission is proposed. A geographically consistent and compact spatial index is generated through joint optimization of lightweight visual semantic features and global navigation satellite system/inertial measurement unit measurements, and a device-edge-cloud collaborative architecture is constructed around this index. Three functions are simultaneously served by semantic anchors: spatial registration primitives, cross-session persistent indices, and spatial containers for user-generated content. In conjunction with edge-side lightweight quality assessment, context-aware recommendation, and dynamic caching strategies, system availability and content distribution efficiency under weak-network conditions are effectively improved. The comprehensive performance of the method is validated through prototype deployments in the Yangzhou and Hangzhou sections of the Grand Canal.

2 SYSTEM MODEL AND PROBLEM DEFINITION

A linear heritage region of length L along the Grand Canal is considered, along which M mobile user terminals and K edge nodes are deployed. The edge nodes are mobile edge computing servers or 5G small base stations with computing capability. A three-tier device-edge-cloud collaborative architecture is adopted. Environmental sensing, augmented reality interaction, and local pose tracking are undertaken by

the terminals; anchor aggregation, content caching, lightweight quality assessment, and local content distribution are handled by the edge nodes; and the global anchor index, the intangible cultural heritage content repository, and the user interest model are maintained by the cloud. As the core of this architecture, a class of lightweight and geographically consistent spatial indexing units, referred to as semantic anchors, is introduced. Semantic anchors are not pre-calibrated fixed identifiers; rather, they are generated online through joint optimization of terminal visual, global navigation satellite system, and inertial measurement unit observations during user mobility. Each anchor is associated with a physical component or spatial location exhibiting stable semantic attributes in the canal scene. Spatial registration, cross-session recovery, and precise anchoring of user-generated content are intended to be simultaneously supported at minimal storage and transmission cost. For an intuitive illustration of the overall technical workflow, the overall framework of the semantic-anchor-driven mobile augmented reality method for living intangible cultural heritage transmission is presented in Figure 1. In this framework, with semantic anchors as the core, multi-source sensing, spatial registration, cross-session recovery, and device-edge-cloud collaborative content updating and distribution are uniformly organized.

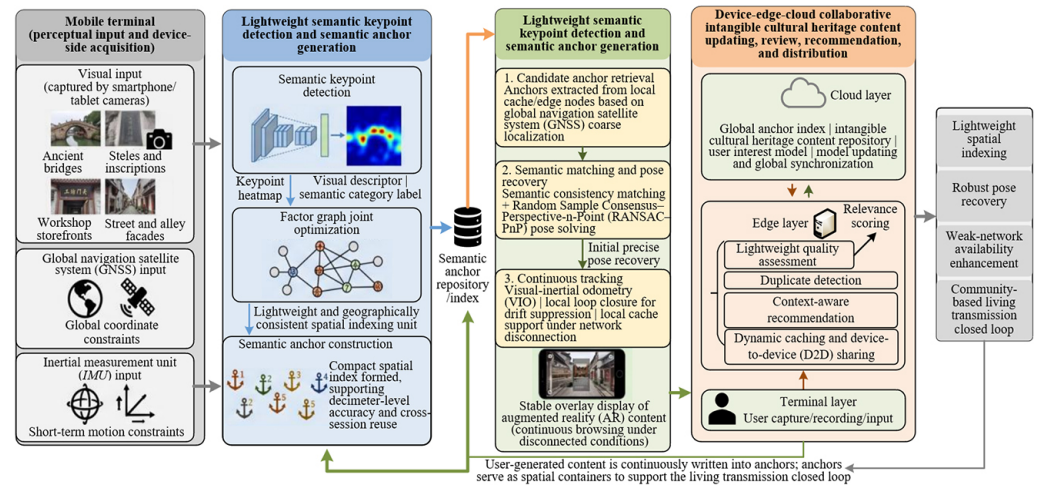


Fig. 1. Overall framework of the semantic-anchor-driven mobile augmented reality method for living transmission of intangible cultural heritage

A semantic anchor A_i is defined as a five-tuple $A_i = \{p_{w,i}, d_i, l_i, C_i, t_i\}$, where $p_{w,i} \in \mathbb{R}^3$ is the three-dimensional coordinate of the anchor in the world coordinate frame, determined by joint optimization and endowed with global geographic consistency; $d_i \in \mathbb{R}^D$ is the aggregated visual descriptor, with dimension typically set to 64, used for cross-session matching and pose solving; l_i is the semantic label indicating the component category to which the anchor belongs, such as an ancient bridge balustrade, stele inscription, or intangible cultural heritage workshop storefront; C_i is a set of pointers to associated intangible cultural heritage digital content, allowing multiple user-generated micro-narratives to be aggregated at a single anchor; and t_i is the last update timestamp, used for version control and incremental synchronization. Intangible cultural heritage content $c \in C$ encompasses short videos, audio, and text, and each content item is accompanied by contextual metadata such as location, time, and user identifier. User context $u = \{p_u, v_u, h_u, J_u\}$ denotes the user's current position, velocity, historical interaction behavior vector, and interest vector, respectively. Edge node e_j is constrained by cache capacity B_j and computing capability F_j , and its caching decision is required to maximize cache hit utility under limited resources.

Under the above system model, the core problem can be formalized as a multi-objective strategy design that jointly accounts for spatial registration and content distribution. The optimization variable is defined as a strategy set P , which includes anchor generation and matching parameters, edge caching decisions, and recommendation ranking weights. Localization accuracy, recovery timeliness, content quality, and cache efficiency are simultaneously considered in the objective function, which is expressed as:

$$\min_P E[\tau_{recover} + \lambda_1 \cdot \epsilon_{reg}] - \lambda_2 \cdot E[Q_{content} + U_{cache}] \quad (1)$$

where, $\tau_{recover}$ is the first augmented reality content recovery latency after a user enters the neighborhood of an anchor, ϵ_{reg} is the spatial registration error, $Q_{content}$ is the content quality score output by the edge-side quality assessment model, U_{cache} is the cache hit utility, and λ_1 and λ_2 are weight coefficients balancing localization accuracy and content experience. Constraints include the parameter count of the terminal semantic feature extraction model not exceeding the mobile storage threshold; the cache occupancy of an edge node not exceeding B_j ; and the end-to-end content acquisition latency does not exceed the interaction tolerance threshold. This formalization is used to uniformly characterize the system design objectives of spatial registration and content distribution. Subsequently, semantic anchor generation and matching, cache placement, and recommendation ranking are implemented separately, and overall performance is evaluated through experiments.

3 SEMANTIC-ANCHOR-DRIVEN SPATIAL REGISTRATION METHOD FOR MOBILE AUGMENTED REALITY

Reliable spatial localization capability under extremely low computational and storage overhead is required by the real-time constraints of mobile devices. Therefore, a lightweight semantic keypoint detector Φ based on a MobileNetV3 backbone is adopted. For a single-frame input, a keypoint heatmap, a 128-dimensional visual descriptor d , and a semantic category probability distribution $\hat{s}$ for each keypoint are simultaneously output by this network. Components with stable discriminability in canal scenes, such as ancient bridge balustrades, stele inscriptions, and intangible cultural heritage workshop storefronts, are covered by the semantic categories. The detector is trained through a self-distillation strategy, by which a lightweight student network is made to fit the output distribution of a large-scale teacher network with respect to semantic consistency. The training loss is defined as:

$$L_{det} = L_{heatmap} + \alpha L_{desc} + \beta L_{semantic} \quad (2)$$

where, $L_{heatmap}$ is the per-pixel regression loss of the keypoint heatmap, L_{desc} is a contrastive loss used to pull matched keypoint descriptors closer and push unmatched ones apart, $L_{semantic}$ is the semantic classification cross-entropy loss, and α and β are balancing coefficients. The network parameter count is controlled within 2M, and the per-frame inference latency on mobile devices is maintained below 15 ms so that a stable and low-latency visual front end is provided for subsequent factor graph optimization.

After visual-semantic observations are obtained, global navigation satellite system, inertial measurement unit, and visual information are jointly incorporated

into a factor graph for joint optimization so that global navigation satellite system multipath errors in urban canyons and along the canal, as well as cumulative drift of visual odometry, are mitigated. The optimization variable X consists of the terminal pose sequence $\{T_t\}$ and the three-dimensional coordinates $\{p_{w,t}\}$ of candidate anchors in the world coordinate frame. The objective function is:

$$\min_X \sum_{i \in G} \|r_{GNSS,i}\|_{\Omega_{g,i}}^2 + \sum_{j \in J} \|r_{IMU,j}\|_{\Omega_{i,j}}^2 + \sum_{k \in V} \|r_{vis,k}\|_{\Omega_{v,k}}^2 \quad (3)$$

where, $r_{GNSS,i}$ is the global navigation satellite system position residual, $r_{IMU,j}$ is the inertial measurement unit pre-integration residual, $r_{vis,k}$ is the visual-semantic residual, and $\Omega_{g,i}$, $\Omega_{i,j}$ and $\Omega_{v,k}$ are the corresponding information matrices that reflect the noise levels of the observations. The visual residual is defined as the reprojection error between the projection of a matched anchor into the current frame and the observed semantic keypoint. Unlike conventional visual matching, a semantic consistency constraint is introduced when matches are established: a valid match is considered only when the two keypoints share the same semantic category and the descriptor distance is below a threshold τ_d . The mismatch rate caused by repetitive textures and symmetric structures in canal scenes is significantly reduced by this constraint. The optimization is solved using the incremental smoothing and mapping backend in Georgia Tech Smoothing and Mapping (GTSAM), and the per-frame optimization time is kept below 20 ms.

When a keypoint is stably observed across N consecutive frames and its semantic confidence exceeds a preset threshold, it is promoted to a semantic anchor. A compact structure is adopted for anchor storage: the descriptor is reduced to 64 dimensions by principal component analysis, coordinates are stored at float32 precision, and semantic labels are encoded using one-hot encoding. Each anchor is represented in a compact form at the hundreds-of-bytes level, and storage and transmission overheads are significantly reduced relative to cloud anchors that contain hundreds of local three-dimensional points. The anchor index is organized as a geographic grid-based hash table, and range queries centered at the current user position are supported. The computational complexity of the overall pipeline is $O(T \cdot N_k \cdot D)$, where T is the number of frames involved in optimization, N_k is the number of semantic keypoints per frame, and D is the dimension of the reduced descriptor. Under typical configurations, $D = 64$ and $N_k < 200$, so real-time operation on mobile devices is enabled. The storage complexity is $O(|A|)$, scaling linearly with the number of anchors in the scene and is far lower than that of methods based on dense point clouds or global maps. Thus, a necessary prerequisite is provided for local caching and cross-session recovery under weak-network conditions.

4 CROSS-SESSION PRECISE RECOVERY AND WEAK-NETWORK ADAPTIVE TRACKING

The core challenge of cross-session recovery is that, after a scene has been left and is subsequently re-entered, a precise pose must be rapidly recovered without a preloaded map and possibly under weak-network or disconnected conditions so that stable overlay of intangible cultural heritage content is ensured. For this purpose, a coarse-to-fine recovery pipeline is adopted. Coarse localization is first performed at the terminal through tightly coupled global navigation satellite system/inertial

measurement unit integration, the approximate region in which the user is located is determined, and a candidate anchor set A_{cand} centered at that position with radius R , is retrieved from the local cache or an edge node. Subsequently, semantic keypoints and descriptors are extracted from the current frame on the terminal side and matched against the candidate anchors through semantic-guided nearest-neighbor matching. A bidirectional consistency check is performed during matching, and the two keypoints are simultaneously required to share the same semantic label, so that a high-confidence match set M is obtained. Based on this match set, the current camera pose T is estimated using the Random Sample Consensus–Perspective-n-Point (RANSAC–PnP) algorithm, and the optimization objective is:

$$\min_T \sum_{(i,j) \in M} \left(\left\| u_i - \pi(T, p_{w,j}) \right\|^2 \right) \quad (4)$$

where, u_i is the pixel coordinate of the i -th matched keypoint in the current frame, $p_{w,j}$ is the three-dimensional coordinate of the j -th anchor in the world coordinate frame, $\pi(\cdot)$ is the camera projection function, and $\rho(\cdot)$ is the Huber robust kernel function. Since the anchor coordinates have been optimized by the global factor graph, their spatial positions are geographically consistent, and the reduced-dimensional descriptors remain stable across sessions. Even if some anchors are not visible due to viewpoint changes or occlusion, reliable pose solving is still supported by the remaining matched points.

After initial pose recovery has been completed, a switch is made to visual-inertial odometry mode for continuous tracking so that inter-frame stability of augmented reality content is maintained. A sliding-window tightly coupled framework is adopted in this tracking stage, and the state vector is composed of accelerometer and gyroscope biases of the inertial measurement unit, velocity, position, and orientation quaternion. When the network is completely disconnected, cloud and edge nodes are no longer relied upon; instead, short-term pose propagation is performed using the locally cached anchor index and inertial measurement unit pre-integration. The recursion relation is given by:

$$T_{t+\Delta t} = T_t \text{Exp}(\Delta \xi_t) \quad (5)$$

where, T_t and $T_{t+\Delta t}$ are the poses at time t and $t + \Delta t$, respectively; $\Delta \xi_t$ is the relative pose increment estimated from inertial measurement unit pre-integration within the sliding window, containing rotation and translation components; and $\text{Exp}(\cdot)$ denotes the exponential mapping from the Lie algebra to the special Euclidean group. At the same time, local loop closure detection is performed using cached historical keyframes, by which accumulated drift is constrained. During network disconnection, augmented reality content display is driven entirely by local tracking, and intangible cultural heritage digital content around registered anchors can still be browsed and interacted with by users. Under experimental conditions, augmented reality content can be continuously displayed for more than 15 minutes.

After network connectivity is restored, incremental synchronization is performed between terminals and edge nodes so that data discrepancies generated during disconnection are eliminated. Newly observed anchor descriptor statistics, newly generated user-generated intangible cultural heritage micro-narratives, and local motion trajectories are uploaded by the terminals, whereas an updated version of the anchor index and new content submitted by other users in the same region are returned by the edge nodes. To handle conflicts caused by concurrent updates, a vector-clock-based version control mechanism is adopted. A version number is carried by each anchor and each content item, and during merging, timestamps

and user priorities are used as arbitration criteria so that consistency of the global index and content states is ensured in multi-terminal environments. Only incremental data is transmitted in this synchronization process, communication overhead is low, and completion can be achieved within seconds after network restoration so that seamless continuity of the cross-session experience is ensured.

5 DEVICE-EDGE-CLOUD COLLABORATIVE UPDATING AND DISTRIBUTION OF LIVING INTANGIBLE CULTURAL HERITAGE CONTENT

The living transmission of intangible cultural heritage is critically dependent on enabling users to contribute content in real-world scenes and on ensuring that such content is precisely anchored to corresponding spatial locations. When any semantic anchor A_i is approached by a user, a content creation entry is opened by the system. Short videos, audio, or text recorded by the user are automatically associated with that anchor, and metadata including user position p_u , timestamp t , user identifier u_{id} , and contextual description are appended. Because a semantic label and globally consistent coordinates are possessed by the anchor itself, micro-narratives generated by different users at different times are enabled to converge onto the same physical component so that a continuously accumulating digital cultural layer is formed. After content submission, lightweight quality assessment is immediately performed by the edge node. The assessment module is pretrained in the cloud and then compressed via knowledge distillation to an edge-deployable scale. Text content is directly fed into the text branch; audio content is first transcribed into text and then fed into the same branch. In the text branch, semantic relevance is discriminated by a student network. In the video branch, topic features are extracted by temporal three-dimensional convolution, and a relevance score $q \in [0,1]$ between the content and the intangible cultural heritage theme is output. Meanwhile, perceptual hash fingerprints are extracted by the edge node for duplicate detection, and completely identical or highly similar content is discarded. Only when q is not lower than a preset threshold q_{th} is the micro-narrative published and written into the content pointer set C_i of the anchor. By this design, content review is sunk from the cloud to the edge so that rapid response is enabled even under weak-network conditions, and the inference latency per content item is maintained below 50 ms.

After content has been published, candidate micro-narratives are required to be ranked according to user context so that the precision of intangible cultural heritage content delivery and user engagement are improved. A user interest model J_u is maintained in the cloud, and the model is updated online through historical user interaction behavior. Semantic relevance between content and anchors, anchor popularity, user interest similarity, and temporal freshness are jointly incorporated into the ranking function, which is formalized as:

$$S(u,c,a,t) = \alpha \cdot rel(c,a) + \beta \cdot pop(a) + \gamma \cdot sim(u,c) + \delta \cdot fresh(t) \quad (6)$$

where, u denotes user context, c denotes candidate content, a denotes the semantic anchor to which the content is associated, and t denotes the current time. $rel(c,a)$ measures the degree of matching between content and the semantic label of the anchor; $pop(a)$ represents recent visit popularity of the anchor; $sim(u,c)$ calculates cosine similarity between the user interest vector and the content feature vector; $fresh(t)$ is a temporal decay term; and α , β , γ , and δ are adjustable weight coefficients. This ranking is executed locally at the edge node so that raw user data is prevented

from being uploaded to the cloud. At the same time, low-latency personalized recommendation is achieved by leveraging local contextual information at the edge so that intangible cultural heritage content most relevant to the current spatial context is preferentially presented to users at different locations.

To ensure content reachability under weak-network and bandwidth-constrained conditions, popular content associated with anchors along the route is prefetched in advance by edge nodes according to user trajectory prediction, and device-to-device sharing is performed among multiple terminals in the same region via Wi-Fi Direct or Bluetooth. Maximization of cache utility is taken as the objective of the caching strategy, and the utility function is defined as:

$$U_{cache} = \sum_{a \in A_{cache}} w_a \cdot \lambda_a \cdot P(hit|u,a) \quad (7)$$

where, A_{cache} is the set of anchors covered by the current cache of the edge node; w_a is the average request frequency of content associated with anchor a ; λ_a is a transmission cost factor related to content size; and $P(hit|u,a)$ denotes the probability that a content request to anchor a is initiated given user context u . This probability is provided by a user trajectory prediction model. Trajectory prediction is performed using a lightweight Markov model or a long short-term memory (LSTM) network, which is executed locally at the edge node. Caching priorities are dynamically adjusted through online learning so that cached content remains spatially consistent with the region to be visited by the user, and thus the hit rate is improved and cellular backhaul traffic is reduced. Accordingly, semantic anchors serve not only as spatial registration primitives but also as a unified index for the organization, review, recommendation, and distribution of intangible cultural heritage content so that a complete closed loop for living transmission is formed under the device-edge-cloud architecture.

6 EXPERIMENTS AND PERFORMANCE EVALUATION

6.1 Experimental setup

Experiments were conducted in linear regions of approximately 2 km along the Grand Canal from Dongguan Street to the Ancient Canal in Yangzhou and approximately 1.5 km in the Gongchen Bridge section in Hangzhou. Six edge nodes were deployed along the route, all employing the NVIDIA Jetson Xavier NX platform with 8 GB of memory. Twenty mobile terminals were included, of which 12 were mid-range Android models and 8 were iPhone 12 or later models. The self-collected dataset comprised 5,000 keyframes, 300 semantic anchors, and 200 user-generated intangible cultural heritage micro-narratives. The comparison baselines included Google Cloud Anchors, ARKit World Map, a pure global navigation satellite system overlay, and an augmented reality system based on pre-built three-dimensional reconstruction. Evaluation metrics included spatial registration error, first recovery latency, duration of operation under network disconnection, content quality F1 score, cache hit rate, cellular traffic, device-side frame rate, and whole-device power consumption. All experiments were repeated five times, and means and standard deviations were reported. Statistical significance was assessed using a two-tailed t-test, and the significance level was set to 0.05.

6.2 Experiment 1: Semantic anchor registration accuracy and stability

The spatial consistency of the semantic anchor method under different illumination, weather, and viewpoint conditions was validated in this experiment. Repeated registration was performed for 50 preselected semantic anchors during three time periods along the canal, including clear-sky noon, overcast afternoon, and low-illuminance nighttime. Ground-truth coordinates were acquired using Real-Time Kinematic Global Positioning System (RTK-GPS) with a horizontal accuracy of ± 2 cm and a total station. The spatial registration errors and repeated localization success rates of the different methods are summarized in Table 1.

Table 1. Comparison of spatial registration error and repeated localization success rate among different methods

Method	Clear-Sky Error (m)	Overcast Error (m)	Nighttime Error (m)	Repeated Localization Success Rate (%)
Semantic anchors	0.21 ± 0.08	0.25 ± 0.10	0.31 ± 0.12	98.2 ± 1.5
Google Cloud Anchors	0.35 ± 0.15	0.42 ± 0.18	0.55 ± 0.22	91.6 ± 3.2
ARKit World Map	0.28 ± 0.12	0.33 ± 0.14	0.47 ± 0.19	93.4 ± 2.8
Pure global navigation satellite system overlay	1.82 ± 0.95	2.10 ± 1.20	2.45 ± 1.35	64.3 ± 8.7
Pre-built three-dimensional reconstruction augmented reality	0.18 ± 0.07	0.20 ± 0.08	0.24 ± 0.09	96.8 ± 1.9

As shown in Table 1, superior spatial registration accuracy under all three illumination conditions was achieved by the semantic anchor method relative to Google Cloud Anchors and ARKit World Map. Under clear-sky conditions, a median error of 0.21 m was obtained, corresponding to a reduction of approximately 38% relative to the 0.35 m error of Google Cloud Anchors. Under low-illuminance nighttime conditions, an error of 0.31 m was obtained, representing a reduction of approximately 44% relative to the 0.55 m error of Google Cloud Anchors, and the difference was found to be statistically significant. A repeated localization success rate of 98.2% was still maintained under nighttime low illuminance, whereas a decrease to 91.6% was observed for Google Cloud Anchors, indicating that stronger robustness to illumination variation is exhibited by the semantic descriptors. Slightly higher accuracy was achieved by the pre-built three-dimensional reconstruction augmented reality method; however, its reliance on a dense map constructed by offline laser scanning results in storage and computation overheads far higher than those of the semantic anchor scheme. Errors exceeding 1.8 m were produced by pure global navigation satellite system overlay, and fine-grained alignment of intangible cultural heritage content with specific components cannot be supported. While sub-decimeter-level accuracy is maintained by the semantic anchors, only 256 bytes are required for single-anchor storage, which is reduced by approximately two orders of magnitude relative to dense point cloud schemes, thereby validating the effectiveness of the lightweight spatial index design.

6.3 Experiment 2: Cross-session recovery performance and weak-network adaptability

Typical network fluctuation conditions along the canal were simulated in this experiment. Three bandwidth levels were set: complete disconnection at 0 Mbps,

weak network at 1 Mbps, and normal network at 10 Mbps. The first recovery latency, recovery success rate, and duration of augmented reality content display under network disconnection were measured after a 30-minute absence from the scene and subsequent re-entry. The main results are presented in Table 2.

Table 2. Comparison of cross-session recovery performance under different network conditions

Method	Recovery Latency Under Disconnection (s)	Success Rate Under Disconnection (%)	Duration Under Disconnection (min)	Recovery Latency Under Weak Network (s)	Recovery Latency Under Normal Network (s)
Semantic anchors	2.8 ± 0.4	97.6 ± 1.8	16.2 ± 2.3	2.1 ± 0.3	1.5 ± 0.2
Google Cloud Anchors	Failure	0	0	4.8 ± 0.9	1.2 ± 0.2
ARKit World Map	3.5 ± 0.7	88.4 ± 4.5	4.5 ± 1.2	2.6 ± 0.5	1.4 ± 0.2
Pure global navigation satellite system overlay	1.2 ± 0.3	100	–	1.1 ± 0.3	1.0 ± 0.2
Pre-built three-dimensional reconstruction augmented reality	Failure	0	0	4.2 ± 0.8	1.3 ± 0.2

The essential differences among the methods under network disconnection conditions are directly revealed by the results in Table 2. For the semantic anchor method, recovery can still be completed within 2.8 seconds under complete network disconnection, a success rate of 97.6% is achieved, and augmented reality content can be continuously displayed for 16.2 minutes. Google Cloud Anchors and pre-built three-dimensional reconstruction augmented reality become completely ineffective after network disconnection, because cloud anchor services and local dense maps are respectively relied upon by the two methods; a network connection cannot be established in the former, and matchable spatial primitives are lacking in the latter after the pre-scanned region is left. Although recovery can be achieved by ARKit World Map relying on a local world map, the duration under network disconnection is only 4.5 minutes, and the success rate is reduced to 88.4% due to accumulated drift. The fastest recovery is achieved by pure global navigation satellite system overlay, but its registration error has been confirmed to be meter-level in Table 1, and fine-grained overlay of intangible cultural heritage content cannot be satisfied. Under weak-network conditions, a recovery latency of 2.1 seconds is achieved by the semantic anchor method, which is superior to the 4.8 seconds of Google Cloud Anchors and the 4.2 seconds of the pre-built reconstruction method. The reason is that only hundreds of bytes are required to be transmitted for the anchor index, whereas local point cloud data must be downloaded for cloud anchors. These results indicate that, by the compact persistent design of semantic anchors, spatial recovery capability is no longer dependent on network availability, which is crucial for practical deployment in linear heritage scenarios.

6.4 Experiment 3: Quality assessment and distribution efficiency of living content

This experiment was composed of two parts: performance testing of the quality assessment model and effectiveness testing of the recommendation model. First, 30 intangible cultural heritage enthusiasts were invited to generate 300 micro-narratives, consisting of 100 short videos, 100 audio clips, and 100 text entries. These were annotated as relevant or not by three intangible cultural heritage experts.

The comparison results between the edge-side lightweight quality assessment model and the cloud-based large model and rule-based baseline are reported in Figure 2.

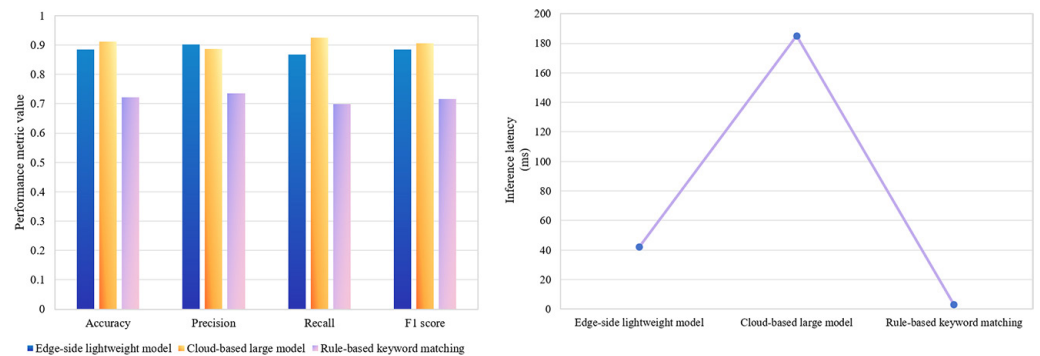


Fig. 2. Performance comparison of quality assessment models for intangible cultural heritage micro-narratives

As can be seen from Figure 2, an F1 score of 0.884 is achieved by the edge-side lightweight model, which is slightly lower than the 0.905 achieved by the cloud-based large model, whereas the inference latency is reduced from 185 ms to 42 ms, corresponding to a decrease of 77%. It is noteworthy that a precision of 0.901 is attained by the edge model, which is higher than the 0.887 attained by the cloud model, indicating that better performance in suppressing false positives is exhibited by the distilled lightweight network. This is particularly critical for filtering low-quality or irrelevant content in intangible cultural heritage scenarios. An F1 score of only 0.716 is achieved by rule-based keyword matching, demonstrating that semantic associations in intangible cultural heritage micro-narratives cannot be captured by shallow rules. When both accuracy and latency are considered, an optimal balance between utility and overhead is provided by the edge-side lightweight model in the device-edge collaborative architecture so that quality review can be completed in real time at edge nodes without reliance on cloud round trips.

The effectiveness of the recommendation model was evaluated through a two-week A/B test. Users were randomly assigned to three groups and were presented with random recommendation, popularity-based recommendation, and the proposed context-aware recommendation, respectively. User interaction rate, average dwell time, and satisfaction score are presented in Figure 3.

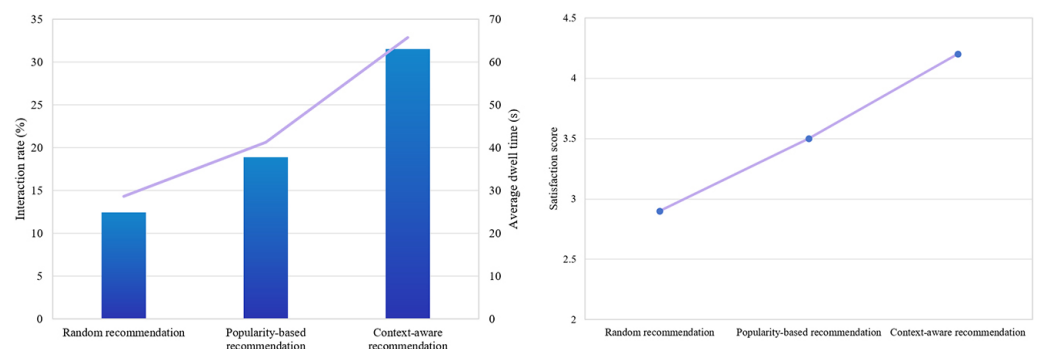


Fig. 3. User engagement metrics under different recommendation strategies

As shown in Figure 3, an interaction rate of 31.5% was achieved by the context-aware recommendation, corresponding to an increase of 66.7% over the popularity-based recommendation and 154% over the random recommendation, and the

magnitude of improvement was pronounced. Average dwell time was increased from 41.2 seconds under the popularity-based recommendation to 65.8 seconds, an increase of nearly 60%, indicating that the recommended micro-narratives were highly consistent with the user's current spatial context and interest preferences and that more time was willingly invested by users in browsing and interaction. A satisfaction score of 4.2 was achieved, approaching the satisfactory level, whereas only 2.9 was achieved by the random recommendation. The synergistic effectiveness of the four types of signals in the ranking function—anchor semantic relevance, anchor popularity, user interest similarity, and temporal freshness—was validated by these data, demonstrating that the precision of intangible cultural heritage content delivery and the depth of user participation can be substantially improved by the contextual metadata carried by semantic anchors.

6.5 Experiment 4: Device-Edge collaborative caching and network overhead optimization

This experiment was continuously run for one week in a real weak-network environment with an average downlink bandwidth of 1.2 Mbps. Network performance under three strategies—no caching, traditional least recently used caching, and anchor-based dynamic caching—was compared. Key metrics are summarized in Table 3.

Table 3. Comparison of network performance under different caching strategies

Caching Strategy	Cache Hit Rate (%)	Cellular Traffic (GB/week)	End-to-End Latency (s)	Device-to-Device Sharing Ratio (%)
No caching	0	48.6	4.8 ± 1.2	0
Least recently used caching	62.3 ± 4.5	22.1	2.9 ± 0.7	0
Anchor-based dynamic caching	84.7 ± 3.2	12.4	1.6 ± 0.4	21.3 ± 2.6

As shown by the results in Table 3, a cache hit rate of 84.7% was achieved by anchor-based dynamic caching, which is 22.4% higher than that of least recently used caching. This improvement is attributed to two mechanisms. Proactive prefetching of anchor content in regions to be visited by users is enabled by trajectory prediction, rather than passive waiting for requests. Additionally, cache replacement decisions are made more spatially targeted by the association between anchor popularity and spatial location. Cellular traffic was reduced from 48.6 GB/week under no caching to 12.4 GB/week, a decrease of 74.5%, greatly alleviating back-haul link pressure in weak-network environments. End-to-end content acquisition latency was reduced from 4.8 seconds to 1.6 seconds so that almost instantaneous responses are provided after intangible cultural heritage content is clicked by users, and the smoothness of the augmented reality interaction experience is thus significantly improved. A device-to-device sharing ratio of 21.3% was achieved, indicating that duplicate downloads are effectively reduced by content sharing among users in the same region via Wi-Fi Direct or Bluetooth. As a unified content indexing unit, semantic anchors enable edge caching and inter-terminal sharing to be collaboratively organized rather than operating independently, which is the structural reason for the improved cache efficiency.

7 CONCLUSION

To address the core problems of insufficient spatial registration accuracy, poor weak-network adaptability, and static intangible cultural heritage content faced by mobile augmented reality in linear heritage scenarios along the Grand Canal, a device-edge-cloud collaborative method with semantic anchors as a unified spatial index was proposed. Semantic anchors were generated online through joint factor graph optimization of lightweight visual semantic features and global navigation satellite system/inertial measurement unit observations, and compact storage at the hundreds-of-bytes level was adopted so that decimeter-level registration accuracy was maintained while dependence on dense maps and continuous network connectivity was eliminated. The anchors simultaneously served as persistent primitives for cross-session pose recovery and spatial containers for user-generated intangible cultural heritage micro-narratives so that spatial registration, weak-network recovery, and living content accumulation formed a closed loop around the same lightweight structure. It is shown by prototype experiments in real canal sections that a recovery success rate of 97.6% under complete network disconnection, an augmented reality display duration exceeding 16 minutes, a cache hit rate of 84.7%, an average mobile-side frame rate of 28.6 fps, and whole-device power consumption of 2.8 W were achieved by the proposed method, and its engineering feasibility and practical value in large-scale weak-network scenarios were thereby validated.

8 ACKNOWLEDGEMENT

This study was supported by the Research Project of Humanities and Social Sciences in Colleges and Universities of Hebei Province (Grant No.: QN2025809).

9 REFERENCES

- [1] S. Galani and S. Vosinakis, “An augmented reality approach for communicating intangible and architectural heritage through digital characters and scale models,” *Personal and Ubiquitous Computing*, vol. 28, pp. 471–490, 2024. <https://doi.org/10.1007/s00779-024-01792-x>
- [2] F. Luo, “Application of mobile augmented reality technology in urban tourism and optimization of visitor experience,” *International Journal of Interactive Mobile Technologies*, vol. 20, no. 11, pp. 138–151, 2026. <https://doi.org/10.3991/ijim.v20i11.62124>
- [3] I. Kyianytsia, D. Yatsiuk, H. Aldankova, O. Horobets, V. Slipchenko, and V. Dobrovolskyi, “A cyber-physical system approach to adaptive visual branding in cultural institutions,” *Journal of Engineering Management and Systems Engineering*, vol. 5, no. 3, pp. 340–353, 2026. <https://doi.org/10.56578/jemse050304>
- [4] S. Zhang, X. Zhang, and J. Liu, “Spatial distribution and pedigree age of intangible cultural heritage along the Grand Canal of China,” *Heritage Science*, vol. 12, Art. no. 246, 2024. <https://doi.org/10.1186/s40494-024-01357-4>
- [5] Z. Huang and V. Friderikos, “Optimal service decomposition for mobile augmented reality with edge cloud support,” *Computer Communications*, vol. 202, pp. 97–109, 2023. <https://doi.org/10.1016/j.comcom.2023.02.002>
- [6] B. Hannewijk, F. L. Vinella, V. Khan, I. Lykourantzou, K. Papangelis, and J. Masthoff, “Capturing the city’s heritage on-the-go: Design requirements for mobile crowdsourced cultural heritage,” *Sustainability*, vol. 12, no. 6, Art. no. 2429, 2020. <https://doi.org/10.3390/su12062429>

- [7] M. Chen, W. Liu, T. Wang, A. Liu, and Z. Zeng, "Edge intelligence computing for mobile augmented reality with deep reinforcement learning approach," *Computer Networks*, vol. 195, Art. no. 108186, 2021. <https://doi.org/10.1016/j.comnet.2021.108186>
- [8] A. Vidal-Balea, O. Blanco-Novoa, P. Fraga-Lamas, and T. M. Fernández-Caramés, "Off-cloud anchor sharing framework for multi-user and multi-platform mixed reality applications," *Applied Sciences*, vol. 15, no. 13, Art. no. 6959, 2025. <https://doi.org/10.3390/app15136959>
- [9] C. Toft et al., "Long-term visual localization revisited," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 4, pp. 2074–2088, 2020. <https://doi.org/10.1109/TPAMI.2020.3032010>
- [10] Y. Qi, C. Liu, H. Wu, X. Shao, and J. Wei, "A multi-layer fusion image enhancement method for visual odometry under poor visibility scenarios," *International Journal of Applied Earth Observation and Geoinformation*, vol. 115, Art. no. 103118, 2022. <https://doi.org/10.1016/j.jag.2022.103118>
- [11] Y. Wang, L. Chen, P. Wei, and X. Lu, "Visual-inertial odometry of smartphone under manhattan world," *Remote Sensing*, vol. 12, no. 22, Art. no. 3818, 2020. <https://doi.org/10.3390/rs12223818>
- [12] B. Bahadur, "A study on the real-time code-based GNSS positioning with Android smartphones," *Measurement*, vol. 194, Art. no. 111078, 2022. <https://doi.org/10.1016/j.measurement.2022.111078>
- [13] Y. Hou, S. Kenderdine, D. Picca, M. Egloff, and A. Adamou, "Digitizing intangible cultural heritage embodied: State of the art," *Journal on Computing and Cultural Heritage*, vol. 15, no. 3, Art. no. 55, 2022. <https://doi.org/10.1145/3494837>
- [14] G. Koutromanos, D. Koukopoulos, Z. Koukopoulos, and C. Mouzakis, "Cultural heritage content development and dissemination through a participatory platform: Lessons learned from in-service teachers' perception," *Education and Information Technologies*, vol. 28, pp. 3513–3536, 2022. <https://doi.org/10.1007/s10639-022-11336-6>
- [15] E. Kaldeli, O. Menis-Mastromichalakis, S. Bekiaris, M. Ralli, V. Tzouvaras, and G. Stamou, "CrowdHeritage: Crowdsourcing for improving the quality of cultural heritage meta-data," *Information*, vol. 12, no. 2, Art. no. 64, 2021. <https://doi.org/10.3390/info12020064>

10 AUTHORS

Xi Zhao received her bachelor's degree in Fashion Design and Technology from Hebei Normal University in 2012 and her master's degree in Fashion Design from the Accademia di Belle Arti di Napoli, Italy, in 2014. She is currently a lecturer and director of the Fashion and Apparel Design Teaching and Research Section at the Department of Fashion Technology, Hebei Art and Design Academy. Her research interests include fashion design, intangible cultural heritage, and digital fashion (E-mail: zhaoxigymy@126.com).

Zhao Zhang received her bachelor's degree in Fashion Design and Engineering from the College of Arts and Information Engineering, Dalian Polytechnic University. She is currently a lecturer in the Department of Fashion Technology at Hebei Art and Design Academy and holds a Level II vocational qualification as a garment pattern maker. Her research interests include digital fashion design and pattern-making technology (E-mail: Zzzn1112@163.com).

Aimin Li received his bachelor's degree from Tangshan Normal University in 2012 and his master's degree from the Accademia di Belle Arti di Napoli, Italy, in 2014. He currently works at the Student Affairs Office, Tangshan Normal University. His research focuses on cultural heritage conservation, particularly the protection and development of intangible cultural heritage along the Grand Canal (E-mail: liaimintnu@163.com).