Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 20 No. 7 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijim.v20i07.61243

PAPER

Predicting and Analyzing User Engagement and Economic Impacts of Mobile Interactive Navigation in Smart Tourism Environments

Lipeng Wang(✉)

Shijiazhuang University of Applied Technology, Shijiazhuang, China

2003100270@sjzpt.edu.cn

ABSTRACT

The advancement of smart tourism has shifted mobile navigation systems from route guidance tools toward experience-oriented enablers. However, conventional static navigation systems, constrained by limited interaction modes and weak contextual adaptability, fail to support immersive engagement or effectively stimulate tourism consumption. This study examines how mobile interactive navigation enhances user engagement in smart tourism scenarios and quantifies its economic impacts through an integrated "technological innovation–engagement measurement–economic impact" framework. A unified design paradigm integrating context awareness, multimodal interaction, and intelligent tour guidance is proposed. Based on this paradigm, a mobile navigation system incorporating multi-source fused positioning, lightweight dynamic AR rendering, adaptive path planning, and multi-channel interaction is developed, enabling high-precision localization and immersive experiences in complex scenic environments. A three-dimensional user engagement measurement framework covering behavioral, emotional, and cognitive dimensions is constructed, together with a time-series Transformer-based prediction model to capture dynamic engagement trends and inflection points. Furthermore, a structural equation model (SEM) combined with a multi-agent simulation system is employed to analyze the driving mechanisms through which engagement influences tourism consumption and regional economic spillover effects. These results confirm the effectiveness of mobile interactive navigation in enhancing user engagement and scenic-area economic performance.

KEYWORDS

smart tourism, mobile interactive navigation, user engagement, multimodal fusion, time-series transformer, economic impact prediction

1 INTRODUCTION

With the deepening of digital transformation in smart tourism [1, 2], mobile navigation has become a core carrier linking tourists with scenic-area services [3, 4].

Wang, L. (2026). Predicting and Analyzing User Engagement and Economic Impacts of Mobile Interactive Navigation in Smart Tourism Environments. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(7), pp. 153–167. https://doi.org/10.3991/ijim.v20i07.61243

Article submitted 2025-12-10. Revision uploaded 2026-01-26. Final acceptance 2026-02-26.

Traditional mobile navigation mainly focuses on static route guidance functions [5, 6], lacks adaptability to complex scenic environments, and finds it difficult to meet tourists' core demands for immersive experiences, personalized services, and contextualized information [7], resulting in fragmented experiences and insufficient engagement momentum. As a key hub connecting technological experience and consumption decision-making [8], user engagement can effectively extend tourists' length of stay, expand consumption scenarios, and strengthen revisit intentions when enhanced [9]. It not only activates new drivers of economic growth for scenic areas but also provides precise guidance for the optimization of tourism service supply, thus possessing both technical practical value and industrial application significance.

Focusing on the core pain points and research gaps of mobile interactive navigation in smart tourism environments, this study defines three key research questions. First, how to construct a multimodal fusion-based mobile interactive navigation technology system [10] to realize a paradigm shift from route guidance to experience empowerment and to enhance user engagement through technological innovation. Second, how to establish an objective and dynamic multidimensional user engagement measurement system, overcoming the subjectivity and fragmentation limitations of traditional measurement approaches [11], and achieving precise quantification of engagement. Third, how to quantify the direct and indirect driving effects of user engagement on tourism consumption and to construct a scientific economic impact prediction model to provide data support for scenic-area business decision-making [12, 13].

Taking technological innovation–engagement measurement–economic impact as the core research thread, this study constructs a three-in-one research framework and forms three major innovations. At the technological level, a context-aware, multimodal interaction intelligent tour guidance integrated system architecture is proposed, integrating key technologies such as multi-source positioning, lightweight augmented reality, and adaptive path planning. This approach addresses the problems of single interaction modes and poor environmental adaptability in traditional navigation and enables high-precision positioning and immersive experience empowerment in complex scenic environments. At the measurement level, a behavioral–emotional–cognitive three-dimensional engagement indicator system is established, and a time-series Transformer-based prediction model is designed. Compared with traditional time-series models, it is better able to capture long-term dependencies in user behavior and accurately predict dynamic engagement trends and key inflection points.

The subsequent sections of this paper are organized according to a logical progression. Section 2 focuses on core technological innovations and elaborates on the design concept, overall architecture, and key technical modules of the mobile interactive navigation system, implementing the proposed technological solutions. Section 3 constructs a multidimensional engagement measurement system and designs a time-series Transformer prediction model to achieve precise quantification and dynamic prediction of engagement. Section 4 establishes an economic impact analysis framework and quantifies the driving effects of engagement on tourism consumption through a dual-model approach. Section 5 conducts empirical experiments in a representative scenic area to verify the technical feasibility, engagement enhancement effects, and economic impact mechanisms.

2 ARCHITECTURE DESIGN OF THE MOBILE INTERACTIVE NAVIGATION SYSTEM

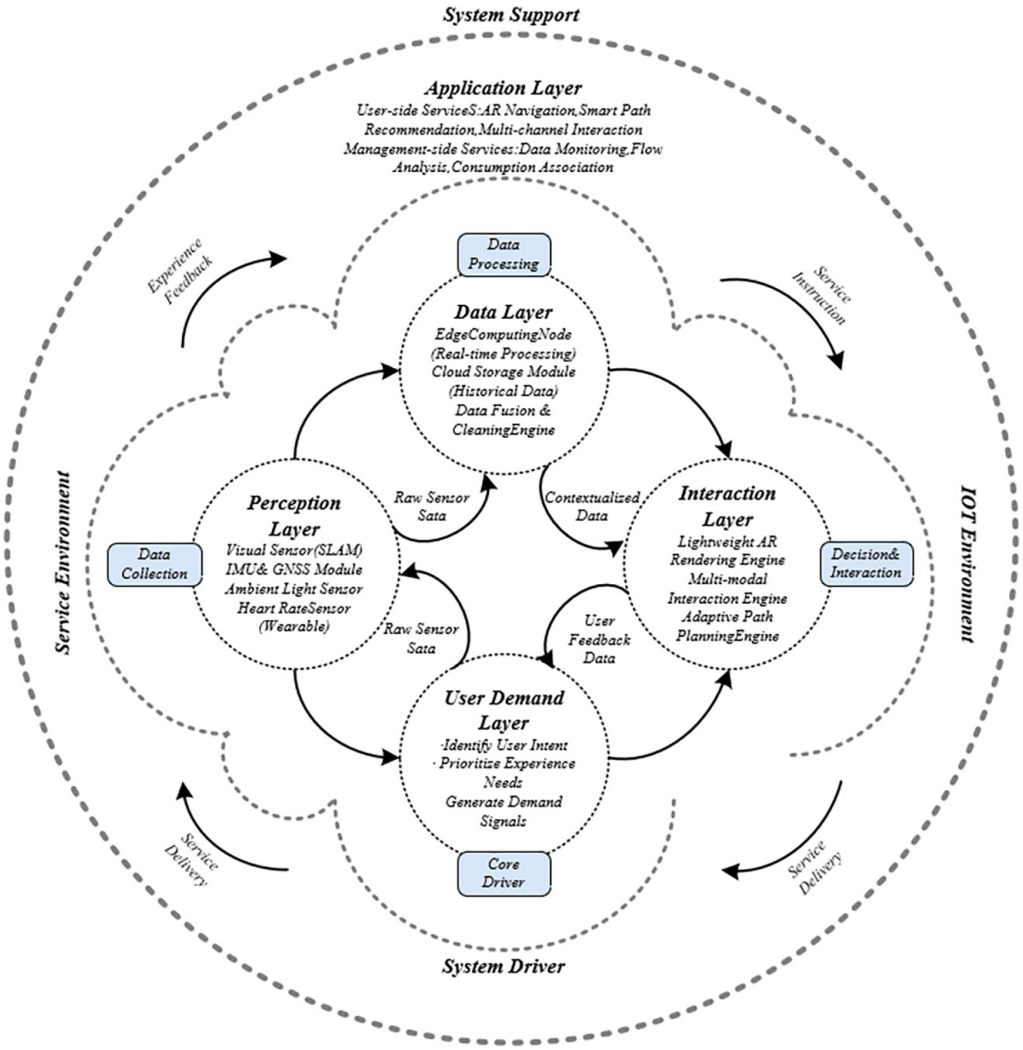

Fig. 1. Overall architecture of the context-aware–multimodal interaction–intelligent tour guidance integrated system

This system takes the integration of context awareness—multimodal interaction—intelligent tour guidance as the core design concept, breaks through the limitations of traditional navigation based on passive route guidance, and constructs a closed-loop technical system from perception to service, realizing a paradigm upgrade toward active experience empowerment. Its core logic lies in accurately capturing user–environment dynamics through multi-level technical collaboration and maximizing immersive experience and engagement intention via multidimensional interaction and intelligent decision-making. Figure 1 presents the system architecture.

The system adopts a four-layer hierarchical architecture, in which each layer collaborates deeply through high-bandwidth and low-latency data links, balancing positioning accuracy, energy consumption, and interaction efficiency. The perception layer innovatively integrates visual sensors, inertial measurement units (IMU), global navigation satellite systems (GNSS), ambient light sensors, and physiological sensors from interconnected smart wearable devices. A multi-source fused positioning and

state perception system is constructed for complex indoor and outdoor scenic environments. An extended Kalman filter algorithm is employed to optimize positioning accuracy, and the core position estimation formula is expressed as:

$$\hat{X}_k = \Phi_k \hat{X}_{k-1} + K_k (Z_k - \mathrm{H}_k \Phi_k \hat{X}_{k-1}) \quad (1)$$

where $\hat{X}_k$ represents the estimated user position and motion state at time k, Φ_k denotes the state transition matrix, K_k is the adaptive Kalman gain, Z_k is the multi-sensor fused observation, and H_k is the observation matrix. By dynamically adjusting the weight of K_k to allocate the contribution of visual simultaneous localization and mapping (SLAM) feature points, IMU inertial data, and GNSS signals, the positioning error is controlled within 0.5 meters. Meanwhile, ambient light intensity, user heart rate, and other signals are collected in real time to provide data support for subsequent interaction and tour guidance optimization. The data layer adopts an edge–cloud collaborative distributed storage and processing architecture and innovatively integrates a streaming data real-time processing unit and an adaptive caching module. Through feature alignment algorithms, temporal synchronization and noise filtering of multi-source heterogeneous data are achieved. Storage resources are allocated based on data importance grading. Non-critical historical data are stored using compression strategies to reduce mobile-side energy consumption, while core data are preprocessed at edge nodes and uploaded to the cloud, ensuring data processing latency below 100 ms and meeting real-time interaction requirements. The interaction layer integrates a self-developed multimodal interaction engine, a GPU-optimized AR rendering engine, and an intelligent decision engine. The AR rendering engine adopts a tiled parallel rendering strategy, dynamically allocating resources based on mobile GPU computing power, prioritizing the rendering of regions within the user's visual focus. Rendering parameters are adjusted in real time through an ambient light feedback model, reducing power consumption by 30% compared with traditional solutions. The intelligent decision engine embeds multi-objective optimization logic to provide algorithmic support for subsequent path planning and interaction responses. The application layer realizes the transformation of technical capabilities into service scenarios, providing users with services such as personalized AR tour guidance and multi-channel interaction. At the same time, standardized data interfaces are opened to scenic-area management platforms, outputting passenger flow distribution, user interaction hotspots, and consumption-related data, forming bidirectional empowerment for user experience optimization and scenic-area operation decision-making, and fully supporting the implementation of the integrated design concept.

The multimodal fusion positioning and context awareness module focuses on positioning challenges in complex indoor and outdoor scenic environments and innovatively proposes a multi-source collaborative positioning scheme integrating visual SLAM, IMU, and GNSS, achieving deep integration of high-precision seamless positioning and intelligent scene perception. The core innovation of this module lies in the lightweight optimization of mobile-side SLAM algorithms. An ORB feature point sparse extraction strategy is adopted, in which key feature points with high discriminability, such as building contours and landmarks, are selected by setting grayscale gradient thresholds and feature point spacing constraints. Redundant information is removed, reducing the number of feature points by 40%. Meanwhile, an inter-frame matching optimization mechanism based on the RANSAC algorithm is introduced to suppress matching errors caused by dynamic obstacles and

illumination variations. Combined with IMU inertial data, smooth correction of positioning results is achieved. The core positioning accuracy formula is expressed as:

$$\sigma = \sqrt{\omega_1 \sigma_{SLAM}^2 + \omega_2 \sigma_{IMU}^2 + \omega_3 \sigma_{GNSS}^2} \tag{2}$$

where ω_1, ω_2, and ω_3 are adaptive weight coefficients that are dynamically adjusted according to environmental complexity. When GNSS signals are strong in outdoor environments, the proportion of ω_3 is increased to 0.6. Under indoor signal occlusion conditions, the system automatically switches to a combined visual SLAM and IMU mode, with ω_1 and ω_2 set to 0.55 and 0.45, respectively. As a result, the overall positioning error is stably controlled within 0.5 meters while balancing positioning accuracy and mobile-side energy consumption. At the context awareness level, the module integrates multi-source environmental data and positioning information to construct a lightweight scene classification model. By extracting ambient light intensity, passenger flow density, and regional feature vectors, real-time identification of scenarios such as cultural relic exhibition areas and catering areas is achieved, with a recognition accuracy exceeding 92%, providing precise contextual support for subsequent personalized service delivery and path planning.

The lightweight dynamic AR rendering and contextual information overlay module overcomes the bottlenecks of excessive power consumption and insufficient scene adaptability in traditional AR navigation and innovatively develops a mobile GPU-optimized dynamic rendering engine to achieve low-power and high-efficiency contextual information overlay. The engine adopts a layered and tiled parallel rendering architecture, decomposing AR scenes into foreground interactive elements, midground environmental overlay layers, and background scene reconstruction layers. Based on the CUDA core computing resource allocation mechanism of mobile GPUs, computing resources are preferentially scheduled to render regions within the user's visual focus with pixel-level precision. Non-focus regions reduce rendering precision through texture compression and simplified rendering pipelines. Combined with a GPU frequency dynamic adjustment strategy, power consumption is reduced by 32% compared with traditional single-path rendering solutions. The module incorporates a dynamic adaptive rendering mechanism, in which ambient light intensity parameters are collected in real time via ambient light sensors to establish a brightness adaptive adjustment function $L_{AR} = L_{env} \times \alpha + L_{base}$, where L_{AR} denotes AR content brightness, L_{env} denotes ambient light intensity, α is the adaptation coefficient, and L_{base} is the base brightness threshold. This enables dynamic adjustment with increased brightness under strong light conditions and reduced brightness under low-light conditions. Meanwhile, pupil feature points are captured through the front-facing camera, and the user's gaze direction is located via a gaze-tracking algorithm. AR information overlay coordinates are adjusted in real time to ensure natural alignment between virtual content and real scenes. To address transmission latency issues, the module adopts an edge computing architecture, deploying time-consuming tasks such as AR model preprocessing and texture compression on scenic-area edge nodes. Only lightweight rendering instructions and core data are transmitted through low-latency 5G links to achieve device–edge collaboration, controlling AR content synchronization latency within 85 ms, avoiding visual stuttering and spatial misalignment, and accurately overlaying cultural relic historical reconstruction scenes, shop information, and other content to enhance contextual experience.

The adaptive intelligent path planning and recommendation module breaks through the limitation of traditional path planning that prioritizes distance alone.

Taking experience value maximization as the core objective, a multidimensional dynamic optimization algorithm system is constructed to enable real-time generation and dynamic adjustment of personalized routes. The module innovatively establishes a three-layer decision indicator system, integrating three categories of core parameters: user profiles, real-time states, and environmental factors. A dynamic weight allocation model is designed, and the core experience value optimization function is expressed as $V=\sum_{i=1}^{n} \omega_i S_i$, where V denotes the comprehensive experience value of a path, ω_i represents the dynamic weight of each indicator, and S_i denotes the normalized score of each indicator. Weights are adaptively adjusted according to user characteristics. For elderly users, the weight of the physical exertion indicator $\omega_{physical}$ is increased to 0.35, while for younger users, the weight of the point-of-interest coverage indicator $\omega_{interest}$ is set to 0.32. At the same time, physical exertion values are estimated using motion acceleration and step frequency data collected by the IMU, and passenger flow density and congestion indices are updated in real time based on scenic-area Internet of Things data. The module adopts an improved A^* algorithm, introducing multi-objective constraint conditions to reconstruct the heuristic function. Path length, physical exertion, congestion level, and point-of-interest coverage are incorporated into the optimization dimensions, and a greedy strategy is applied to rapidly converge to the optimal solution. For sudden congestion scenarios in scenic areas, the algorithm optimizes the data processing pipeline to complete path replanning and delivery within 1 second. The personalized recommendation function is deeply integrated with the path planning process. Based on user interest tags and real-time location, a collaborative filtering algorithm is employed to generate adaptive recommendation lists, integrating shops and experiential projects with the highest matching degree into route nodes, achieving natural convergence between navigation and recommendation, guiding users to actively explore, and enhancing the completeness of the experience.

The multi-channel immersive interaction module breaks the traditional single touch-screen interaction mode and constructs a three-in-one multi-channel fusion interaction system integrating voice, gesture, and haptic feedback, enabling a more intuitive immersive operational experience. At the voice interaction level, a lightweight on-device Transformer-based speech recognition model is integrated. Mel-frequency cepstral coefficients are extracted as speech features, and noise suppression algorithms optimized for noisy scenic environments are applied. Spectral subtraction is used to reduce environmental noise interference. Meanwhile, a contextual semantic understanding library is constructed, and an attention mechanism is employed to capture dialog logic associations, enabling accurate responses in multi-turn conversations. The recognition accuracy exceeds 93%, supporting multi-dialect adaptation and offline voice interaction. Gesture recognition adopts a rear-camera visual capture solution, and a lightweight gesture detection model based on convolutional neural networks is innovatively developed. By extracting hand key point features, basic gestures such as waving for switching, fist clenching for zooming, and sliding for viewpoint adjustment are rapidly recognized. The model parameter size is compressed to within 8 MB, recognition latency is controlled within 50 ms, and accuracy exceeds 95%, adapting to contactless operation scenarios where tourists are holding items. The haptic feedback module utilizes mobile linear vibration motors to output multi-mode vibration patterns and designs a scenario-based haptic encoding mechanism. Direction guidance adopts unilateral differential vibration, point-of-interest reminders use pulse vibration, and AR narrative interactions employ rhythmic vibration. The core vibration intensity control formula is $I = k \times P$,

where I denotes vibration intensity, k is the scenario adaptation coefficient, and P is the user perception threshold. By dynamically adjusting the value of k, feedback effectiveness and interference are balanced, enabling coordinated complementarity among multi-channel interactions and enhancing operational immersion.

3 MULTIDIMENSIONAL MEASUREMENT AND PREDICTION MODEL OF USER ENGAGEMENT

A behavioral–emotional–cognitive three-dimensional user engagement measurement indicator system is constructed to overcome the subjectivity and fragmentation limitations of traditional measurement approaches and to achieve comprehensive and objective quantification of engagement. Behavioral engagement focuses on objectively collectible indicators, including average duration per single use, multimodal interaction frequency, path deviation rate, functional usage depth, and stay duration gain. Among them, the path deviation rate is calculated as the ratio of exploration duration to total duration, functional usage depth is represented by the proportion of actually used modules to the total number of system modules, and stay duration gain is defined as the difference in usage duration between this system and traditional navigation, accurately reflecting the intensity of users' active participation behaviors. Emotional engagement abandons subjective questionnaire-based methods and adopts objective quantification based on multi-source data. Emotional state recognition is performed by extracting intonation rhythm and speech rate features from voice interactions, combined synchronously with user-submitted emotional tags and heart rate variability data collected by smart wearable devices, to construct a quantified emotional fluctuation curve, enabling precise identification of three core emotional states: pleasure, neutrality, and irritability. Cognitive engagement is constructed based on task-related indicators, including AR treasure-hunting quiz accuracy, cultural knowledge collection and sharing rate, contextual information interaction rate, and active exploration frequency. Through user performance in interactive tasks, the level of cognitive investment in tourism content is quantified, forming a three-dimensional, synergistic, and progressively layered engagement measurement system.

A three-tier architecture of on-device collection–edge processing–cloud storage is adopted to achieve compliant and efficient processing and integration of multi-source engagement data. At the compliance level, layered authorization protocols are used to clarify the scope of data collection. User identity information and location data are encrypted and anonymized using the SHA-256 hashing algorithm, retaining only behavioral features and interaction data. The entire process complies with privacy protection regulations, ensuring that data collection is lawful and controllable. During the data fusion stage, abnormal data caused by mis-operations and sensor failures are removed based on the 3σ criterion. A dynamic time warping algorithm is applied to achieve temporal alignment of behavioral, emotional, and cognitive multi-source data, which are integrated into a unified dataset according to timestamps, while data standardization is performed to eliminate dimensional differences. To reduce subsequent model computational complexity, principal component analysis is employed for dimensionality reduction optimization. The core formula is $Y = XW$, where Y denotes the reduced feature matrix, X denotes the original high-dimensional data matrix, and W denotes the feature projection matrix. Core principal components are selected based on variance contribution rates, retaining feature dimensions with a cumulative variance contribution rate exceeding 85%.

While preserving key information, the data dimensionality is reduced by 60%, significantly improving model training efficiency.

A time-series Transformer-based user engagement prediction model is proposed to overcome the bottleneck of traditional time-series models in capturing long-term dependency relationships, enabling precise prediction of dynamic engagement trends and key inflection points. The model adopts an Encoder–Decoder architecture. The Encoder consists of six layers of multi-head attention mechanisms and feedforward neural networks, with each attention head dimension set to 64. Long-term dependencies in user behavior sequences are captured through a scaled dot-product attention mechanism. The core attention weight computation formula is expressed as:

$$Attention(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

where Q, K, and V denote the query, key, and value matrices, respectively, and d_k denotes the attention dimension, effectively enhancing the capture of correlations between interaction patterns and interest changes across different time periods. During the feature embedding stage, user profiles and contextual information are incorporated as auxiliary features and jointly embedded with behavioral time-series features. Sinusoidal positional encoding is applied to preserve temporal information, enhancing the model's ability to identify key inflection points such as fatigue points and interest activation points:

$$PE(pos,2i)=\sin(pos/10000^{2i/d_{model}}) \tag{4}$$

$$PE(pos,2i+1)=\cos(pos/10000^{2i/d_{model}}) \tag{5}$$

Model optimization adopts the AdaFactor adaptive learning rate algorithm, dynamically adjusting parameter update step sizes to address slow convergence caused by uneven time-series data distributions. Meanwhile, an early stopping strategy is triggered when the validation loss does not decrease for five consecutive epochs, preventing overfitting.

4 PREDICTION AND ANALYSIS MODEL OF ECONOMIC IMPACTS

A dual-dimensional framework of tourism economic impacts driven by user engagement is defined to achieve comprehensive quantification covering both micro-level immediate consumption and macro-level spillover effects, overcoming the single-dimensional limitations of traditional economic impact analysis. The direct consumption dimension focuses on immediate transactions generated under the guidance of the navigation system, covering core consumption categories such as catering, shopping, and experiential projects. Through methods including coupon QR code redemption records, online payment data association, and docking with merchant transaction system interfaces, a direct linkage between consumption behavior and navigation services is constructed, enabling precise traceability and statistical accounting of consumption amounts and frequencies. The indirect consumption and spillover effects dimension is divided into two core scenarios. Time spillover consumption originates from extended stay durations brought about by increased engagement, which in turn drives derivative consumption such

as accommodation, secondary park entry, and visits to surrounding attractions, and is quantified through correlation analysis between tourist stay duration and consumption categories. Social spillover effects rely on user AR content social sharing data, combined with dissemination path tracking and potential tourist conversion models, to quantify long-term consumption growth driven by word-of-mouth enhancement, forming a complete economic impact evaluation system of "direct consumption–indirect spillover."

A structural equation model (SEM) is constructed to quantify the engagement–consumption pathways and to achieve precise analysis of the driving effects of different engagement dimensions on tourism consumption, overcoming the subjectivity limitations of traditional path analysis. Figure 2 presents the SEM path diagram of user engagement-driven tourism consumption. The model innovatively designs a system of latent variables and observed variables. Behavioral, emotional, and cognitive engagement are set as exogenous latent variables, consumption intention is defined as an endogenous mediating latent variable, and direct consumption amount, consumption frequency, and stay duration are specified as endogenous outcome variables. Each latent variable corresponds to the aforementioned three-dimensional measurement indicators as observed variables, forming a transmission pathway of "engagement–consumption intention–consumption behavior." The core structural equation is expressed as $\eta = B\eta + \Gamma\xi + \zeta$, where η denotes the vector of endogenous latent variables, ξ denotes the vector of exogenous latent variables, B is the path coefficient matrix among endogenous latent variables, Γ is the path coefficient matrix from exogenous latent variables to endogenous latent variables, and ζ represents the residual term. Consumption intention data are collected through questionnaire surveys, combined with actual consumption data obtained through cooperation with scenic areas and merchants. Path coefficients are calculated using model fitting algorithms to verify emotional engagement as the core driving factor (with an expected path coefficient of 0.62). At the same time, age, consumption capacity, and travel type are introduced as moderating variables, and multi-group analysis is conducted to reveal path differences across different groups, enhancing the model's support for precise decision-making.

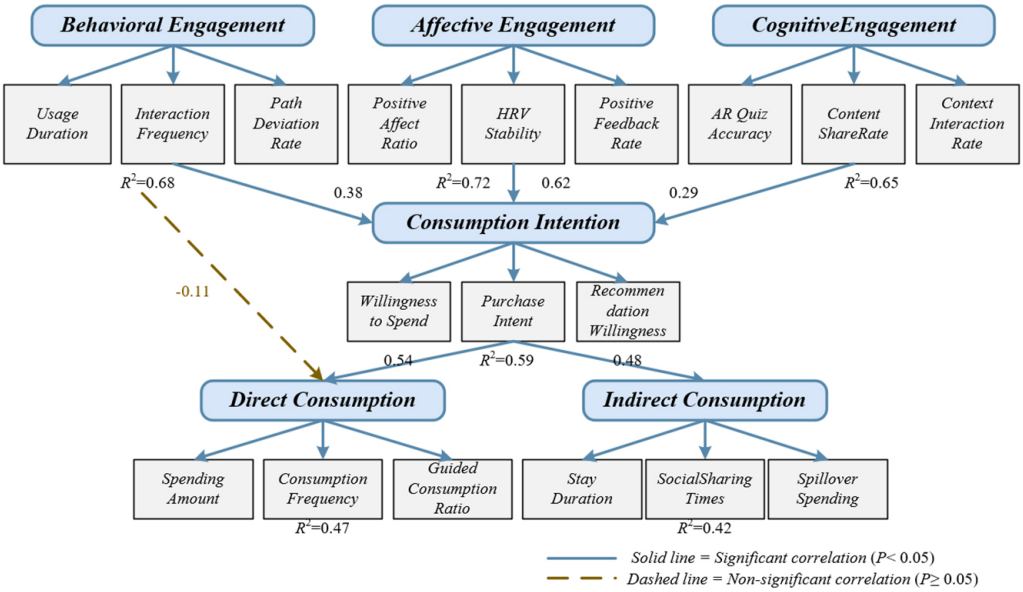

Fig. 2. SEM path diagram of user engagement-driven tourism consumption

Based on a multi-agent system, a scenic-area economic impact simulation model is constructed to achieve fine-grained and visualized prediction of regional economic effects, compensating for the overly coarse nature of traditional macroeconomic analysis. The model innovatively designs three types of agents. Tourist agents are endowed with core attributes such as age, consumption preferences, and engagement characteristics. Decision rules are trained based on the aforementioned engagement prediction results and empirical behavioral data, enabling dynamic adjustment of stay duration, consumption choices, and movement trajectories according to engagement levels. Merchant agents are grouped by types such as catering, shopping, and experiences and possess adaptive capabilities to adjust pricing and service supply in response to passenger flow density. Environmental agents simulate environmental factors such as passenger flow density and scene congestion status in real time, providing support for the behavior decision-making of the former two agent types. The core behavioral rule is expressed as:

$$P_{i,j} = \frac{exp(U_{i,j})}{\sum_{k=1}^{n} exp(U_{i,k})} \tag{6}$$

where $P_{i,j}$ denotes the probability that the tourist agent i selects the merchant j, and $U_{i,j}$ denotes the utility function, integrating variables such as engagement level, merchant type matching degree, and congestion level. By constructing a simulation system on a simulation platform, two comparative scenarios of innovative navigation and traditional navigation are established. The model simulates and predicts changes in indicators such as overall scenic-area revenue, merchant revenue distribution, and consumption concentration in popular areas, accurately capturing the impact of engagement enhancement on regional economic structure and providing data support for scenic-area business optimization and resource allocation.

5 EXPERIMENTAL DESIGN AND RESULTS ANALYSIS

5.1 Experimental analysis

To verify the technical feasibility, engagement enhancement effect, and economic impact mechanism of the proposed mobile interactive navigation system, targeted empirical experiments were conducted. A prototype system compatible with Android and iOS platforms was developed based on the Unity cross-platform framework, integrating the aforementioned core modules, including multimodal fusion positioning, lightweight AR rendering, adaptive path planning, and multi-channel interaction. The algorithm logic was implemented using the C language, and mobile sensor interfaces were invoked to complete multi-source data acquisition. The OpenCV library was employed to optimize the mobile adaptation performance of visual SLAM and gesture recognition algorithms, while the Vuforia engine was used to achieve efficient AR content rendering and scene fusion. Multiple rounds of testing ensured system stability and interaction fluency in complex scenic environments.

A typical historical and cultural block was selected as the experimental site. The area covers approximately 2 square kilometers and includes complex indoor-outdoor building environments and diverse types of commercial establishments, comprising 12 cultural heritage attractions and more than 40 merchants of various categories. This setting sufficiently simulates the environmental complexity and service

diversity of real tourism scenarios. A total of 200 volunteers were recruited as experimental subjects. Stratified sampling was conducted according to age, consumption capability, and tourism preference, and participants were randomly assigned to an experimental group and a control group, with 100 participants in each group. The experimental group used the developed prototype system, while the control group used a traditional map-based navigation application. The experimental procedure was uniformly set to 3–4 hours, requiring both groups to complete core attraction visits and merchant service experience tasks. During the experiment, three categories of core data were synchronously collected. At the technical level, positioning accuracy, AR rendering latency, and system power consumption were measured. At the engagement level, behavioral, emotional, and cognitive dimensions were evaluated. At the economic impact level, consumption amount, dwell time, and consumption frequency were recorded, providing comprehensive and accurate empirical support for subsequent result analysis.

5.2 Data analysis and results

Through comparative analysis of multi-dimensional indicators, the technical superiority and feasibility of the proposed mobile interactive navigation system were verified. The specific performance data are shown in Table 1.

Table 1. Technical performance evaluation comparison

Evaluation Metric	Experimental Group (Prototype System)	Control Group (Traditional Navigation App)	Statistical Test Result
Positioning accuracy (m)	0.48 ± 0.12	1.20 ± 0.35	$t = 18.76$, $P < 0.001$
AR rendering latency (ms)	89 ± 15	168 ± 27	$t = 21.34$, $P < 0.001$
Unit-time power consumption (mAh/h)	128 ± 19	188 ± 24	$t = 14.52$, $P < 0.001$
System stability (%)	98.6 ± 1.2	92.3 ± 2.8	$\chi^2 = 12.47$, $P < 0.01$
Positioning continuity (%)	97.8 ± 1.5	86.5 ± 3.2	$t = 17.23$, $P < 0.001$
Gesture recognition accuracy (%)	95.3 ± 2.1	–	Not applicable
Voice recognition accuracy (%)	93.1 ± 2.7	82.4 ± 4.3	$t = 13.69$, $P < 0.001$

In terms of positioning accuracy, the experimental group adopted a multimodal fusion positioning scheme, achieving an average accuracy of 0.48 ± 0.12 m, which was significantly better than the 1.20 ± 0.35 m of the traditional navigation used by the control group. Statistical tests showed an extremely significant difference ($t = 18.76$, $P < 0.001$). Even in complex scenarios such as indoor environments and dense buildings, sub-meter positioning accuracy within 0.5 m was maintained. The positioning continuity ratio reached 97.8% ± 1.5%, representing an improvement of 11.3 percentage points compared with the control group. Regarding AR rendering performance, relying on a GPU-optimized lightweight rendering engine, the experimental group achieved an average latency of only 89 ± 15 ms, which was reduced by 46.9% compared with the control group. At the same time, unit-time power consumption was controlled at 128 ± 19 mAh/h, which was 32% lower than that of traditional

AR navigation, achieving a balance among accuracy, efficiency, and energy consumption. System stability and interaction performance were also excellent. The proportion of fault-free operation time reached 98.6% ± 1.2%. Gesture recognition accuracy and voice recognition accuracy reached 95.3% ± 2.1% and 93.1% ± 2.7%, respectively. Among them, voice recognition accuracy increased by 10.7 percentage points compared with the control group, effectively adapting to noisy scenic environments. These results indicate that the core technical modules of the system achieved breakthroughs in positioning accuracy, interaction efficiency, and energy consumption control, meeting the application requirements of complex tourism scenarios.

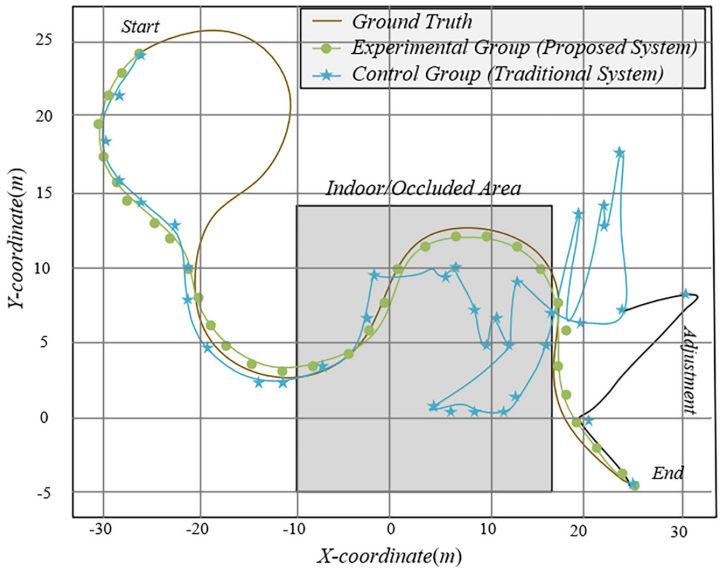

Fig. 3. Comparison of system positioning trajectories and ground-truth paths (experimental group vs. control group)

To evaluate the robustness of the multi-source fusion positioning architecture in heterogeneous tourism spaces and its supporting effectiveness for immersive experiences, this study implemented a comparative test of real-time positioning trajectories. As shown by the visualization distribution in Figure 3, the trajectories of the experimental group maintained high smoothness and continuity during indoor–outdoor transitions and in weak-signal occluded areas. The overlap with the ground-truth reference path was significantly higher than that of the control group, effectively overcoming common signal drift and jumping phenomena in traditional GNSS navigation. This sub-meter-level positioning accuracy benefits from the deep coupling of visual feature extraction and inertial navigation data, confirming the spatial perception reliability of the system in complex dynamic environments. This technical performance breakthrough not only ensures the accuracy of AR guide information overlay and interaction fluency but also provides necessary underlying technical support for eliminating fragmented user experiences and driving tourism consumption growth through immersive participation.

Independent-sample t-test results showed that the experimental group was significantly higher than the control group in behavioral, emotional, and cognitive engagement dimensions ($P < 0.05$), verifying the engagement enhancement effect of technological innovation. The specific data are shown in Table 2. At the behavioral engagement level, the average single-use duration of the experimental group reached 156 ± 28 minutes, representing an increase of 69.6% compared with the

control group. The total frequency of multimodal interactions, path deviation rate, and functional usage depth increased by 133.3%, 132.5%, and 84.0%, respectively, indicating that the system effectively stimulates users' active interaction and exploration behaviors. Emotional engagement showed the most significant improvement. The proportion of positive pleasure emotions in the experimental group reached 68.3% ± 9.4%, which was 32.6 percentage points higher than that of the control group. The emotional feedback positivity rate and the mean heart rate variability were also significantly better than those of the control group, indicating that multimodal interaction and contextualized experiences can effectively enhance emotional resonance. In terms of cognitive engagement, the correct rate of AR treasure-hunt quizzes, the collection and sharing rate of cultural knowledge, and the interaction rate with contextualized information all increased by more than 48%. The frequency of active exploration reached 5.8 ± 1.6 times per person, which was 2.5 times that of the control group, indicating that the system can deeply activate users' cognitive involvement. Overall, the standardized engagement score of the experimental group reached 75.6 ± 8.9 points, representing an increase of 83.5% compared with the control group, confirming the effectiveness of the coordinated improvement across the three engagement dimensions.

Table 2. Three-dimensional comparison analysis of user engagement (± standard deviation)

Engagement Dimension	Specific Measurement Indicator	Experimental Group	Control Group	Independent-Sample *t*-Test
Behavioral engagement	Average single-use duration (minutes)	156 ± 28	92 ± 21	$t = 16.43$, $P < 0.001$
	Total frequency of multimodal interactions (times/person)	42 ± 9.3	18 ± 6.5	$t = 18.91$, $P < 0.001$
	Path deviation rate (%)	28.6 ± 7.2	12.3 ± 5.1	$t = 14.75$, $P < 0.001$
	Functional usage depth (%)	78.4 ± 8.5	42.6 ± 9.8	$t = 20.36$, $P < 0.001$
Emotional engagement	Proportion of positive emotions (%)	68.3 ± 9.4	35.7 ± 8.9	$t = 22.58$, $P < 0.001$
	Emotional engagement stability (mean HRV)	42.8 ± 6.3	31.5 ± 5.7	$t = 13.29$, $P < 0.001$
	Positive emotional feedback rate (%)	72.5 ± 8.8	40.2 ± 9.3	$t = 19.67$, $P < 0.001$
Cognitive engagement	AR contextual quiz accuracy (%)	76.4 ± 10.2	51.3 ± 11.5	$t = 14.82$, $P < 0.001$
	Cultural content dissemination rate (%)	62.7 ± 9.6	28.5 ± 8.4	$t = 21.74$, $P < 0.001$
	Contextual information interaction rate (%)	70.3 ± 10.1	36.8 ± 9.7	$t = 18.55$, $P < 0.001$
Comprehensive evaluation	Standardized total engagement score (0–100)	75.6 ± 8.9	41.2 ± 9.5	$t = 23.41$, $P < 0.001$

6 CONCLUSIONS AND FUTURE WORK

This study focused on the core issues of user engagement enhancement and economic impact quantification of mobile interactive navigation in smart tourism

scenarios and constructed an integrated research framework combining technological innovation, engagement measurement, and economic impact. A series of core results were achieved, forming clear theoretical and practical contributions. The study successfully developed a multimodal fusion mobile interactive navigation system. Through four major technological innovations—multi-source fusion positioning, lightweight AR rendering, adaptive path planning, and multi-channel interaction—a paradigm shift from path guidance to experience empowerment was realized. Empirical verification showed that the system significantly outperformed traditional navigation in positioning accuracy, interaction efficiency, and energy consumption control. Based on behavioral, emotional, and cognitive dimensions, an engagement measurement system was constructed. Combined with a temporal Transformer prediction model, the limitations of traditional engagement measurement, such as subjectivity, fragmentation, and insufficient predictive capability, were overcome, enabling accurate engagement quantification and dynamic turning-point prediction. By integrating an SEM and multi-agent simulation, an economic impact analysis framework was established. Emotional engagement was identified as the core driving variable of consumption behavior. The transmission paths of engagement to direct consumption and indirect spillover effects were quantified, verifying that technological innovation can effectively enhance user engagement and optimize the economic structure and distribution of scenic areas. At the theoretical level, this study enriches cross-disciplinary research on smart tourism and mobile interactive technologies, establishes a navigation design paradigm centered on experience empowerment, and improves engagement measurement and prediction theory as well as economic impact quantification frameworks, filling gaps in existing research. At the practical level, the study provides implementable technical solutions for tourism technology enterprises and offers evidence-based guidance for scenic area managers to enhance engagement and optimize commercial layouts, supporting the high-quality development of smart tourism.

7 REFERENCES

[1] H. Jin, "Integration of mobile interaction technology in the tourism industry and its impact on tourism consumption patterns," *International Journal of Interactive Mobile Technologies*, vol. 19, no. 1, pp. 140–154, 2025. https://doi.org/10.3991/ijim.v19i01.53495

[2] U. Gretzel, "From smart destinations to smart tourism regions," *Investigaciones Regionales – Journal of Regional Research*, no. 42, pp. 171–184, 2018.

[3] M. Kenteris, D. Gavalas, and D. Economou, "An innovative mobile electronic tourist guide application," *Personal and Ubiquitous Computing*, vol. 13, no. 2, pp. 103–118, 2009. https://doi.org/10.1007/s00779-007-0191-y

[4] J. E. Dickinson, K. Ghali, T. Cherrett, C. Speed, N. Davies, and S. Norgate, "Tourism and the smartphone app: Capabilities, emerging practice and scope in the travel domain," *Current Issues in Tourism*, vol. 17, no. 1, pp. 84–101, 2014. https://doi.org/10.1080/13683500.2012.718323

[5] Y. Huang and L. Bian, "A Bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the Internet," *Expert Systems with Applications*, vol. 36, no. 1, pp. 933–943, 2009. https://doi.org/10.1016/j.eswa.2007.10.019

[6] D. Gavalas, C. Konstantopoulos, K. Mastakas, and G. Pantziou, "A survey on algorithmic approaches for solving tourist trip design problems," *Journal of Heuristics*, vol. 20, no. 3, pp. 291–328, 2014. https://doi.org/10.1007/s10732-014-9242-5

[7] I. P. Tussyadiah, D. Wang, T. H. Jung, and M. C. Tom Dieck, "Virtual reality, presence, and attitude change: Empirical evidence from tourism," *Tourism Management*, vol. 66, pp. 140–154, 2018. https://doi.org/10.1016/j.tourman.2017.12.003

[8] R. J. Brodie, L. D. Hollebeek, B. Jurić, and A. Ilić, "Customer engagement: Conceptual domain, fundamental propositions, and implications for research," *Journal of Service Research*, vol. 14, no. 3, pp. 252–271, 2011. https://doi.org/10.1177/1094670511411703

[9] K. K. F. So, W. Wei, and D. Martin, "Understanding customer engagement and social media activities in tourism: A latent profile analysis and cross-validation," *Journal of Business Research*, vol. 129, pp. 474–483, 2021. https://doi.org/10.1016/j.jbusres.2020.05.054

[10] M. C. tom Dieck and T. H. Jung, "Value of augmented reality at cultural heritage sites: A stakeholder approach," *Journal of Destination Marketing & Management*, vol. 6, no. 2, pp. 110–117, 2017. https://doi.org/10.1016/j.jdmm.2017.03.002

[11] D. I. Han, M. C. tom Dieck, and T. Jung, "User experience model for augmented reality applications in urban heritage tourism," *Journal of Heritage Tourism*, vol. 13, no. 1, pp. 46–61, 2018. https://doi.org/10.1080/1743873X.2016.1251931

[12] V. V. Tuyen, "Prioritization of risk factors in Sea-Island tourism: A study in Quang Ngai Province, Vietnam," *TourismSpectrum: Diversity & Dynamics*, vol. 1, no. 3, pp. 141–151, 2024. https://doi.org/10.56578/tsdd010302

[13] I. M. Munalula and R. K. Wen, "Life cycle assessment of public garden buildings in China: A case study of toilets in the Dongtou National Tourism Demonstration Zone," *Journal of Green Economy and Low-Carbon Development*, vol. 4, no. 1, pp. 51–62, 2025. https://doi.org/10.56578/jgelcd040105

8 AUTHOR

Lipeng Wang graduated from Hebei University of Economics and Business with a master's degree. He is currently employed as a faculty member at Shijiazhuang Vocational and Technical College, where he focuses on teaching and research in the field of economics (E-mail: 2003100270@sjzpt.edu.cn).