Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 19 No. 14 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijim.v19i14.57007

PAPER

Interactive Cinematic Experience Design and User Behavior Analysis Enabled by Mobile Augmented Reality

Fangzhao Wang (✉)

Xizang University, Xizang Autonomous Region, Lhasa, China

m18628140048@163.com

ABSTRACT

With the rapid advancement of mobile augmented reality (MAR) technology and its deep integration into the cultural and entertainment sectors, the traditional model of film viewing is undergoing a paradigm shift from passive reception to interactive engagement. The growing demand for immersive and participatory cinematic experiences has catalyzed the convergence of the film industry with MAR technologies. However, existing research on MAR-based interactive film scenes remains constrained by insufficient capture of real-time user interactions and a superficial understanding of underlying psychological mechanisms. This study focuses on the design of interactive experiences within film scenes enhanced by MAR and addresses two core areas. First, a method for generating interactive film scenes was proposed, integrating semantic alignment between virtual elements and real-world environments, interactive logic structuring, and immersion optimization to achieve spatially dynamic and interactive cinematic representations. Second, through the application of eye-tracking and behavioral log analysis, user behavioral patterns during interaction were investigated, with the aim of mapping these behaviors to their corresponding psychological needs. The study is intended to overcome the methodological limitations of prior research, offering a theoretical foundation for the precise design of MAR-based cinematic interaction environments. Additionally, it provides practical insights for the development of interactive film products, aiming to enhance user satisfaction and promote a paradigm innovation in cinematic interaction research through a "technology + art" integrated perspective.

KEYWORDS

mobile augmented reality (MAR), interactive film scenes, experience design, user behavior analysis

1 INTRODUCTION

Amid the rapid development of digital technologies, mobile augmented reality (MAR) has emerged as an innovative technology that seamlessly integrates virtual information with the physical world [1–3], progressively permeating various

Wang, F. (2025). Interactive Cinematic Experience Design and User Behavior Analysis Enabled by Mobile Augmented Reality. *International Journal of Interactive Mobile Technologies (iJIM)*, 19(14), pp. 151–165. https://doi.org/10.3991/ijim.v19i14.57007

Article submitted 2025-04-14. Revision uploaded 2025-05-27. Final acceptance 2025-06-03.

domains and offering users novel forms of interactive engagement. With the continual evolution of the film industry [4, 5], audience expectations have shifted beyond passive viewing toward more immersive and interactive cinematic experiences. The emergence of MAR technologies [6–9] has introduced new possibilities for the presentation of film scenes and the facilitation of real-time user interaction. By enabling virtual cinematic elements to be overlaid onto real-world environments through mobile devices such as smartphones, MAR significantly enhances engagement, the fun, and the sense of participation during film viewing. In this context, the investigation of interactive experience design and user behavior within film scenes enhanced by MAR holds considerable practical significance.

Existing research on augmented reality (AR) and scene interaction has largely relied on traditional user experience evaluation methods, such as questionnaires and interviews [10–13]. However, these approaches often fail to comprehensively capture real-time user behavior and emotional dynamics during interaction. For instance, in the study of AR interaction experiences, Zhang et al. [14] collected user feedback solely through post-experience questionnaires, neglecting the dynamic behavioral data generated during interaction, thereby limiting the depth and accuracy of user experience analysis. Moreover, several studies in user behavior analysis have primarily employed basic statistical approaches, without adequate exploration of the psychological mechanisms underlying user actions [15, 16]. For example, Williams et al. [17] confined the analyses of user interactions with AR to surface-level data—such as click frequency and dwell time—without further investigation into the relationship between user behavior and psychological needs, thus constraining the practical applicability of the research findings.

The present study comprises of two principal research components. The first focuses on the generation of interactive film scenes based on MAR. Through an integrated analysis of cinematic content and the technical affordances of MAR, methods were explored for constructing highly immersive and interactive film environments. This includes the design of virtual elements, the spatial layout of scenes, and the development of interactive logic. The second component centers on the analysis of user behavior in interactive film scenes enhanced by MAR. Advanced data collection techniques—such as eye-tracking and behavioral log recording—were employed to capture detailed user interactions during engagement. These data were used to identify behavioral patterns, user preferences, and influencing factors. Furthermore, an in-depth examination of the relationship between user behavior and psychological needs was conducted.

2 GENERATION OF INTERACTIVE FILM SCENES BASED ON MAR

The core feature of MAR lies in its ability to superimpose virtual digital content onto real-world physical environments in real time through the use of mobile device cameras, sensors, and positioning technologies, thereby forming a hybrid interactive space that fuses virtuality and reality. Based on this principle, the generation of interactive film scenes was structured around a three-layer framework comprising real-world scene anchoring, virtual element integration, and interactive logic association. In this framework, environmental data were first captured through a scene data acquisition module to extract the characteristics of the physical surroundings. These were then semantically aligned with virtual elements derived from intellectual property (IP) content, including characters, props, and narrative segments from the film. Spatial registration between virtual content and the physical scene was achieved through the integration of mobile device-based geolocation, inertial navigation, and computer vision techniques. This ensures that, during user movement, the device screen consistently renders a spatially coherent and dynamically interactive scene that aligns with the actual environment. Figure 1 illustrates a schematic representation of the MAR network architecture designed for interactive cinematic experiences.

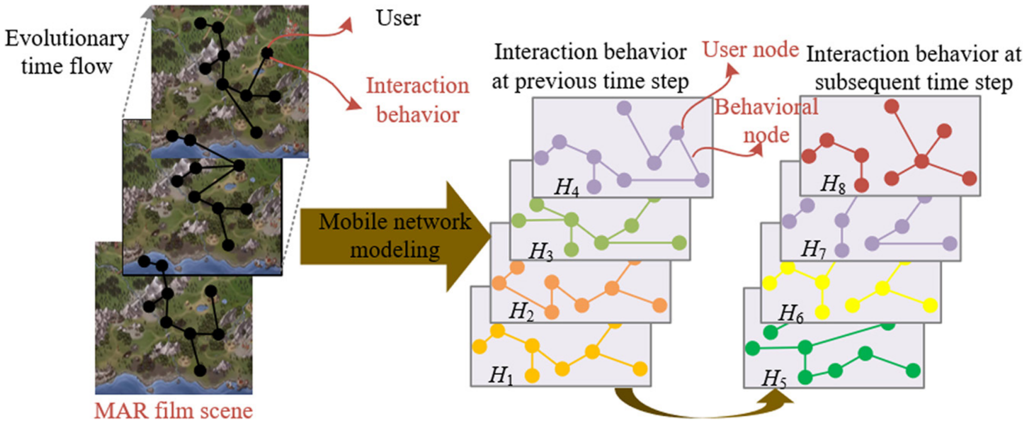

Fig. 1. Schematic diagram of a MAR network for interactive cinematic experiences

To address the dynamic characteristics of MAR scenes and the diversity of user engagement, the generation of interactive cinematic scenes relies on the input and structured processing of multi-source data. During the data acquisition phase, a crowd-sourced collaborative mechanism was employed to integrate professional design input with user-generated creativity. Scene designers define the core virtual elements and interaction rules based on the film's narrative framework, thereby establishing standardized design constraints. Concurrently, crowd-sourced contributors generate diverse scene data by freely combining virtual elements according to the physical features of real-world environments and by scripting interaction scenarios. These data typically include spatial coordinates, element attributes, and event sequences. For instance, in a real cinema corridor, contributors may design an interactive event where a film character emerges from a mural and initiates dialogue as the user approaches the wall. Subsequently, an algorithmic filtering process was applied to eliminate data that violate physical logic or interaction constraints. Only valid scene data—those satisfying spatial mapping requirements in MAR environments—are retained, providing high-quality training inputs for the subsequent generation model.

The generation of MAR cinematic scenes must simultaneously satisfy the dual requirements of real-time performance and interaction diversity. Real-time responsiveness must be achieved despite the computational limitations of mobile devices, while personalized interactive experiences must be delivered to accommodate varying user contexts. To meet these requirements, a generative adversarial network (GAN) was introduced. Through adversarial training between a generator and a discriminator, the model was trained to learn the statistical distribution of real-world scene data, enabling the generation of virtual interaction sequences that are consistent with narrative logic and adapted to real environments. Specifically, the generator receives crowd-sourced scene data as input and employs a multi-layer neural network to synthesize new scene sequences. The discriminator distinguishes between real and generated data, thereby compelling the generator to refine its output toward greater semantic coherence and spatial plausibility. For example, in a task involving the discovery of hidden cinematic props in a real plaza, the generator dynamically determines the virtual placement and trigger effects of the props based on current lighting conditions and crowd distribution. The discriminator ensures that the generated prop locations do not conflict with physical obstacles and that the interaction logic remains faithful to the film's narrative structure.

The final rendering of MAR scenes depends on the coordinated optimization of three-dimensional modeling techniques and mobile device hardware capabilities. On the unity3D platform, the core mechanism for scene generation involves the transformation of structured scene data into lightweight models compatible with the rendering constraints of mobile devices. Initially, spatial mapping technologies were employed to acquire three-dimensional meshes of real-world environments, which serve as anchor points for the placement of virtual elements. Subsequently, high-fidelity models derived from film IP were subjected to model simplification processes to ensure smooth performance on mobile central processing units (CPUs) and graphics processing units (GPUs). Finally, real-time sensor data from mobile devices were used to continuously update the virtual scene's perspective and positioning, thereby establishing a closed-loop interaction flow defined as user movement–scene tracking–interaction triggering. For instance, when users navigate a real street environment, the Unity engine utilizes device-based geolocation data to adjust the relative position of virtual film characters, enabling the characters to appear as if they are walking along the actual street surface. Simultaneously, touch gesture recognition was employed to detect users' tap operations, which can trigger dialogue sequences or initiate narrative branches within the cinematic experience.

3 USER BEHAVIOR ANALYSIS IN INTERACTIVE FILM SCENES BASED ON MAR

An analysis of user behavior within interactive film scenes enhanced by MAR was conducted through a three-stage analytical process: a) detection of user decision-making zones during scene interaction; b) statistical analysis of user interaction behaviors; and c) extraction of behavioral rules underlying user interaction. The overall analytical framework is illustrated in Figure 2. A detailed explanation of each stage is provided below.

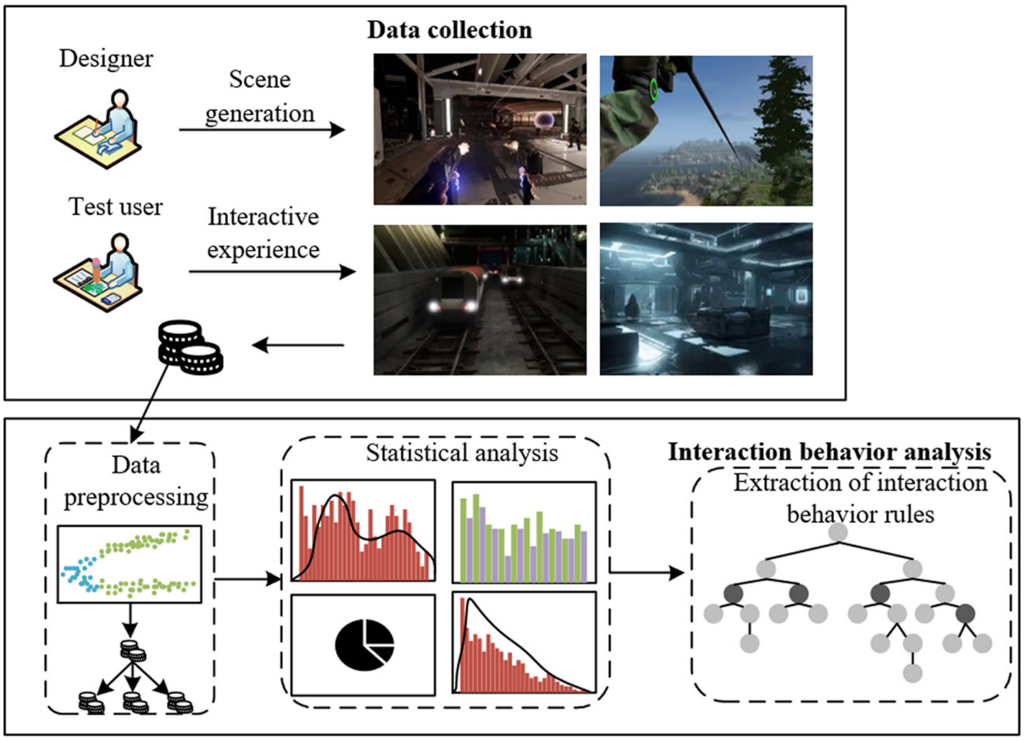

Fig. 2. Analytical framework for user behavior in interactive film scenes based on MAR

3.1 Detection of user decision-making zones in interactive film scenes

In MAR environments, users interact with virtual elements through mobile devices while navigating real-world spaces. Their decision-making behavior is highly dependent on the coupling between physical and virtual spatial domains. Users must continuously adapt their interaction strategies based on the physical constraints of the real environment and the informational density of the overlaid virtual content. Due to the inherently unstructured nature of MAR interactive scenes, user interactions—such as walking, pausing, or turning—tend to cluster within specific physical regions of the environment. These regions serve not only as primary anchor points for virtual content but also as focal areas of user attention. To detect such regions, affinity propagation clustering was employed. This method does not require predefined cluster centers and is capable of autonomously identifying key interaction decision zones based on a similarity matrix computed between data points. For example, repeated triggering of a virtual narrative branch at a physical corner may be algorithmically recognized as a high-value interaction "hotspot" within the blended physical-virtual environment.

The algorithm iteratively exchanges two types of messages—responsibility (attractiveness) and availability (belongingness)—between data points to estimate the likelihood that each point serves as a cluster center. Ultimately, data points exhibiting similar interaction patterns are grouped into clusters, identifying core regions of user interaction decision-making. For instance, if users repeatedly pause and tap to trigger virtual Easter eggs in the lounge area of a real cinema, the algorithm identifies this area as a high-value interaction region. Sensor and geolocation data from mobile devices are also integrated to distinguish between high-engagement interactions occurring during stationary periods and rapid interactions occurring while in motion. Formally, let the set of spatial coordinate samples be denoted as $F = \{a_1, a_2, \ldots, a_v\}$, and let the similarity matrix be denoted as t, where $t(u, k) > t(u, j)$ if and only if the similarity between coordinates a_u and a_k is greater than that between a_u and a_j. The algorithm iteratively updates two message matrices: the responsibility matrix E: $e(u, j)$, and the availability matrix X: $x(u, j)$. All values in matrices E and X were initially set to zero. The iterative procedure of the algorithm proceeds as follows:

Step 1: The responsibility message, denoted as $e_{s+1}(u, j)$, is updated in each iteration according to the following equation:

$$e_{s+1}(u,j) = \begin{cases} t(u,j) - \underset{k \neq j}{MAX}\{x_s(u,k) + e_s(u,k)\}, u \neq j \\ t(u,j) - \underset{k \neq j}{MAX}\{t(u,k)\}, u = j \end{cases} \tag{1}$$

Step 2: The availability message, denoted as $x_{s+1}(u, j)$, is updated in each iteration using the following equation:

$$x_{s+1}(u,j) = \begin{cases} MIN\left\{0, e_{s+1}(j,j) + \sum_{k \neq u,j} MAX\{e_{s+1}(u,k), 0\}\right\}, u \neq j \\ \sum_{k \neq u,j} MAX\{e_{s+1}(u,k), 0\}, u = j \end{cases} \tag{2}$$

Step 3: The above two steps are iteratively executed until either the interaction decision patterns remain unchanged or the maximum number of allowed iterations is reached, at which point the algorithm terminates.

To suppress oscillations during the training process, a damping coefficient η was introduced. The values of $e(u, j)$ and $x(u, j)$ at iteration $s + 1$ were updated using the following formulation:

$$\begin{aligned} e_{s+1}(u,j) &\leftarrow (1-\eta)e_{s+1}(u,j) + \eta e_s(u,j) \\ x_{s+1}(u,j) &\leftarrow (1-\eta)x_{s+1}(u,j) + \eta x_s(u,j) \end{aligned} \tag{3}$$

3.2 Statistical analysis of user interaction behavior

In MAR cinematic scenes, user interaction behavior is influenced by the intertwined effects of physical proximity, virtual density, and interactional effort. Physical proximity determines the ease with which users can approach virtual elements; virtual density affects the efficiency of information acquisition; and interactional effort is directly constrained by the hardware limitations of mobile devices. To model user preferences within continuous interaction spaces, kernel density estimation was employed as the statistical analysis method. This approach enables the smoothing of discrete user behavior data to reveal nonlinear relationships among distance thresholds, density distributions, bandwidth parameters, and interaction choices in MAR scenes. For instance, when analyzing user-triggered dialogue interactions with AR film characters, kernel density estimation reveals that users tend to engage with virtual characters in semi-enclosed real-world environments in order to minimize ambient interference. Such empirical insights provide a theoretical basis for the interaction logic design of virtual elements during scene generation, ultimately enhancing the fluidity and naturalness of physical-virtual interactions.

Prior to statistical modeling, behavioral feature clustering was conducted through a multidimensional tagging system to decompose interaction data and identify user groups with similar behavioral patterns. This stratified analysis serves as a foundation for differentiated scene generation strategies. For example, based on mobile device interaction characteristics, users can be categorized into precision-operating users and gesture-based users; further classification based on real-world mobility patterns yields deeply immersed users and fragmented-experience users. This classification framework reflects the dual nature of MAR interaction—device dependency and environmental dynamism. For precision-operating users, scene generation can be optimized by refining the spatial layout of virtual buttons. For exploratory walking users, narrative trigger points can be preconfigured at key nodes along real-world pathways, enabling a spatial correspondence between the user's physical trajectory and the distribution of virtual content. This alignment enhances the coherence and natural flow of the interactive experience.

Kernel density estimation enables the construction of a probabilistic density distribution model of user selection preferences by smoothing discrete interaction data. This method allows for the quantification of nonlinear relationships between influencing factors and interaction decisions. Specifically, in the dimension of physical proximity, kernel density estimation can identify the "optimal interaction radius" between users and virtual elements, thereby avoiding visual compression caused by excessive proximity or recognition difficulties due to excessive distance. In terms of virtual density, the critical threshold of virtual element quantity within a unit spatial area can be detected, providing quantitative guidance for the density control of scene elements during generation. Regarding interaction width, the effective range of gesture-based interaction can be analyzed, enabling the optimization of interface response sensitivity. Formally, let $a_1, a_2, \ldots, a_v$ represent v independently and identically distributed sample points drawn from distribution D, and let d denote the underlying probability density function. The kernel density estimation is expressed as follows:

$$\hat{d}_g(a) = \frac{1}{v}\sum_{u=1}^{v} J_g(a - a_u) = \frac{1}{vg}\sum_{u=1}^{v} J\left(\frac{a - a_u}{g}\right) \tag{4}$$

3.3 Extraction of user interaction behavior rules

The defining characteristics of interaction in MAR environments are personalization of user behavior and real-time scene responsiveness. Different users may apply varying interaction strategies to the same virtual element, and the interaction rules must dynamically adapt to changes in the physical environment. A fuzzy rule-based classifier optimized through a co-evolutionary algorithm was employed to address the uncertainty and multi-objective optimization challenges inherent in MAR interaction. On one hand, fuzzy rules enable the modeling of imprecise behavioral logic—for example, "when user movement speed is high, prioritize simple tap-based interaction"—to accommodate dynamic variations in user status, device conditions, and environmental context. On the other hand, the co-evolutionary algorithm simulates collaboration and competition among populations to automatically optimize the antecedent parameters and consequent strategies of the fuzzy rules. Specifically, a triangular membership function was adopted. Let the variable be denoted as a, and the coordinate points on the axis be represented by x, y, and z.

$$d(a,x,y,z)\begin{cases} 0, a \le x \\ \dfrac{a-x}{y-x}, x \le a \le y \\ \dfrac{z-a}{z-y}, y \le a \le z \\ 0, z \le a \end{cases} \tag{5}$$

In MAR environments, the variability in user interaction behavior arises from the complex coupling of individual characteristics, device parameters, and environmental conditions. To precisely bind these multidimensional features to corresponding interaction decisions, a structured rule representation was employed. A fuzzy logic-based "feature–decision" conditional rule format was adopted, expressed as: IF ⟨user features + environmental parameters⟩ THEN ⟨interaction strategy⟩. Through this formulation, discrete user labels, continuous sensor data, and interaction parameters of the virtual scene were dynamically linked. This rule-based representation supports heterogeneous, multi-source data inputs and enables explicit binding between features and decisions. As a result, the scene generation system is capable of invoking adaptive interaction strategies in real time based on detected user states, thereby achieving personalized interaction optimization in the form of "one strategy per user."

Given that user characteristics and environmental parameters in MAR interaction scenarios often exist as continuous numerical values, membership functions are required to map precise data into fuzzy linguistic variables—such as "high/medium/low," "fast/slow," or "strong/weak"—in order to construct a decision feature space aligned with human cognitive patterns. For continuous variables, Gaussian membership functions were employed to define fuzzy sets. For example, when the user's actual distance is 1.2 meters, a membership degree of 0.84 may be assigned to the "close-range interaction tendency" fuzzy category, thereby triggering a strategy for enlarging the virtual element display. In contrast, for discrete variables, step-type membership functions were applied to partition decision intervals directly. For instance, on small-screen devices, a membership degree of 1.0 may be assigned to the fuzzy category "low interface element density," thereby enforcing a constraint to reduce the number of virtual elements per unit screen area. Through this hierarchical mapping, precise data from the physical world are transformed into semantic inputs that can be processed by fuzzy rule systems, allowing interaction decisions to be adapted both to the hardware limitations of MAR devices and to users' perceptual and cognitive characteristics.

To ensure the effectiveness and robustness of the fuzzy rule set in MAR environments, a multi-objective fitness function was introduced to guide iterative rule optimization. The core evaluation metrics include:

a) Interaction efficiency metric: The accuracy and response time of user taps on virtual buttons were recorded to evaluate the rule's alignment with device interaction characteristics. For example, on large-screen devices, button spacing rules may be optimized to reduce unintended inputs.

b) User experience satisfaction metric: Eye-tracking data and questionnaire responses were used to quantify the rule's impact on immersion. For instance, if users are observed to frequently skip a certain type of interaction, the corresponding rule weight is reduced.

c) System resource metric: Device computational limitations were considered to constrain rule complexity. For example, on low-end devices, gesture recognition rules are automatically simplified, and lightweight interaction strategies are prioritized. Through simulation of competition and cooperation among rule subsets, the co-evolutionary algorithm dynamically adjusts both the antecedent parameters and consequent strategies based on the defined fitness function.

This data-driven, rule-optimization, and feedback-loop mechanism ultimately enables the adaptive alignment of interaction rules with heterogeneous user demands and dynamically changing scene conditions, thereby providing a decision-making engine with self-adaptive capabilities for interactive cinematic scene generation.

To evaluate the accuracy of predictions, the Kappa coefficient was employed. The Kappa coefficient is calculated as follows:

$$KAPPA = \frac{o_p - o_r}{1 - o_r} \tag{6}$$

Let x_u denote the number of actual samples in class u; y_u the number of predicted samples in class u; and x_{uu} the number of correctly predicted samples in class u. Let j represent the total number of classes and v the total number of samples. The observed agreement rate o_p and the expected agreement rate o_r are defined as:

$$\begin{gathered} o_p = \frac{x_{11} + x_{22} + \ldots + x_{jj}}{n} \\ o_p = \frac{x_1 \times y_1 + x_2 \times y_2 + \ldots + x_j \times y_j}{v \times v} \end{gathered} \tag{7}$$

To quantify the complexity of fuzzy rules, a rule complexity metric was introduced and defined as:

$$COMPLEXITY = \frac{MAX\_N - 0.5NUM\_U - 0.5}{MAX\_N - 1} \tag{8}$$

Assuming that the respective weights for the two measures are represented by α and ε, the final fitness function is defined as:

$$FITNESS = \alpha * KAPPA - \varepsilon * COMPLEXITY \tag{9}$$

4 EXPERIMENTAL RESULTS AND ANALYSIS

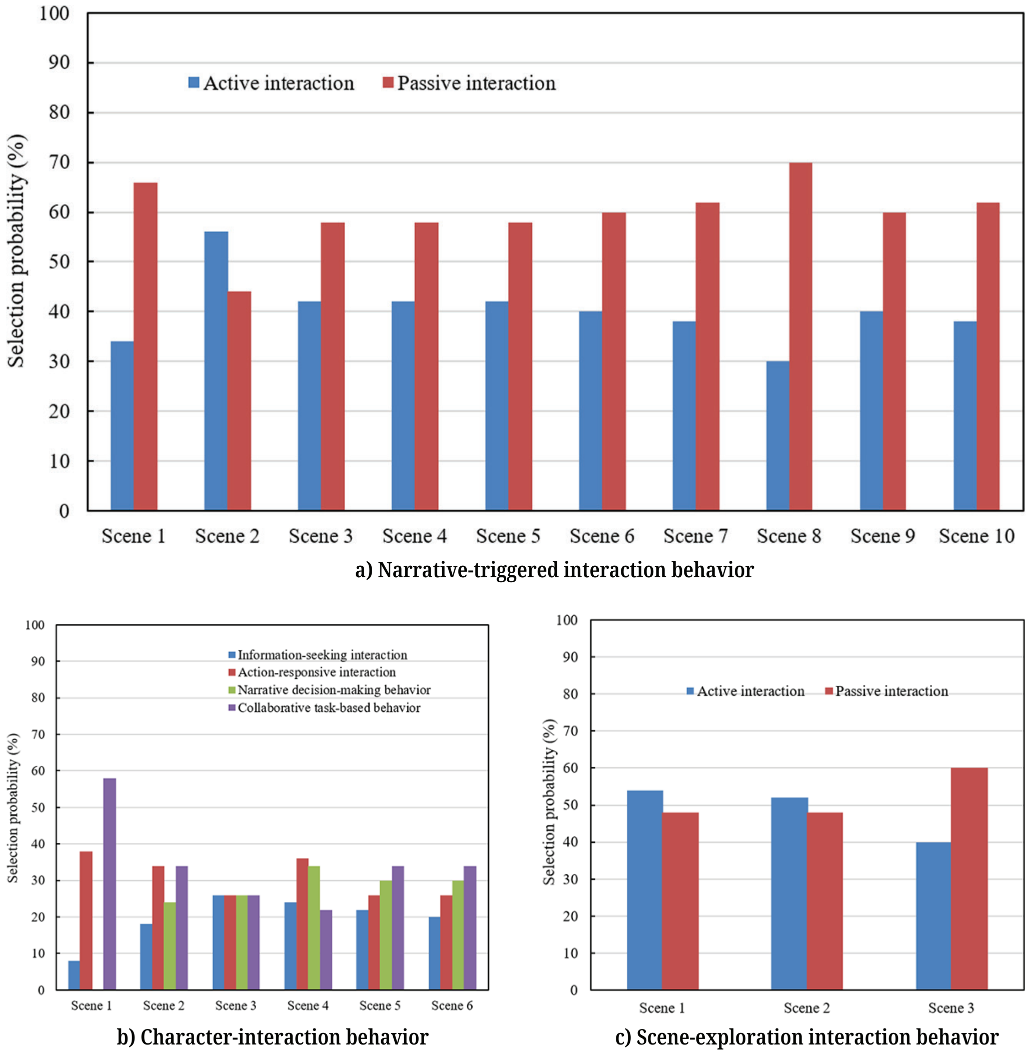

Fig. 3. Experimental results related to user behavior selection in MAR cinematic scenes

In narrative-triggered interaction behavior (Figure 3a), passive interactions exhibited a relatively high selection probability in certain scenes, reaching approximately 65%, while active interactions accounted for about 33%. In other scenes, passive interactions also occupied a substantial proportion, indicating that the design of some narrative-triggered interactions failed to fully align with user expectations.

In character-interaction behavior (Figure 3b), interaction patterns were more diverse. In Scene 1, collaborative task-based interaction exhibited a selection rate close to 60%, while information-seeking interaction remained relatively low. In Scene 4, narrative decision-making interaction accounted for approximately 34%, suggesting varying levels of user acceptance across different types of character interactions and scene contexts. In scene-exploration interaction behavior (Figure 3c), Scene 3 demonstrated a 60% selection rate for passive interactions, exceeding that of active interactions. This outcome indicates that factors negatively affecting user experience may be present in the exploration interaction design of that scene. The experimental findings indicate that the elevated proportion of passive interactions in narrative-triggered scenarios suggests potential deficiencies in narrative integration or interaction logic, which may have limited user engagement. The high selection rate of collaborative task-based interactions in Scene 1 implies a strong user preference for goal-oriented interaction formats. Moreover, the variation in interaction types across different scenes underscores the influence of interaction form–scene alignment on user behavior. The dominance of passive interactions in Scene 3's exploration behavior likely reflects limitations in environmental design or interaction mechanisms.

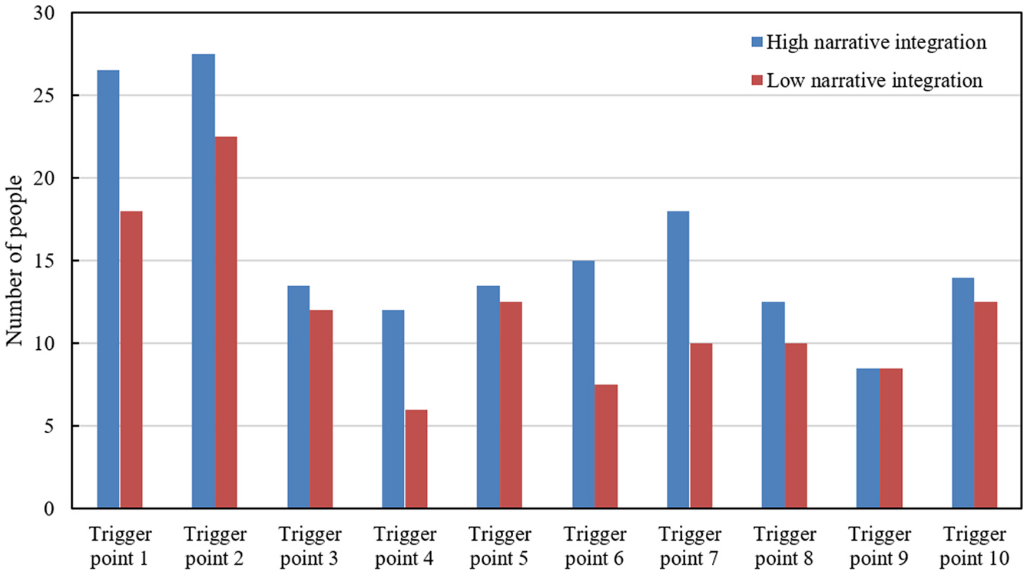

Fig. 4. Influence of narrative integration on interaction behavior decision-making

According to the data presented in Figure 4, at narrative trigger point 1, 26 participants selected the high narrative integration condition, while 18 selected the low integration condition. At trigger point 2, approximately 27 participants selected the high integration condition, and 22 selected the low integration condition. For trigger point 3, the respective values were 14 and 12; for trigger point 4, they were 12 and 6. From trigger points 5 through 10, the number of participants selecting the high narrative integration condition was 14, 15, 18, 13, 9, and 15, respectively, while the corresponding low integration values were 13, 8, 10, 10, 8, and 13. It can thus be observed that at most narrative trigger points, the number of participants choosing the high narrative integration condition exceeded those selecting the low integration condition, with only a few points showing minimal differences. These experimental results indicate that narrative integration exerts a significant influence on user interaction decision-making. Across the majority of narrative trigger points, higher integration levels were associated with greater user engagement in interactive behaviors. This suggests that stronger narrative coherence more effectively encourages user participation by aligning with user expectations for immersive and logically consistent experiences. These findings empirically validate the importance of narrative integration in the generation of interactive scenes, as previously proposed. Although some trigger points exhibited marginal differences—likely due to the interference of external variables—the overall trend underscores the conclusion that enhancing narrative integration is a key factor in fostering active user participation in interactive cinematic experiences.

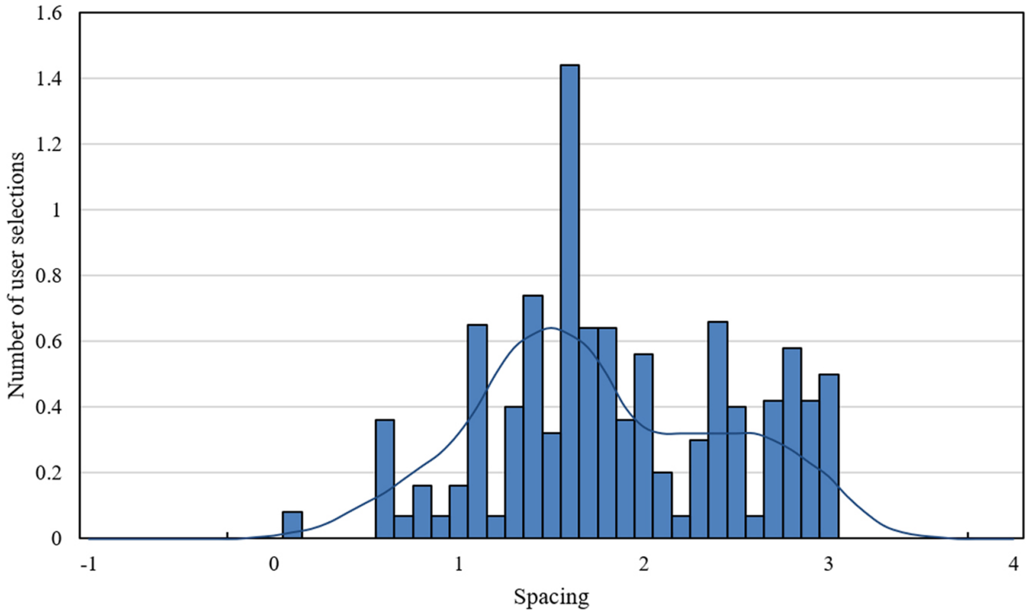

Fig. 5. Influence of narrative trigger point spacing on user interaction decision-making

According to the data presented in Figure 5, when the spatial distance between narrative trigger points was zero, the number of selections was nearly zero. As the spacing increased, the number of user selections fluctuated between 1 and 3, with a peak observed at a spacing of approximately 2, where the number of selections reached nearly 1.5. At other spacing intervals, the number of selections remained relatively low and was more diffusely distributed. These results indicate that user interaction decisions vary significantly with changes in the spatial distribution of narrative trigger points. A higher degree of selection concentration was observed at specific spatial intervals. According to the experimental results, the narrative trigger point spacing has a significant impact on users' interactive decision-making behavior. When the spacing is approximately 2, it becomes the focal range for user selection, indicating the existence of an "optimal interval" for user experience in MAR film scenes. If the trigger points are placed too closely or too far apart, users may experience difficulty in recognition or become confused. In contrast, moderate spacing enables users to focus more easily and engage more effectively in interactive experiences.

From the data presented in Table 1, it can be observed that under active interaction, the number of positive samples for information-seeking behavior is 66, while negative samples total 1148. For creative interaction, the number of positive samples is 169 and negative samples 912. Social sharing behavior shows a relatively higher number of positive samples at 435, with 628 negative samples. Collaborative interaction includes 162 positive and 917 negative samples. For passive interactions, experience disruption behavior contains 36 positive samples and 1231 negative samples. Behavioral disordering is associated with 34 positive and 1152 negative samples, while excessive distraction shows 32 positive and 1124 negative samples. Atmosphere disruption demonstrates a comparatively higher number of positive samples at 128, with 932 negative samples. It is evident that, across both active and passive interaction categories, negative samples significantly outnumber positive samples. These findings suggest that the actual occurrence rate of those interaction behaviors remains relatively low or that the current detection method has limitations. Among active interactions, the relatively high number of positive samples for social sharing indicates a moderate level of user engagement with this interaction type. Conversely, the extremely low number of positive samples for information-seeking behavior reflects a potential shortcoming in either the design's ability to attract user participation or the system's ability to detect this interaction accurately. Among passive behaviors, the higher number of positive samples for atmosphere disruption implies that such behavior is more easily triggered or more reliably detected in actual scenes.

Table 1. Behavior prediction dataset in MAR-based interactive cinematic scenes

<table>
<tr><th colspan="2">Interaction Behavior Type</th><th>Narrative Trigger Point Pair Count</th><th>Positive Samples</th><th>Negative Samples</th></tr>
<tr><td rowspan="4">Active interaction</td><td>1. Information-seeking</td><td rowspan="4">1125</td><td>66</td><td>1148</td></tr>
<tr><td>2. Creative interaction</td><td>169</td><td>912</td></tr>
<tr><td>3. Social sharing</td><td>435</td><td>628</td></tr>
<tr><td>4. Collaborative interaction</td><td>162</td><td>917</td></tr>
<tr><td rowspan="4">Passive interaction</td><td>1. Experience disruption</td><td rowspan="4">1125</td><td>36</td><td>1231</td></tr>
<tr><td>2. Behavioral disordering</td><td>34</td><td>1152</td></tr>
<tr><td>3. Excessive distraction</td><td>32</td><td>1124</td></tr>
<tr><td>4. Atmosphere disruption</td><td>128</td><td>932</td></tr>
</table>

Table 2. Prediction results for interaction behavior using different user scene decision zone detection methods

Algorithm	Precision	Accuracy	Recall	F1-Score	RMSE
K-means	0.684	0.715	0.724	0.715	0.278
Fuzzy C-Means	0.715	0.726	0.678	0.726	0.265
DBSCAN	0.816	0.824	0.789	0.814	0.214
Spectral clustering	0.725	0.726	0.685	0.716	0.268
OPTICS	0.736	0.728	0.721	0.728	0.245
Proposed method (affinity propagation clustering)	0.825	0.818	0.823	0.815	0.178

As shown in Table 2, the K-means algorithm yielded a precision of 0.684, accuracy of 0.715, and recall of 0.724. The fuzzy C-means algorithm achieved a precision of 0.715, accuracy of 0.726, and recall of 0.678. Density-based spatial clustering of applications with noise (DBSCAN) achieved a precision of 0.816, accuracy of 0.824, and recall of 0.789. Spectral clustering produced values of 0.725, 0.726, and 0.685, respectively, while ordering points to identify the clustering structure (OPTICS) achieved 0.736 in precision, 0.728 in accuracy, and 0.721 in recall. The proposed affinity propagation clustering method achieved a precision of 0.825, accuracy of 0.818, and recall of 0.823. Furthermore, this method yielded the lowest root mean square error (RMSE) at 0.178, significantly outperforming the baseline algorithms. These results demonstrate that the proposed method performs robustly in the prediction of interaction behavior. The higher precision and recall values indicate superior accuracy and coverage in identifying diverse interaction patterns, while the minimal RMSE suggests reduced prediction error and improved model stability.

Table 3. Prediction results using different methods for user interaction behavior rule extraction

Algorithm	Precision	Accuracy	Recall	F1-Score	RMSE
ANFIS	0.985	0.942	0.965	0.965	0.178
CART	0.986	0.962	0.725	0.974	0.125
SVM	0.984	0.967	0.726	0.973	0.116
AdaBoost	0.982	0.715	0.846	0.826	0.224
Proposed method (fuzzy rule classifier based on a co-evolutionary algorithm)	0.987	0.956	0.968	0.985	0.112

As shown in Table 3, the adaptive neuro-fuzzy inference system (ANFIS) algorithm yielded a precision of 0.985, accuracy of 0.942, and recall of 0.965. The classification and regression tree (CART) algorithm achieved a precision of 0.986, accuracy of 0.962, and recall of 0.725. The support vector machine (SVM) algorithm produced a precision of 0.984, accuracy of 0.967, and recall of 0.726. Adaptive boosting (AdaBoost) demonstrated a precision of 0.982, accuracy of 0.715, and recall of 0.846. The proposed fuzzy rule classifier based on a co-evolutionary algorithm attained a precision of 0.987, accuracy of 0.956, and recall of 0.968. Notably, this method also recorded the lowest RMSE at 0.112, outperforming ANFIS (0.178), CART (0.125), SVM (0.116), and AdaBoost (0.224). These experimental results demonstrate the superiority of the proposed co-evolutionary fuzzy rule classification method in predicting user interaction behavior. The higher precision, recall, and F1-score indicate enhanced accuracy and comprehensiveness in extracting interaction behavior rules. The lower RMSE further confirms the method's robustness and lower prediction error.

5 CONCLUSION

An in-depth investigation into interactive cinematic experiences based on MAR was conducted in this study, addressing two principal dimensions. First, high-immersion and high-interactivity film scenes were constructed using MAR technologies, encompassing virtual element design, scene spatial layout, and interaction logic modeling. Second, user interaction behaviors—including patterns, preferences, and influencing factors—were analyzed through eye-tracking data and behavior log records to reveal the intrinsic relationships between behavior and psychological needs. The experimental results demonstrated that the proposed method outperformed traditional algorithms such as K-means, DBSCAN, and ANFIS in terms of precision, recall, and RMSE in both behavior prediction and rule extraction tasks. These outcomes provide a strong foundation for accurately modeling user interaction and optimizing MAR cinematic scene design, highlighting the method's significant potential for enhancing interaction experience and fulfilling user psychological expectations.

However, several limitations have also been identified. The diversity and scale of experimental scenes and user groups remain limited, constraining the generalizability of the findings in complex real-world contexts. The adaptability and efficiency of the algorithm under extreme or dynamic environmental conditions require further validation. Moreover, the exploration of users' psychological needs remains superficial, with several latent influencing factors not yet fully addressed. Future research could be expanded in the following directions: a) broaden the experimental scope to incorporate more diverse scenarios and user groups, thereby enhancing the universality of the findings. b) further optimize the algorithmic framework to improve robustness and computational efficiency in complex environments. c) integrate theories and methods from psychology to deepen the interpretation of user behavior and psychological demand. d) explore the convergence of MAR with emerging technologies to further enhance the immersion and interactivity of cinematic experiences, promoting the development of research in this field to a higher level.

6 REFERENCES

[1] P. Wongklang and J. Wipatsopakron, "The development of problem-based mobile augmented reality application to enhance creative problem-solving skills for undergraduate students," *International Journal of Interactive Mobile Technologies*, vol. 18, no. 7, pp. 53–67, 2024. https://doi.org/10.3991/ijim.v18i07.42261

[2] G. Papagiannakis, G. Singh, and N. Magnenat-Thalmann, "A survey of mobile and wireless technologies for augmented reality systems," *Computer Animation and Virtual Worlds*, vol. 19, no. 1, pp. 3–22, 2008. https://doi.org/10.1002/cav.221

[3] M. Zaifri, H. Khalloufi, F. Z. Kaghat, A. Benlahbib, A. Azough, and K. A. Zidani, "Enhancing tourist experiences in crowded destinations through mobile augmented reality: A comparative field study," *International Journal of Interactive Mobile Technologies*, vol. 17, no. 20, pp. 92–113, 2023. https://doi.org/10.3991/ijim.v17i20.42273

[4] H. G. Dixon, D. J. Hill, R. Borland, and S. J. Paxton, "Public reaction to the portrayal of the tobacco industry in the film The Insider," *Tobacco Control*, vol. 10, no. 3, pp. 285–291, 2001. https://doi.org/10.1136/tc.10.3.285

[5] A. Dadlani *et al.*, "Leading by the nodes: A survey of film industry network analysis and datasets," *Applied Network Science*, vol. 9, 2024. https://doi.org/10.1007/s41109-024-00673-9

[6] M. Rusiñol, J. Chazalon, and K. Diaz-Chito, "Augmented songbook: An augmented reality educational application for raising music awareness," *Multimedia Tools and Applications*, vol. 77, pp. 13773–13798, 2018. https://doi.org/10.1007/s11042-017-4991-4

[7] N. Nugegoda, K. Hasan, and P. Lasserre, "Exploring the effect of viewing attributes of mobile AR interfaces on remote collaborative and competitive tasks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 11, pp. 7288–7298, 2024. https://doi.org/10.1109/TVCG.2024.3456177

[8] T. Siu and V. Herskovic, "Mobile augmented reality and context-awareness for firefighters," *IEEE Latin America Transactions*, vol. 12, no. 1, pp. 42–47, 2014. https://doi.org/10.1109/TLA.2014.6716491

[9] J. Joo-Nagata and J. Rodríguez-Becerra, "Mobile pedestrian navigation, mobile augmented reality, and heritage territorial representation: Case study in Santiago de Chile," *Applied Sciences*, vol. 15, no. 6, p. 2909, 2025. https://doi.org/10.3390/app15062909

[10] H. Herwin, L. D. Prasojo, B. Saptono, and S. C. Dahalan, "Analyzing the impact of augmented reality on student motivation: A time series study in elementary education," *Ingénierie des Systèmes d'Information*, vol. 28, no. 5, pp. 1197–1203, 2023. https://doi.org/10.18280/isi.280507

[11] L. Kürvers and J. Manske, "Enhancing the retail experience through augmented reality: The role of flow in brick-and-mortar stores," *Journal of Intelligent Management Decision*, vol. 4, no. 1, pp. 53–65, 2025. https://doi.org/10.56578/jimd040104

[12] G. M. N. Taira, A. C. Sementille, and S. R. R. Sanches, "Influence of the camera viewpoint on augmented reality interaction," *IEEE Latin America Transactions*, vol. 16, no. 1, pp. 260–264, 2018. https://doi.org/10.1109/TLA.2018.8291482

[13] J. Cha, B. C. Lee, J. P. Kim, S. Kim, and J. Ryu, "Smooth haptic interaction in broadcasted augmented reality," in *Human-Computer Interaction – INTERACT 2005*, in Lecture Notes in Computer Science, M. F. Costabile and F. Paternò, Eds., vol. 3585, 2005, pp. 1046–1049. https://doi.org/10.1007/11555261_101

[14] B. J. Jo, S.-K. Kim, and S. Kim, "Enhancing virtual and augmented reality interactions with a MediaPipe-based hand gesture recognition user interface," *Ingénierie des Systèmes d'Information*, vol. 28, no. 3, pp. 633–638, 2023. https://doi.org/10.18280/isi.280311

[15] B. Balajinath and S. V. Raghavan, "Intrusion detection through learning behavior model," *Computer Communications*, vol. 24, no. 12, pp. 1202–1212, 2001. https://doi.org/10.1016/S0140-3664(00)00364-9

[16] S. Möller, N. Ben-Asher, K. P. Engelbrecht, R. Englert, and J. Meyer, "Modeling the behavior of users who are confronted with security mechanisms," *Computers & Security*, vol. 30, no. 4, pp. 242–256, 2011. https://doi.org/10.1016/j.cose.2011.01.001

[17] A. S. Williams, J. Garcia, and F. Ortega, "Understanding multimodal user gesture and speech behavior for object manipulation in augmented reality using elicitation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 12, pp. 3479–3489, 2020. https://doi.org/10.1109/TVCG.2020.3023566

7 AUTHOR

Fangzhao Wang, graduated from the University of Xizang University with a doctoral degree, majoring in ancient Chinese literature (doctoral degree). He is currently a Lecturer at Xizang University and a national Putonghua tester. His research interests include ancient Chinese literature, film and television literature. He presided over one philosophy and social science project in the Xizang Autonomous Region, presided over and completed two teaching reform and curriculum ideological and political projects in Xizang University, published more than ten papers, and edited one textbook of College Chinese (E-mail: m18628140048@163.com).