SIN-CHN SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2025, 22(2), 745.
https://doi.org/10.62617/mcb745

Article

Algorithms for digital cultural tourism ecological model with biomechanical considerations in VR scene interactions

Jing Peng

School of Visual Arts, Hunan Mass Media Vocational and Technical College, Changsha 410100, Hunan, China; 364807136@qq.com

CITATION

Peng J. Algorithms for digital cultural tourism ecological model with biomechanical considerations in VR scene interactions. Molecular & Cellular Biomechanics. 2025; 22(2): 745.
https://doi.org/10.62617/mcb745

ARTICLE INFO

Received: 7 November 2024
Accepted: 15 November 2024
Available online: 17 January 2025

Abstract: Digital cultural tourism is an emerging form of tourism that presents cultural heritage and tourism resources digitally to users, providing immersive travel experiences. However, traditional methods of constructing virtual scenes often rely on manual modeling, leading to low efficiency and high costs. The existing digital cultural tourism platforms mostly provide static and pre-set content, lacking interaction with users, making it difficult to achieve personalized recommendations and interactive experiences. In response to these issues, this article is based on VR (Virtual Reality) scene intelligent generation and interactive algorithms, and aims to optimize the overall synergy between the presentation of cultural resources and user experience by constructing a digital cultural tourism ecological model. Drawing on biomechanical principles, the study emphasizes the importance of natural user interactions and physical engagement in enhancing the immersive experience. Firstly, the Lindenmayer system (L-system) and parameterized generation rules are used to generate complex natural landscapes and architectural structures. Natural and textured scene details are added using the Perlin noise algorithm. Using GANs (Generative Adversarial Networks) technology, generative and discriminative networks are trained to generate more realistic VR scenes, further enhancing the realism and detail representation of the scenes. At the same time, a gesture recognition technology combining CNN (Convolutional Neural Network) and LSTM (Long Short-Term Memory) models, along with a speech recognition algorithm based on DNN (Deep Neural Networks), is adopted to enhance the natural interaction between users and virtual scenes. By combining collaborative filtering algorithms with user behavior data, personalized content recommendations are realized, enhancing user engagement and satisfaction. The efficiency test of scene modeling, the total time required to generate scenes using the model in this article is only 84 hours, which is much lower than manual modeling. In the interactive test, the highest success rate of the model in this article in gesture recognition reaches 94%. The experimental results have verified the advantages of the model in this article in improving scene modeling efficiency and enhancing immersive experiences through biomechanically informed interactions.

Keywords: virtual reality; biomechanics; scene intelligent generation; interactive algorithm; convolutional neural network; deep neural networks

1. Introduction

The current digital transformation in the field of cultural tourism faces multiple challenges. The creation of virtual environments relies on manual operation, which limits production speed and increases economic burden. Existing digital tourism platforms often provide static content and lack dynamic interaction, making it difficult to meet personalized user needs and reducing participation interest. To address these issues, it is necessary to explore more efficient content production methods. Currently, the best solution is to use automated tools and intelligent algorithms to achieve dynamic updates. In addition, the platform should provide flexible interactive mechanisms to meet user needs. Although VR (Virtual Reality) technology brings new experiences to cultural tourism, its popularity and cost remain obstacles, and the threshold needs to be lowered to improve usability and accessibility.

This article focuses on the challenges faced by digital cultural tourism and studies a digital cultural tourism ecological model based on VR scene intelligent generation and interactive algorithms. This article optimizes virtual scene generation through intelligent algorithms, reducing costs and improving efficiency. Gesture and speech recognition technology are applied to enhance user interaction experience and meet user needs through personalized recommendations, promoting the development of digital cultural tourism. This study aims to optimize the overall synergy between cultural resource display and user experience by improving specific indicators such as user satisfaction, increasing interaction frequency, and enhancing the accuracy of personalized recommendations.

The article first analyzes the current situation and challenges of digital cultural tourism in terms of structure, studies a digital cultural tourism ecological model based on VR scene intelligent generation and interactive algorithm, and details the various technologies used in the model. Subsequently, the efficiency, diversity, and interactivity of the model's scene modeling are evaluated through experiments, and the research results and potential impact on the cultural tourism field are summarized in the end.

Main contribution: an intelligent scene generation model based on VR is established, which achieves efficient and realistic automatic construction of virtual landscapes through L-system and Perlin noise; gesture recognition and speech recognition technologies are integrated to enhance the natural interaction experience between users and virtual scenes; a personalized content recommendation based on user behavior data is implemented, enhancing users' personalized experience and satisfaction.

2. Related work

Previous studies have explored improvements and optimizations in digital cultural tourism. Ammirato et al. [1] established a classification framework for value creation, value proposition, and value acquisition by analyzing the business model and key features of cultural tourism mobile applications. The research found that although technological advancements provided new opportunities for the cultural tourism industry, digital enterprises did not fully utilize these technologies to meet user needs. They provided directions for optimizing mobile application services for digital enterprises in their research. Kerdpitak [2] studied the potential improvement effects of innovative management methods such as collaborative networks, digital marketing, service quality, and supply chain management on cultural tourism performance in northeastern Thailand. Through questionnaire surveys and random sampling, the study found that these management innovations were key factors in improving cultural tourism performance. The results emphasized that optimizing these management strategies could significantly improve performance in digital cultural tourism. Guo et al. [3] explored the tourist response to immersive digital museums as an innovative tourism experience. The study revealed three dimensions of pleasure, personal escapism, and locality, and found that visual and auditory cues were key to enhancing tourist experience. Emotional state and presence play a mediating role between sensory experience and tourist experience. These findings contribute to understanding the digital museum experience and provide guidance for optimizing the digital tourism experience. Yanti et al. [4] analyzed social media content and online comments, and explored how digital information promoted communication in the tourism industry and stimulated innovation in rural communities. The study found that effective use of the Internet platform could change the negative view of rural tourism, promote sustainable development, protect cultural heritage, and promote economic development. Preko et al. [5] explored the impact of digital tourism experience on revisiting tourist attractions. Technology driven service innovation can significantly enhance service value, increase traffic, and promote experience sharing. The research results provide insights for tourism website operators on how to use technological innovation to improve customer revisit rates and share experiences, which is of great significance to policy makers and practitioners in the tourism industry. Marwan et al. [6] emphasized the importance of digital technology in providing tourism information, forming a positive destination image, facilitating accessibility, and improving infrastructure. These factors help improve the performance of digital tourism destinations by promoting information sharing between tourism providers and potential tourists. These studies emphasize the shortcomings in technology application and user demand satisfaction in the current field of digital cultural tourism. The importance of management innovation reflects that optimizing content production and user interaction strategies is the key to enhancing the attractiveness of digital cultural tourism. The current personalized needs of tourists are not being taken seriously, and the overall user experience is poor.

Some scholars have attempted to use advanced algorithms to improve the effectiveness of scene generation and increase interaction between the scene and users. Gao et al. [7] proposed a hierarchical graph network for generating 3D indoor scenes. This method considered a complete hierarchical structure from room level to object level and then to object part level, and simplified the learning process by applying functional areas as intermediate agents. Using a variational autoencoder based on conditional recurrent neural networks, furniture objects with fine-grained geometric shapes and their layouts were directly generated. Wu [8] et al. proposed a diffusion-based 3D scene generation model, BlockFusion, which extended the scene by generating unit blocks and seamlessly merging new blocks. BlockFusion can generate geometrically consistent and unbounded large-scale scenes in indoor and outdoor environments through extrapolation and 2D layout adjustment mechanisms, achieving high-quality shape generation. Yang et al. [9] explored the influencing factors of personalized travel recommendations based on the stimulus organism response theory. Research has found that perceived personalization, visual appearance, and information quality are key factors affecting consumers' perception of personalized travel recommendations. Zeng et al. [10] explored how VR tourism experiences can stimulate tourists' willingness to engage in cultural communication behavior by enhancing experiential value and increasing tourist pride. The study not only expanded the theoretical understanding of VR tourism from an emotional perspective, but also had important practical guidance for VR design and destination marketing. These studies have identified the shortcomings in the current field of digital cultural tourism and provided reference for the exploration of improving scene generation efficiency and user interaction experience in this article.

3. Intelligent scene generation and interactive optimization modeling

This article explores the application of intelligent scene generation and interactive optimization modeling in the field of virtual museums. The advancement of digital technology has given birth to virtual museums, a new form of cultural dissemination that is not constrained by time and space, allowing people from all over the world to easily access and appreciate rich cultural heritage. Virtual museums effectively enhance the effectiveness of education and cultural dissemination by providing diverse exhibition content and immersive experiences. The existing virtual museums still have shortcomings in terms of interactivity and personalized experience, which limits the improvement of user experience. Therefore, optimizing existing models to enhance user engagement and experience quality is particularly urgent.

3.1. Landscape and structure generation

Generating complex and realistic landscapes and structures is of great help in enhancing the user experience. This article combines L-system with parameterized generation rules to efficiently construct natural landscapes and architectural structures, achieving enhanced realism and immersion.

L-system [11,12] is a string replacement-based generative system used to simulate complex structures of natural forms. In the environment of a virtual museum, the first step is to set up a basic symbol set, where each symbol represents a specific element. The basic elements and generation rules set in this article are shown in **Table 1**:

Table 1. Symbol set for landscape and structure generation.

Symbol	Element type	Generative rule	Interpretation
T	Tree	T → T[+T] [−T]	The recursive growth of trees, producing branches. [+T] and [−T] represent generating new tree symbols in both directions.
F	Flower	F → F F | F[F]	F F means the flowers are repeated in the same place. F[F] indicates that a flower is formed on a branch of another flower.
G	Grassplot	G → G G | FG	G G means the grass is repeated in the same place. FG indicates the combination of grass and flowers.
R	Rock	R → R R | FR | GR	R R means the rock is repeated in the same place. FR means the rock formed next to the flower. GR indicates that rocks are generated adjacent to grassland.
B	Building	B → BB | B[B]	BB indicates that the building is repeated in the same location. B[B] indicates that the building adds more layers to the existing building.

Table 1 shows the symbol sets and their generation rules for landscape and structure generation used in virtual museum environments. These symbols and rules are based on the L-system, which constructs natural landscapes and architectural structures through string replacement. Each symbol represents a specific element, such as basic trees, flowers, grasslands, rocks, and buildings, and rules define the generation methods and interrelationships of these elements.

The rule definition is completed and the iterative process is carried out. L-system gradually constructs rich natural landscapes and architectural structures by repeatedly applying defined replacement rules. Taking the generation of landscapes and structures in forests as an example, the iterative generation first sets an initial symbol string X. In the virtual museum environment, X represents a defined basic element, namely T. The current symbol string is replaced according to the preset generation rules. The replacement process gradually applies more details and complexity through specific generation rules. A forest is generated, and the T symbol is further replaced with a complex branching structure to achieve the effect of natural growth.

The entire generation process is a process of iterative use of rules in multiple rounds. In the first round, the initial symbol string X is replaced and becomes T. In the second round, according to the rule $T \rightarrow T[+T]\ [-T]$, the symbol string becomes $T[+T]\ [-T]$. In the third round, the symbol string becomes $T[+T]\ [-T]\ [+T]\ [-T]$, and so on. This process is repeated continuously, ultimately forming a complete forest structure.

The recursive nature enables the generated landscape and structure to have a high degree of detail and variation, which can reflect the complexity of plants in nature. After multiple iterations, the generated trees have rich branches and diversity in height, width, and morphology. The generated landscape and structure enhance visual appeal and enhance users' experience in virtual museums.

In order to flexibly adjust landscape features according to different environmental requirements, natural landscapes and buildings suitable for specific scenes are generated. This article extends the L-system by parameterizing the generation process. The core of parameterized generation rules lies in dynamically adjusting various features during the generation process. In the generation of forest landscape and structure, the adjusted features include height, width, branch angle and color.

The height of trees directly affects the visual effect and the sense of hierarchy in landscape and structure. The height parameter H for trees is set, and the value of H for different types of trees is limited. Shorter shrubs are suitable for growing in grasslands, while tall pine trees can provide shade in forest environments, increasing the richness of space. The width of trees affects the coordination between their appearance and their surrounding environment, and the width parameter W is set. The angle of branching affects the growth morphology of trees, and the branching angle parameter A is set. The color of trees affects the visual effect and is closely related to the theme of the virtual museum, and the color parameter C is set. **Table 2** is a regularized table of forest landscape and structural parameters:

Table 2. Regularization table of forest landscape and structural parameters.

Parameter	Description	Value range	Sample
Height (H)	Affects the visual effect and spatial layering of trees	3 m–10 m	Shrub (2.5 m–4 m) Pine tree (7 m–9 m) Beech tree (5 m–7 m) Birch tree (6 m–8 m)
Width (W)	Influences the appearance and harmony with the surrounding environment	0.5 m–2 m	Willow tree (0.5 m–0.6 m) Oak tree (1.2 m–1.8 m) Red maple (0.9 m–1.1 m) Pine tree (1 m–1.4 m)
Branching angle (A)	Affects the growth form of trees	20°–60°	20°-22° (Upright growth) 40°–50°(Spreading growth) 28°–32° (Slightly tilted) 48°–52° (Significant branching)
Color (C)	Influences visual effects and thematic consistency	Seasonal variation	Spring (bright green) Summer (dark green) Autumn (golden) Winter (brownness) Cherry Blossom (pink)

Table 2 shows the parameter regularization in forest landscape and structure generation. By setting features such as height, width, branching angle, and color, the characteristics of trees can be flexibly adjusted to meet specific environmental needs.

3.2. Detail enhancement

In the environment of a virtual museum, in order to enhance the realism of the landscape and structure, this article applies the Perlin noise algorithm to add natural textures and details. Perlin noise [13,14] is a progressive noise generation technique that produces results that are coherent and natural. The Perlin noise algorithm can effectively simulate the irregularity in nature, making the generated elements more vivid and realistic. When applying the Perlin noise algorithm, the base frequency of the noise is set to 0.01 and the amplitude is set to 1. To generate richer terrain details, a multi-layer noise superposition strategy is adopted, and the frequency and amplitude of each layer of noise decrease in the proportion of 1/2^n, where n represents the number of noise layers. The smoothness parameter of the noise function is set to 0.5 to ensure that the generated terrain is both natural and smooth.

The application of Perlin noise first determines the coordinate system generated by the noise. For the natural landscape and structure in the virtual museum, the three-dimensional coordinate (x, y, z) are used to represent the spatial position in the museum. The continuous noise value $P(x, y, z)$ is generated using the Perlin noise function:

$$P(x,y,z) = (1 - \mathrm{F}(t)) \times \mathrm{g}(a) \times (t - a) + \mathrm{F}(t) \times \mathrm{g}(b) \times (t - b) \tag{1}$$

Among them, $F(t)$ represents the interpolation function, and $g(a)$ and $g(b)$ represent the gradient vectors at points a and b.

To achieve more complex effects, this article adopts a fractal noise generation strategy by stacking multiple Perlin noise layers to form richer details. This process involves superimposing various types of noise with different amplitudes and frequencies.

The left side of **Figure 1** shows the normalized images of 5 different amplitudes and frequencies of noise. The horizontal and vertical coordinates are frequency and amplitude, respectively, and the unit a.u. stands for arbitrary units, indicating the relative values of frequency and amplitude. The right figure shows the noise density distribution after superposition processing, which has also been normalized to demonstrate the comprehensive effect of multiple noise layers superposition. The color depth of each point represents the noise density at that location, reflecting the complexity and diversity of natural textures.

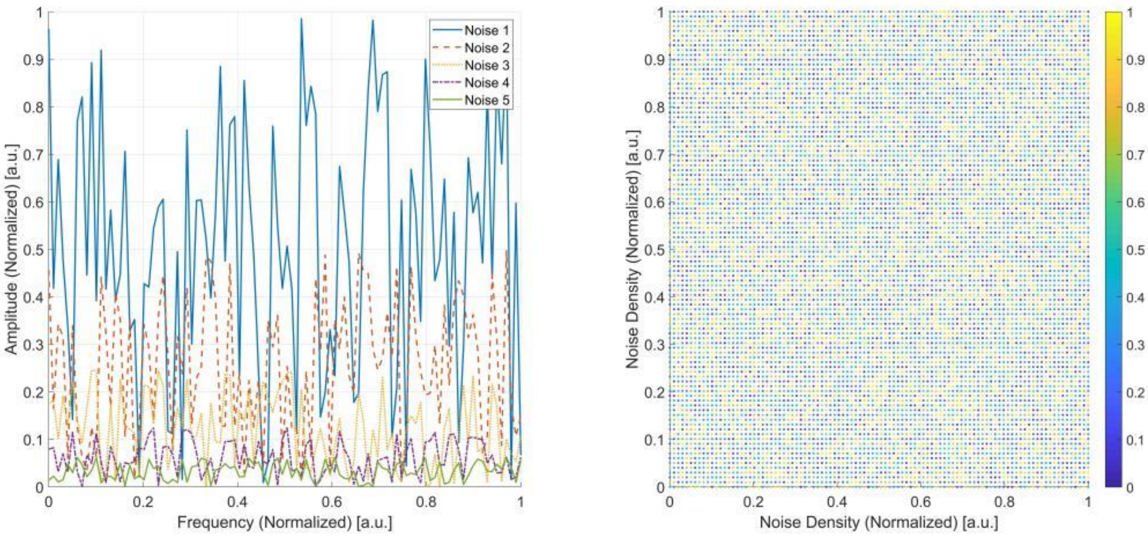

Figure 1. Noise superposition processing.

The generated Perlin noise values are used in virtual museums to control the texture of natural landscape elements. By mapping the noise values to different texture features, rich visual effects can be achieved, enhancing the user's interactive experience.

The terrain is the fundamental part of the entire scene, using Perlin noise to generate height maps, creating undulating terrain, and simulating real-world terrain changes. The height of each ground point is adjusted according to the corresponding noise value, forming a natural high-low fluctuation. By enhancing local details, the distribution of vegetation and stones can be further controlled, ensuring the diversity and richness of the scene, allowing visitors to experience the real natural environment during exploration.

Perlin noise values are analyzed and vegetation textures are generated. The density of trees, shrubs, and grasslands are adjusted based on changes in noise levels to make the vegetation layout more reasonable and natural. The noise values are mapped to different green tones, representing different types of plants and enhancing the layering of vegetation.

To enhance the realism of VR scenes, material and lighting treatments are also required. The Perlin noise is combined for material mapping to generate different textures for the ground, including sand, grass, and stones. Perlin noise provides dynamic shadow effects for elements in the scene. The intensity and position of the light source affect the brightness and reflection of the elements, ensuring coordination with the surrounding environment. Combined with the shadows generated from the height map information, the shadow effect under natural lighting is simulated to enhance the three-dimensionality of the scene.

3.3. Realistic scene generation

In the environment of a virtual museum, this article uses GAN [15,16] technology to train generative and discriminative networks to generate 3D scenes. The architecture design of GAN in this article is shown in **Figure 2**:

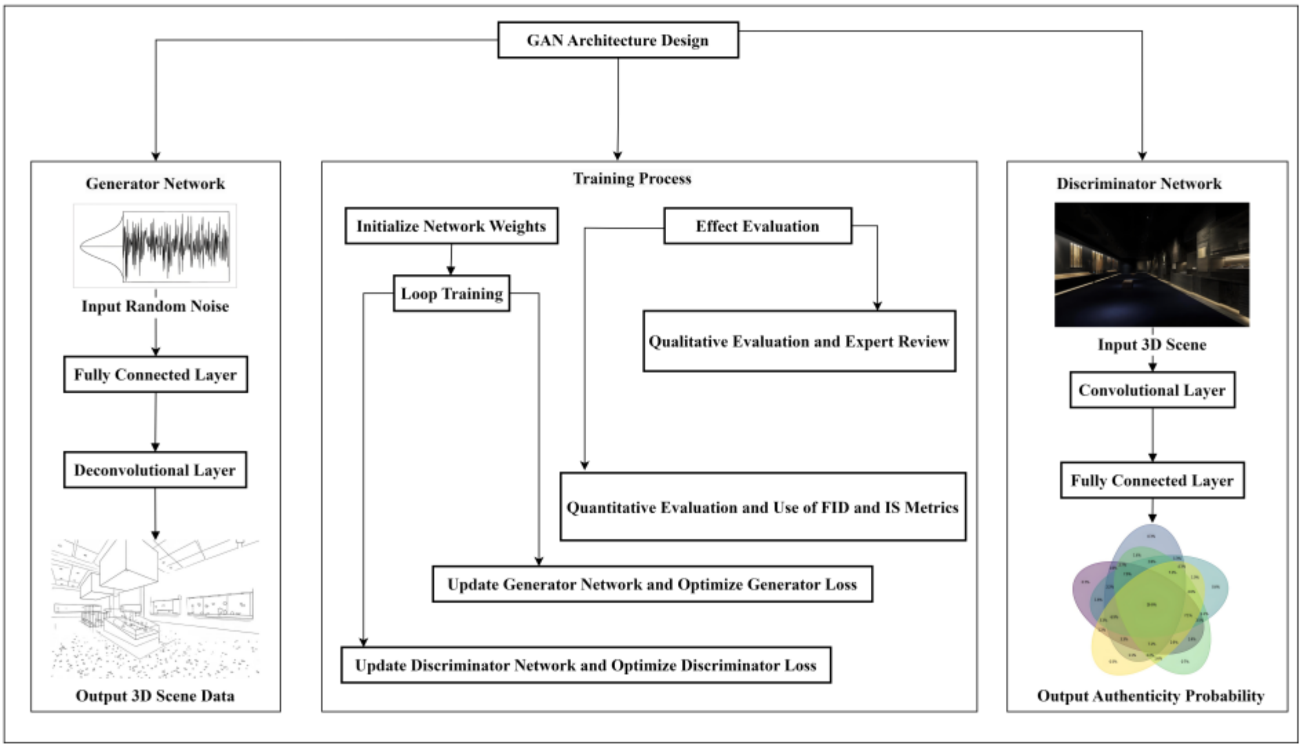

Figure 2. GAN architecture design.

Figure 2 shows the architecture design of using GAN to generate 3D scenes in this article. The generative network converts random noise into a virtual museum scene, which includes an input layer receiving random noise, followed by a fully connected layer and a deconvolution layer gradually upsampling, and finally outputting three-dimensional scene data such as exhibits, environment, and background.

The discriminative network is responsible for determining whether the input virtual museum scene is a real exhibition or generated virtual content. Its structural design is similar to that of a generative network, including multiple convolutional layers for extracting features of exhibits, environment, and background, gradually reducing the spatial dimension of the input, and finally outputting the probability of authenticity through a fully connected layer.

During the training process, after initializing the weights of the generative network and discriminative network, the network is trained iteratively. Firstly, the discrimination network is updated with real exhibits, environment and background samples, and generated virtual samples to optimize the discrimination loss. In the discrimination loss, this article uses cross entropy loss:

$$\mathrm{L_D} = -\mathbb{E}[\mathrm{logD(x)}] - \mathbb{E}\left[\log\left(1 - \mathrm{D}\big(\mathrm{G(z)}\big)\right)\right] \tag{2}$$

Among them, $\mathrm{D(x)}$ represents the probability that the discriminative network judges the input real virtual museum scene as real. $\mathrm{D}\big(\mathrm{G(z)}\big)$ represents the probability that the discriminative network judges the generated virtual scene as real. $\mathbb{E}$ represents the expected value of the input sample.

The output of the discriminative network is used to update the generative network and optimize the generation loss. The formula for generating losses is:

$$L_G = -\mathbb{E}\big[\mathrm{log}D\big(G(z)\big)\big] \tag{3}$$

Among them, $G(z)$ represents the virtual museum scene generated by the generation network on the input random noise vector z. $D\big(G(z)\big)$ has the same meaning as discriminant loss. Through multiple iterations, the network gradually converges and continuously improves the quality of the generated scenes.

The generated virtual museum scene needs to undergo quantitative and qualitative evaluations to verify its quality. Quantitative evaluation uses Fréchet Inception Distance (FID) and Inception Score (IS) metrics to measure the differences and diversity between generated exhibition scenes and real scenes, with lower FID and higher IS indicating higher generation quality. At the same time, feedback is collected through expert review or user research to evaluate the realism, details, and visual appeal of the generated scene.

3.4. Natural interaction

In virtual museums, achieving natural interaction between users and virtual scenes requires the combination of gesture recognition and speech recognition technology. This article chooses the combination of CNN and LSTM for gesture recognition.

Gesture recognition first collects user gesture training data. Gesture videos are recorded using cameras and depth cameras, and the recognized gesture categories are defined, including “pointing to exhibits”, “clicking on information icons”, and “scrolling through”. The data is recorded under different lighting and environmental conditions, which can improve the diversity and robustness of the dataset, and annotate each gesture. The data preprocessing includes extracting keyframes from the video, normalizing image size and pixel values, and improving the accuracy of gesture recognition.

Efficient gesture feature extraction is achieved through multiple convolutional and pooling layers. The input layer of the model receives processed image frames, and the input is a three channel image. Convolutional layers use convolution kernels of different sizes to capture local features of gestures, combined with ReLU function to apply complexity. The pooling layer reduces the feature dimension, preserves the most important information, and reduces computational complexity. The fully connected layer outputs the probability distribution of each gesture category, allowing users to quickly and accurately recognize gestures when interacting with the virtual museum, enhancing the user experience.

To capture the temporal characteristics of gestures, the gesture recognition system of the virtual museum integrates an LSTM model. The feature extraction results of multiple consecutive frames are input into the LSTM unit to form time series data, and recognize the temporal changes in the start, duration, and end states of gestures. The LSTM layer is capable of processing complex temporal information and combining its output with the results of CNN to ultimately generate gesture recognition results. The integration method ensures that when users interact dynamically in the virtual museum, the system can recognize their gestures in real-time, enhancing the overall immersion and interactivity of digital cultural tourism.

This article uses DNN for speech recognition. Tourists' voice commands are recorded through a microphone, and the recorded voice samples are annotated to ensure that each sample corresponds to the correct text label. Voice features are extracted through short-time Fourier transform. The input layer of DNN receives extracted speech feature vectors and learns features through multiple fully connected hidden layers, while ReLU helps process complex speech patterns. Finally, the probability distribution of each voice command is generated to identify the specific instructions of the tourists.

3.5. Personalized recommendations

In personalized recommendations for virtual museums, the behavioral data from visitors is collected, including their browsing history, interaction history, rating data, and search history. Besides, the data is deduplicated to ensure its uniqueness; timestamps are standardized for unified analysis; the various behaviors of tourists are encoded to ensure consistent data format.

The recommendation of virtual museums is achieved through collaborative filtering algorithms, which are divided into two methods: visitor-based and exhibit-based. The collaborative filtering based on tourists first constructs a similarity matrix by calculating the similarity between tourists. This article uses cosine similarity to calculate the similarity of each visitor entering the virtual museum:

$$S(u,v)=\frac{\sum_{i\in I} r_{u,i}\times r_{v,i}}{\sqrt{\sum_{i\in I} r_{u,i}^2}\times\sqrt{\sum_{i\in I} r_{v,i}^2}} \tag{4}$$

Among them, $r_{u,i}$ and $r_{v,i}$ respectively represent the ratings of exhibit i by the tourist u and the tourist v. The tourist who is most similar to the target tourist is selected as the neighbor based on the calculated similarity matrix.

Based on the behavior data of these neighbors, the rating of tourists who have not viewed the exhibits is predicted:

$$r_{u,j} = \frac{\sum_{v \in N(u)} S(u,v) \times r_{v,j}}{\sum_{v \in N(u)} |S(u,v)|} \quad (5)$$

Among them, $r_{u,j}$ represents the predicted rating of exhibit j by the tourist u, and $N(u)$ represents the collection of neighbors of the tourist u. Based on the behavior data of these neighbors, the ratings of tourists who have not viewed the exhibits are predicted, generating personalized recommendations accordingly. The exhibit-based similarity calculation process is similar to the visitor-based process, which calculates the similarity between exhibits and in turn generates a list of recommendations for visitors.

Periodic surveys and feedback are conducted on tourists to collect their satisfaction with recommendations. Visitors are asked about their perceptions of the relevance, interest, and novelty of the recommended content. Personalized recommendations provide virtual museum visitors with highly relevant content recommendations. It not only improves visitors' browsing experience, but also enhances their engagement and satisfaction.

4. Model utility evaluation

4.1. Scene modeling efficiency testing

Table 3. Modeling time for different method scenes.

Scene Type	Method	Generation Time (Hours)	Detail Enhancement Time (Hours)	Total Time (Hours)
Natural Landscape	Manual Modeling	68	20	88
	Blender Procedural Modeling	45	15	60
	Unity ProBuilder	44	10	54
	The model of this article	32	8	40
Architectural Structure	Manual Modeling	54	15	69
	Blender Procedural Modeling	38	10	48
	Unity ProBuilder	32	8	40
	The model of this article	26	4	30
Exhibit Layout	Manual Modeling	41	10	51
	Blender Procedural Modeling	20	8	28
	Unity ProBuilder	25	6	31
	The model of this article	12	2	14
Overall	Manual Modeling	163	45	208
	Blender Procedural Modeling	103	33	136
	Unity ProBuilder	101	24	125
	The model of this article	70	14	84

The experiment compares the efficiency of traditional manual modeling, Blender Procedural Modeling, Unity ProBuilder and the model proposed in this paper in virtual museum scene modeling. Virtual museum scenes of the same complexity are selected for modeling, ensuring coverage of natural landscapes, architectural structures, and exhibit arrangements. Traditional manual modeling uses

Maya tools to record the time required for scene modeling. The time required for two methods to generate the same scene is compared when using the model in this article for automated modeling. The results obtained are shown in **Table 3**:

The results in **Table 3** retain the integer part. The total time for manual modeling of natural landscapes, architectural structures, and exhibition arrangements is 88 h, 69 h, and 51 h, respectively, for a total of 208 h. The total time for Blender Procedural Modeling and Unity ProBuilder is 136 h and 125 h, respectively. The model in this article takes less time than manual modeling in various scene types, with a total time of only 84 h. In order to further explore the reasons why manual modeling takes a long time, the specific time data for each part of manual modeling are shown in **Table 4**:

Table 4. Time required for manual modeling of various parts in different scenes.

Scene Type	Step	Time (Hours)	Scene Type	Step	Time (Hours)
Natural Landscape	Terrain Creation	35	Exhibit Layout	Exhibit Model Creation	25
	Vegetation Design	26		Layout Design	11
	Detail Layout	15		Information Panel Creation	9
	Texture Application	12		Detail Adjustment	6
Architectural Structure	Base Structure Setup	22		Material Application	15
	Detail Decoration	19		Lighting Setup	13

Table 4 shows the time data for manually modeling various parts in different scenes, including detailed steps for natural landscape and exhibit layout, as well as building structure. The terrain creation of natural landscapes takes the longest time, reaching 35 h. The production time for the exhibit model in the exhibit arrangement is the longest, at 25 h. The construction of the base structure in the building takes the longest time, taking 22 h. From the experimental results, it can be seen that the model proposed in this article significantly reduces time consumption in modeling scenes of the same complexity. The time-consuming use of manual modeling in some stages results in low manual efficiency.

4.2. Assessment of differences and diversity

The differences and diversity between the virtual museum exhibition scenes generated using this article's model and the real scenes are evaluated, and images of real exhibition scenes are selected as baseline samples. Virtual exhibition images corresponding to these scenes are generated. Training is performed using this article's model, and the differences and diversity between the generated images and the real images are quantified by calculating the FID and IS metrics. Five types of exhibition halls are selected for the experiment, generating 50, 100, 150, and 200 images for each exhibition hall, and their FID and IS values are calculated. The FID statistical results are shown in **Figure 3**:

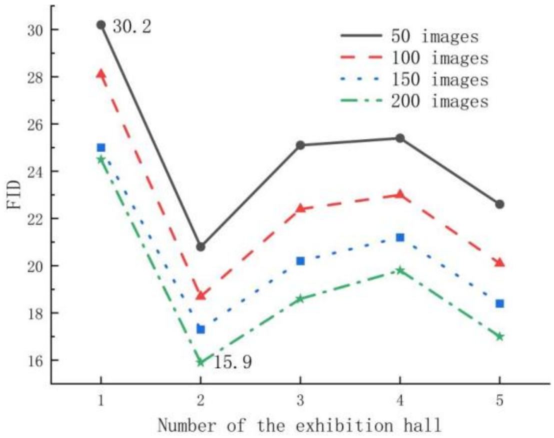

Figure 3. FID value statistical results.

Figure 3 shows the results of FID testing. The FID value represents the difference between the generated image and the real image, and the smaller the value, the smaller the difference. A FID value below 50 indicates good model performance, while a FID value below 20 indicates excellent model performance. The FID value in the experiment decreases as the number of images increases. When testing 50 images, the highest FID value in exhibition hall 1 exceeds 30. As the number of images increases, the FID value gradually decreases. In exhibition hall 2, the lowest FID of 200 images tested reaches 15.9. **Figure 4** shows the statistical results of IS:

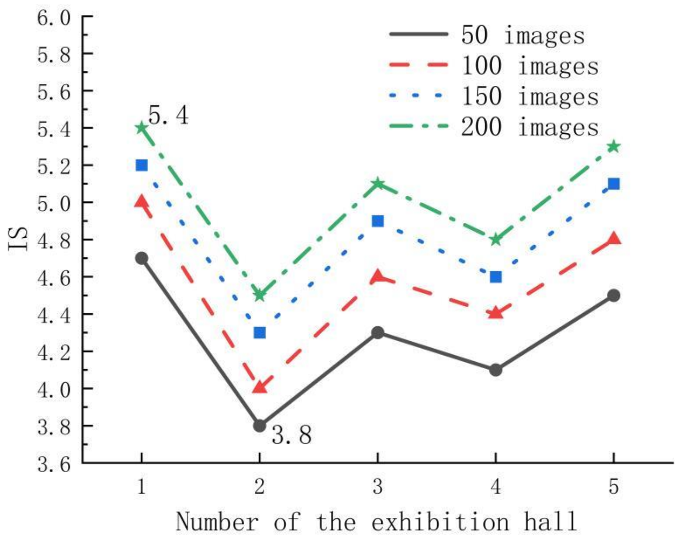

Figure 4. IS statistical results.

The results of the IS test are shown in **Figure 4**. The IS value reflects the diversity and quality of the generated image, with a higher value indicating better quality and diversity of the generated image. An IS value between 4–7 indicates good model performance, while a value above 7 indicates excellent model performance. From the perspective of IS value, the model in this article does not have excellent performance, with the highest IS value being 5.4. Overall, the model presented in this article demonstrates good performance in generating virtual museum exhibition scenes, with low variability and good diversity.

4.3. Interactive testing

The purpose of this experiment is to evaluate the application effect of a digital cultural tourism ecological model based on VR scene intelligent generation and interactive algorithm on interactivity in virtual museums. 20 participants are recruited to ensure a diverse range of browsing habits in the virtual museum, including users of different ages and cultural backgrounds. Participants have one hour of free browsing in the virtual museum, and the model adjusts the recommended content in real-time based on their behavior. The success rate of each participant in each interaction method is recorded. The results obtained are shown in **Figure 5**:

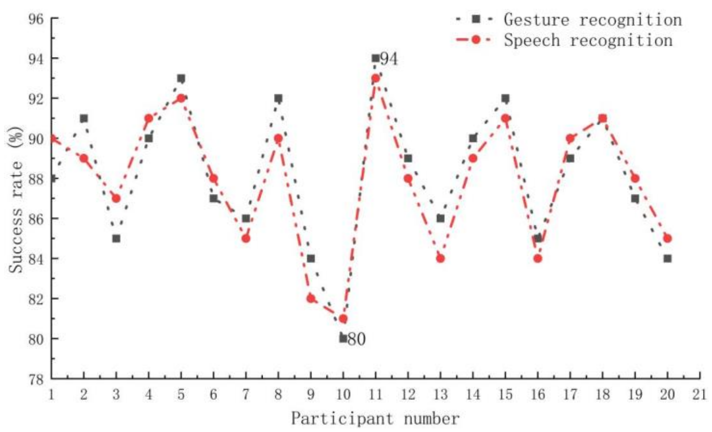

Figure 5. Interactive evaluation results.

From **Figure 5**, it can be seen that the success rate of this article's model in gesture recognition and speech recognition is the lowest at 80%, while the highest success rate in gesture recognition reaches 94%. The results show that the model in this article performs well in user interaction, accurately recognizing users' gestures and speech, and providing a better interactive experience.

For the purpose of comprehensively evaluating the interactive effect of the model, the experiment also collected the participants' specific feelings, problems encountered, and suggestions for improvement during use. Most users said that the immersive experience of the virtual museum is very attractive, and gesture recognition and voice recognition technology make the interaction more natural and smooth. Some users also reported that under certain lighting conditions, gesture recognition occasionally misjudged; some users believed that voice recognition did not perform well in noisy environments. In response to these problems, users suggested that the development team should consider enhancing the system's adaptability in different environments and adding more diverse interaction methods to meet the needs of different users.

5. Conclusions

This article effectively addresses the challenges faced in the field of digital cultural tourism by constructing a digital cultural tourism ecological model based on VR scene intelligent generation and interactive algorithms. The research adopts L-system and parameterized generation rules to automatically construct complex landscapes, utilizes Perlin noise and GAN technology to enhance the realism and details of the scene, and integrates gesture recognition and speech recognition technology to enhance user interaction experience. In addition, the application of collaborative filtering algorithms has enabled personalized content recommendation, further enhancing user satisfaction. The experimental results show that the model has significant advantages in improving scene modeling efficiency, reducing costs, and enhancing interactivity and personalized experience. Future research can further explore how to combine social media data to optimize personalized recommendation algorithms and consider integrating multimodal interactions such as tactile feedback into virtual tourism experiences to achieve broader user participation and deeper cultural tourism experiences.

Ethical approval: Not applicable.

Conflict of interest: The author declares no conflict of interest.

References

1. Ammirato S, Felicetti A M, Linzalone R, Carlucci D. Digital business models in cultural tourism. International Journal of Entrepreneurial Behavior & Research, 2022, 28(8): 1940-1961.
2. Kerdpitak C. The effects of innovative management, digital marketing, service quality and supply chain management on performance in cultural tourism business. Uncertain Supply Chain Management, 2022, 10(3): 771-778.
3. Guo K, Fan A, Lehto X, Day J. Immersive digital tourism: the role of multisensory cues in digital museum experiences. Journal of Hospitality & Tourism Research, 2023, 47(6): 1017-1039.
4. Yanti D, Heryadi D Y, Cakranegara P A, Kadyrov M. Developing rural communication through digital innovation for village tourism. Jurnal Studi Komunikasi, 2023, 7(3): 696-712.
5. Preko A, Amoako G K, Dzogbenuku R K, Kosiba J. Digital tourism experience for tourist site revisit: an empirical view from Ghana. Journal of Hospitality and Tourism Insights, 2023, 6(2): 779-796.
6. Marwan J, Utama Z M. The role of digital tourism destination in achieving digital tourism destination performance through digital's role as intervening variables. International Journal of Human Capital Management, 2022, 6(1): 24-37.
7. Gao L, Sun J M, Mo K, Lai Y K, Guibas L J, Yang J. Scenehgn: Hierarchical graph networks for 3d indoor scene generation with fine-grained geometry. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(7): 8902-8919.
8. Wu Z, Li Y, Yan H, Shang T, Sun W, Wang S, et al. Blockfusion: Expandable 3d scene generation using latent tri-plane extrapolation. ACM Transactions on Graphics (TOG), 2024, 43(4): 1-17.
9. Yang X, Zhang L, Feng Z. Personalized tourism recommendations and the E-tourism user experience. Journal of Travel Research, 2024, 63(5): 1183-1200.
10. Zeng Y, Liu L, Xu R. The effects of a virtual reality tourism experience on tourist’s cultural dissemination behavior. Tourism and Hospitality, 2022, 3(1): 314-329.
11. Lim S G. Construction of HMD-based Interactive Realistic Media Algorithm Using L-System. Journal of Korea Multimedia Society, 2020, 23(1): 58-66.
12. Bernard J, McQuillan I. Stochastic L-system inference from multiple string sequence inputs. Soft Computing, 2023, 27(10): 6783-6798.
13. Etherington T R. Perlin noise as a hierarchical neutral landscape model. Web Ecology, 2022, 22(1): 1-6.
14. Andrian R, Permana A A, Kusnadi A. Design Adventure Role Playing Game With Procedural Content Generation Using Perlin Noise Algorithm. J. Theor. Appl. Inf. Technol, 2023, 101(9): 3653-3665.
15. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Communications of the ACM, 2020, 63(11): 139-144.
16. Gui J, Sun Z, Wen Y, Tao D, Ye J. A review on generative adversarial networks: Algorithms, theory, and applications. IEEE transactions on knowledge and data engineering, 2021, 35(4): 3313-3332.