IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 41, No. 5, October, 2024, pp. 2507-2515
Journal homepage: http://iieta.org/journals/ts

Interactive Theater Experience Design Based on Image Recognition in Virtual Reality Environments

Xiaomu Cai[1], Ziqiao Wang[2*]

[1] College of Arts, Yanbian Universtiy, Yanji 133000, China
[2] Academic Affairs Office, Yanbian University, Yanji 133000, China

Corresponding Author Email: zqwang@ybu.edu.cn

https://doi.org/10.18280/ts.410524

Received: 17 March 2024
Revised: 20 August 2024
Accepted: 15 September 2024
Available online: 31 October 2024

Keywords:
Virtual Reality (VR), interactive theater, image recognition, adaptive initial contour, saliency detection

ABSTRACT

With the advancement of Virtual Reality (VR) technology, interactive theater has gained increasing attention as an emerging art form. VR environments provide audiences with immersive experiences, allowing them not only to observe but also to influence the narrative progression. However, existing research primarily focuses on static scenes and simple interaction mechanisms, lacking real-time analysis of dynamic user behavior, which limits engagement and the quality of the experience. Moreover, traditional image recognition techniques often fall short in accuracy and real-time performance when handling complex scenes, making them insufficient for the evolving demands of interactive theater. Therefore, exploring interactive theater experience design based on image recognition—particularly with adaptive initial contouring and saliency detection—becomes crucial. This study aims to enhance the user experience in interactive theater through two main components. First, it investigates adaptive initial contouring of VR images in interactive theater to enable personalized user interactions. Second, it employs superpixel-based contour-aware methods for saliency detection in VR images, aiming to improve the efficiency and accuracy of visual content recognition. Through these studies, this research seeks to provide new technical support and theoretical foundations for creating interactive theater in VR, driving further advancements in the field.

1. INTRODUCTION

With the rapid development of technology, VR technology has gradually become a core driving force for innovation across multiple fields, especially demonstrating broad application prospects in the fields of entertainment and art [1-4]. Interactive theater, as an emerging art form, integrates the characteristics of theatrical performance and audience participation, providing a unique immersive experience for the audience [5-9]. The VR environment offers a new platform for the presentation of interactive theater, allowing the audience not only to be mere observers but also to directly influence the development and outcome of the narrative through their actions and choices. However, how to achieve efficient image recognition and user interaction in a VR environment remains a major challenge in current research.

In this context, research on the design of interactive theater experiences based on image recognition in VR environments has significant theoretical and practical implications [10-15]. First, an in-depth exploration of the interaction between users and virtual environments will help enhance the sense of immersion and participation in theatrical works, promoting emotional resonance among the audience. Second, research on adaptive initial contouring can provide personalized interactive experiences for different users, thus enhancing the appeal of the work. In addition, research on saliency detection based on superpixel and contour awareness can offer a new perspective for optimizing visual content in virtual environments, helping designers better grasp the focus of audience attention [16, 17]. These studies not only enrich the expressive forms of VR art but also provide theoretical support for the development of related technologies.

Although there has been some existing research on VR and interactive theater, current methods still exhibit obvious deficiencies [18-21]. Many studies focus primarily on static scenes or simple interaction mechanisms, lacking real-time analysis and adaptation to dynamic user behavior. Furthermore, some existing image recognition technologies are insufficient in accuracy and real-time performance when processing complex scenes, rendering them ineffective in responding to rapidly changing environments and user demands in interactive theater. This limits audience engagement and experience quality, preventing the full potential of interactive theater from being realized.

This paper aims to address the above issues by proposing a more effective VR image design scheme for interactive theater. The research mainly includes two parts: first, exploring adaptive initial contouring of VR images in interactive theater, using image recognition technology to analyze user characteristics in real time to achieve personalized interactive experiences; second, performing saliency detection of interactive theater VR images based on superpixel and contour

awareness methods, aiming to improve the efficiency and accuracy of visual content recognition. Through these two aspects of research, this paper not only provides new ideas for the creation and design of interactive theater but also lays a solid foundation for enhancing user experience and promoting the application of VR technology in the field of art.

2. ADAPTIVE INITIAL CONTOUR OF VR IMAGES IN INTERACTIVE THEATER

In a VR environment, the design of an interactive theater experience must fully consider the user's sense of immersion and engagement. The main purpose of researching adaptive initial contours for VR images in interactive theater is to enhance the user's personalized experience. When participating in interactive theater, each user may have unique backgrounds, emotional states, and behavioral preferences. Therefore, design teams need to leverage image recognition technology to analyze users' facial expressions, movements, and other physiological feedback in real time to generate initial contours that adapt to the user's characteristics. This adaptive design can not only offer users a narrative progression closer to their psychological expectations but also dynamically adjust scenes and characters, allowing users to feel a stronger sense of engagement and immersion in the virtual environment, thereby significantly improving the overall experience quality of interactive theater.

In the VR environment of interactive theater, the audience's sense of engagement and immersion depends on the quality and precision of image segmentation. Traditional image segmentation methods often require manual setting of initial contours, a process that is both time-consuming and labor-intensive. In practical applications, background interference can lead to suboptimal segmentation effects. For instance, the Active Contour Model (ACM) can be highly influenced by various background features depending on the initial contour, thereby affecting the precision of target extraction. To enhance the segmentation performance of VR images in interactive theater, it is necessary to explore a more efficient and adaptive method for initial contour setting. The Robust Noise via Local Similarity Factor (RLSF) model, based on local similarity information, can conduct segmentation experiments under different initial contours, providing an automated solution that reduces manual intervention and improves segmentation accuracy.

In a VR environment for interactive theater, an ideal initial contour setting must meet specific conditions. (1) Due to the frequent scene changes in interactive theater, the audience's attention and interactive behavior are also highly dynamic. Therefore, designing a system that can automatically generate initial contours can effectively reduce the operational burden on users within the immersive experience, allowing them to focus more on the plot development and character interaction. This adaptive initial contour not only enhances segmentation efficiency but can also quickly adjust to various visual scenes to meet different audience needs, thereby strengthening immersion and engagement. (2) The ideal initial contour should strive to cover the target and maintain an appropriate distance from the target boundary. In VR images for interactive theater, the targets are typically dynamic characters or important scene elements. Ensuring that the initial contour includes these targets can improve segmentation accuracy and real-time performance. If the distance between the initial contour and the target boundary is too large, the segmentation result may be suboptimal, thereby affecting the audience's experience and understanding of the narrative. Thus, designing a system that can intelligently recognize targets and automatically adjust the contour ensures that the initial contour maintains a suitable distance from the target, making the segmentation process more precise.

Interactive theater VR experiences often involve rapidly changing scenes and characters, where audience interactions require the system to respond and adjust swiftly, with the audience's attention typically focused on dynamic characters and key scenes. Hence, effectively capturing the target area while maintaining clear image boundaries is essential. Therefore, this paper adopts Simple Linear Iterative Clustering (SLIC) for coarse segmentation selection of VR images in interactive theater. The advantage of the SLIC algorithm lies in its ability to divide the image into a small number of irregular superpixels based on pixel texture, color, and brightness features. This segmentation approach not only effectively captures the target area but also maintains clear image boundaries. By generating compact and uniform superpixels, SLIC establishes a foundation for subsequent segmentation and recognition steps, ensuring that the audience's immersion experience is not compromised by blurry boundaries, thereby enhancing understanding and engagement with the plot. Furthermore, SLIC's computational speed is relatively fast, enabling it to meet the efficiency requirements in real-time interactive environments. SLIC initializes clustering centers randomly and continually optimizes them during iterations, providing flexibility to adapt to various visual scenes and supporting reliable data for generating adaptive initial contours.

Based on the above analysis, this paper first conducts coarse segmentation of VR images in interactive theater using the SLIC algorithm, aiming to generate adaptive initial contours through fuzzy clustering. Next, by calculating the standard deviation of each superpixel, it quantifies the dispersion of pixel values within each superpixel to identify those with significant features. Suppose the original image I generates an image with v superpixels through the SLIC algorithm, represented as $U_T = [U_{T1}, U_{T2}, U_{T3}, ..., U_{Tv}]$. The calculated standard deviation for each superpixel U_{Tu} is denoted as $\delta = [\delta_1, \delta_2, ..., \delta_v]$, and all superpixel blocks are concentrated into a one-dimensional matrix A_δ.

Then, a Fuzzy Clustering Algorithm (FCM) is applied to classify the one-dimensional matrix A_δ to further refine the feature regions of superpixels. In VR images for interactive theater, the dynamic changes of characters and scenes require the system to rapidly identify and respond to key feature areas. Through clustering of superpixels, fuzzy clustering can aggregate those superpixels containing target features together, thus generating adaptive initial contours. Let z represent the clustering center and l the weighted index, with the FCM loss function given by:

$$M^{DZL} = \sum_{u=1}^{Z}\sum_{k=1}^{v\times 1}\left[i_u\left(a_k\right)\right]^l \left\|a_k - zr_u\right\|^2 \tag{1}$$

The u-th cluster center expression zr_u:

$$zr_u = \frac{\sum_{k=1}^{v\times 1}\left[i_u(a_k)\right]^l a_k}{\sum_{k=1}^{v\times 1}\left[i_u(a_k)\right]^l} \tag{2}$$

Membership function expression $i_u(a_k)$:

$$i_u(a_k) = \frac{\sum_{u=1}^{z}\left|a_k - zr_u\right|^{\frac{-2}{l-1}}}{\sum_{t=1}^{z}\left|a_k - zr_t\right|^{\frac{-2}{l-1}}} \tag{3}$$

3. SALIENCY DETECTION FOR INTERACTIVE THEATER VR IMAGES BASED ON SUPERPIXELS AND CONTOUR AWARENESS

In a VR environment, the core objective of interactive theater experience design is to create an immersive space that evokes strong emotional resonance and engagement from the audience. This study introduces an image saliency detection method based on superpixels and contour awareness as an innovative exploration to achieve this goal. Using superpixel technology, images are divided into more structured regions, facilitating analysis of each region's visual information and features. Contour awareness effectively identifies and extracts the boundaries of key characters and significant scenes within the performance, thereby helping the system more accurately target the audience's focal points. This saliency detection method not only enhances the visual effect of the scene but also enables real-time adjustment of visual content within the dynamic interactive environment, keeping the audience focused on elements with the most emotional and dramatic tension as they engage with the storyline.

The total energy function R of interactive theater VR images serves as a global fitting energy that incorporates overall saliency information of the image, allowing the model to assess the relative importance of target regions across a broader context. For interactive theater VR experience design, constructing R ensures that the audience can capture the dynamic changes of main characters and significant scenes even amid rapidly evolving plots. This total energy function E consists of three components: the local similarity-based fitting energy R_M, the global saliency-based fitting energy R_H, and the energy regularization term R_E. R_M aims to quantify the fit between the contour and the target region by analyzing the local similarity of superpixels within an image, thereby driving the contour to gradually approach the target boundary. The local similarity factor, R_H, ensures that the model fully leverages local structural information within the image, aiding in the identification of salient regions. The energy regularization term, R_E, is used to control the model's smoothness and stability, preventing overfitting and noise interference, thus enhancing the robustness of saliency detection. The constraint term based on gradient similarity is denoted by d, the regional gradient coefficient matrix is represented as $I = [I_1, ..., I_v]$, and the v-dimensional linear combination is expressed as $F = [F_1, \ldots, F_v]^T$. The functional expression is as follows:

$$R = R_M + R_H + R_E \quad s.t.d = \sum_{u=1}^{l} I_u F_u \tag{4}$$

Specifically, during interaction, when the audience engages with characters or objects in the scene, R_M can adjust in real-time to ensure that these interactive elements maintain visual saliency without being overwhelmed by surrounding clutter. This dynamic adjustment not only enhances the audience's sense of immersion but also boosts their interaction with the virtual environment, making each interaction filled with exploratory excitement and surprise. Given the vulnerability of traditional region-based methods to non-Gaussian noise and image feature interference, R_M incorporates a local similarity factor. By analyzing spatial and intensity differences in local regions of the image, it enhances the model's robustness. In the rapidly changing scenes of interactive theater, where the audience's attention often centers on dynamic characters and significant objects, the use of local similarity factors ensures that these key elements are effectively highlighted against complex backgrounds. Assume that a local window centered at pixel a is represented by V_a, the Euclidean spatial distance between two pixels is denoted by f, and the local average intensity value is denoted by mz. The energy computation formula based on region that incorporates the local similarity factor is as follows:

$$MTO(a, mz) = \int_{(b\in V_z)\neq a} \frac{\| U(b) - mz\|^2}{d(a,b)} db \tag{5}$$

Assuming mz_1 and mz_2 represent the intensity averages in the local internal and external regions around contour a, and $G_\gamma(.)$ is the Heaviside function, the local intensity mean z_u is defined as:

$$mz_u(a) = \frac{\int_\Psi L(a,b)U(b)G_u(\Theta(b))db}{\int_\Psi L(a,b)G_u(\Theta(b))db}, u \in [1,2] \tag{6}$$

Assuming the Euclidean spatial distance from pixel b to the local area center a is represented by f, and the parameter defining the size of the local area is e. The expression for the local region $L(a, b)$ based on the local similarity image fitting in interactive theater VR is:

$$L(a,b) = \begin{cases} 1 & f(a,b) < e \\ 0 & other \end{cases} \tag{7}$$

In summary, the expression for the region-based image fitting function R_M is:

$$\begin{aligned} R_M(a) &= \eta_M \sum_{u=1}^{2} \int_\Psi MTD(a, mz_u)\, da \\ &= \eta_M \sum_{u=1}^{2} \int_\Psi \frac{\left|U(b) - mz_u\right|^2}{f(b,a)} dbda \end{aligned} \tag{8}$$

The construction of the global fitting energy function R_H based on saliency aims to enhance the accurate detection of salient areas, especially in natural images and interactive theater scenes. Figure 1 presents a comparison between traditional saliency algorithms and improved saliency algorithms. The traditional local fitting constraint is often affected by various features in the image, leading to suboptimal segmentation results. To address this issue, R_H performs a comprehensive analysis of the image's global features, considering the distribution and relative relationships

of significant information throughout the entire image. This global fitting energy not only focuses on extracting local information but also includes a global evaluation of saliency regions, enabling the model to identify key elements that capture the audience's attention, such as dynamic character movements and changes in important props, within a broader context. Assume that the saliency map is represented by T_v, and the weight of the global fitting energy is represented by η_H. The expression is:

$$R_H(\Theta) = \eta_H \left(\int_\Psi (T_v - t_1)^2 G_\gamma(\Theta) da \right) + \int_\Psi (T_v - t_2)^2 \left(1 - G_\gamma(\Theta)\right) da \Big) \tag{9}$$

The expression for the mean value t_u inside and outside the curve Z in image T_v is:

$$t_u = \frac{\int_\Psi T_v(a,b) G_u(\Psi) dadb}{\int_\Psi G_u(\Psi) dadb}, u \in 1,2 \tag{10}$$

The saliency of interactive theater VR images stems from the uniqueness of VR visuals. The formula for computing image U_{st} in traditional saliency detection algorithms is as follows:

$$U_{st} = \sum_{u=1}^{3} \left| U_u - \mathrm{MEAN}(U_u) \right| \tag{11}$$

Given that certain pixels exhibit very low intensity in at least one-color channel, this paper leverages these points to enhance contrast between the target and background. Figure 2 shows the effect of weighting in RGB image channels. Let the original image be represented as U, with each channel's weight denoted as μ_u. U_1, U_2, and U_3 represent the R, G, and B channels of the original image, respectively. A new saliency detection formula is then proposed as follows:

$$U_{vt} = \sum_{u=1}^{3} \left| U_u - \mu_u \cdot \mathrm{MEAN}(U_u) \right| \tag{12}$$

where,

$$\mu_u = \frac{SUM(U) - SUM(U_u)}{SUM(U)} \quad u = 1,2,3 \tag{13}$$

Consequently, the saliency feature map T_v is expressed as:

$$T_v = \frac{U_{vt} - MIN(U_{vt})}{|ZZ_1 - ZZ_2|} \tag{14}$$

Figure 3 displays a visual comparison of saliency detection in interactive theater scene images.

In interactive theater, where scenes change rapidly, viewers must quickly and accurately focus on dynamic characters or significant elements. The regularization term is introduced to ensure the smoothness of the evolving curve Z, preventing instability in saliency boundaries caused by local noise or irregular shapes. By constraining the curve's perimeter, the regularization term effectively limits the complexity of segmentation results, ensuring that the boundary of the saliency region remains smoother and more natural. This avoids visual distractions due to overfitting, thus enhancing the overall viewing experience. The length regularization term is defined as:

$$M(\Theta(a)) = \sigma_\gamma(\Theta(a)) \, |\nabla \Theta(a)| \tag{15}$$

Since Z is the zero value of the level set function $\Theta(a)$, the energy function can be updated to:

$$R^{TMTRT}(a, \Theta(a)) = \sum_{u=1}^{2} \left(\eta_M \int_\Psi \frac{|U(b) - mz_u|^2}{d(b,a)} G_u(\Theta(a)) + \eta_H \int_\Psi (T - t_u) G_u(\Theta(a)) \right) dbda \quad s.t.d = \sum_{u=1}^{I} I_u F_u \tag{16}$$

Minimizing the energy function with respect to Θ yields the variational level set formula:

$$\frac{\partial \Theta(a)}{\partial s} = \sigma_\gamma(\Theta(a)) \left[\eta_M \left(\int_{(b \in V_a) \neq a} \left(\frac{|U(b) - mz_2|^2}{d(b,a)} - \frac{|U(b) - mz_1|^2}{d(b,a)} \right) \right) db + \eta_H \left((T - t_2)^2 - (T - t_1)^2 \right) + \omega DI \left(\frac{\nabla \Theta(a)}{|\nabla \Theta(a)|} \right) \right] \quad s.t.d = \sum_{u=1}^{I} I_u F_u \tag{17}$$

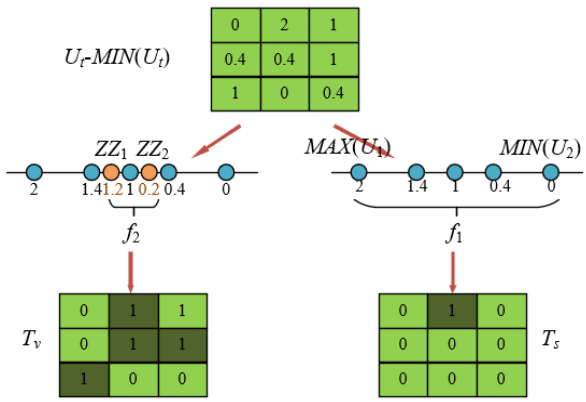

Figure 1. Comparison of traditional saliency algorithm and improved saliency algorithm

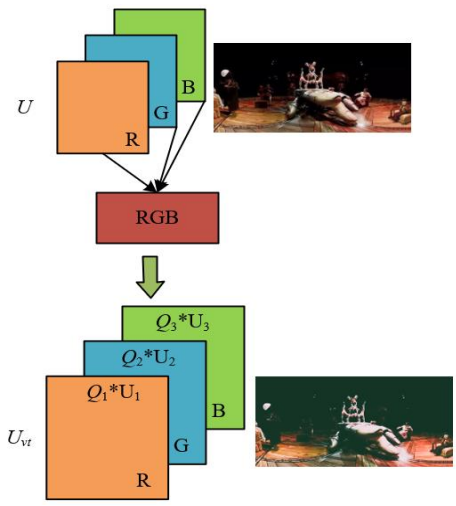

Figure 2. The effect of weighting in RGB image channels

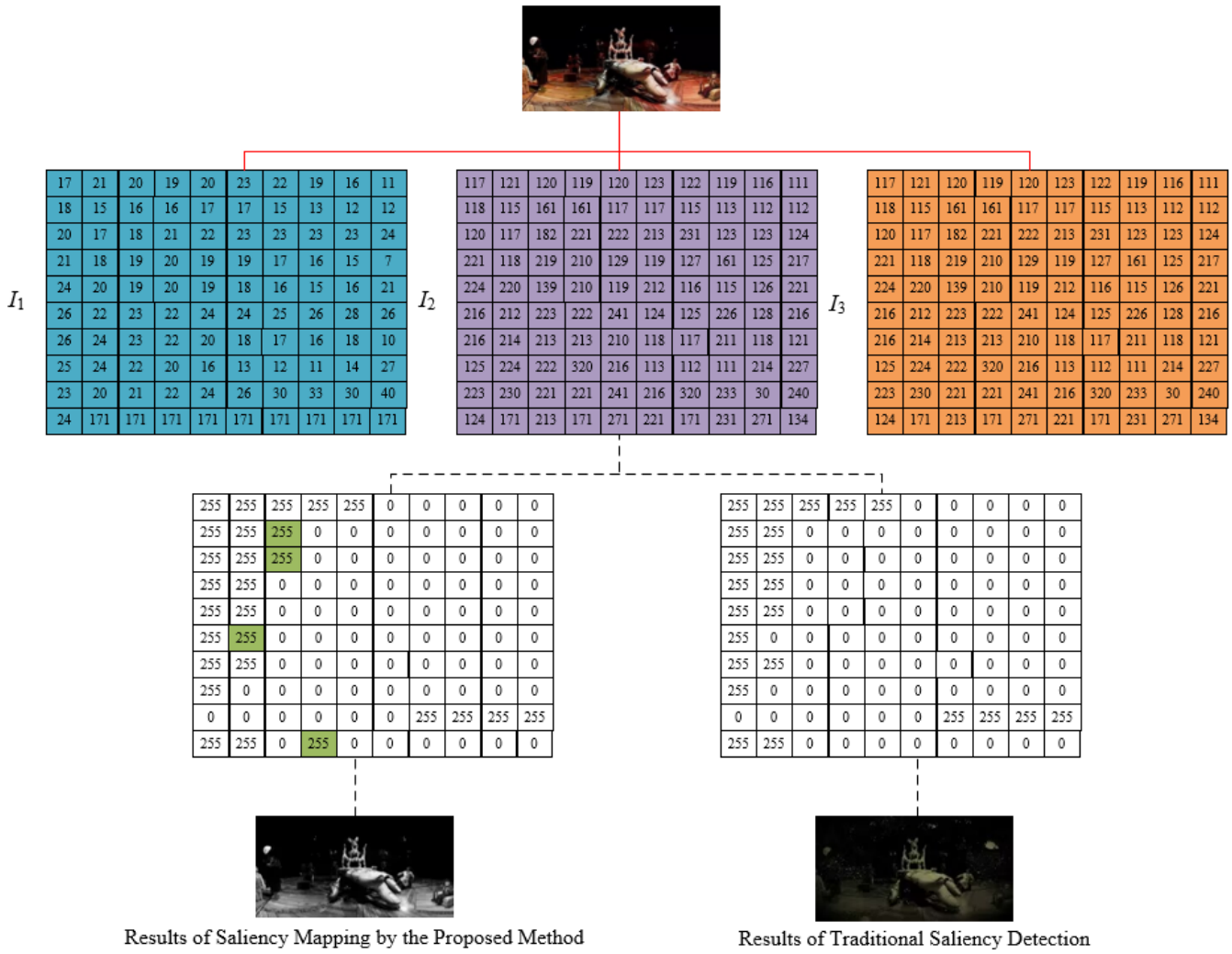

Figure 3. Visual comparison of saliency detection in interactive theater scene images

To effectively extract saliency regions in complex interactive scenes, this study employs the Orthogonal Matching Pursuit (OMP) algorithm to solve the total energy function for image saliency detection. This ensures optimal decomposition results while addressing the diversity and complexity of image features in the interactive theater VR environment. Specifically, by constructing an overcomplete dictionary d containing various possible visual feature atoms, the algorithm can accommodate different types of sparse decomposition requirements. During each iteration, the algorithm first selects the atom I_b from dictionary d that best matches the current signal b and uses region gradient similarity to evaluate the accuracy of each decomposed signal I_u. By calculating the residual $E_u = |I_u - I_b|$, if the residual falls below a preset threshold, the current decomposition is deemed satisfactory, retaining the signal I_u as output; otherwise, the dictionary matrix F_u is updated to enhance the accuracy in the next iteration.

4. ALGORITHM IMPLEMENTATION

The framework for the saliency detection algorithm based on superpixels and contour awareness in interactive theater VR images is shown in Figure 4. The specific implementation steps are as follows:

(1) Preprocessing and Initialization: Begin by inputting the original image of the interactive theater VR scene, setting necessary parameters such as the number of FCM clusters, SLIC seed count, and color space distance-related parameters. These parameters should be selected based on the unique characteristics of the VR environment and the specific demands of interactive theater.

(2) Superpixel Segmentation: Apply the SLIC algorithm to the VR image to perform coarse segmentation, generating a limited number of superpixels. This step helps reduce the complexity of subsequent processing while retaining the main structural information of the image. In a VR interactive theater environment, superpixel segmentation can assist in quickly identifying key elements in the scene, such as characters, props, and background.

(3) Initial Contour Generation: Use the FCM clustering algorithm to cluster superpixel blocks, generating an initial contour. In an interactive theater scene, the initial contour assists in locating potential interactive objects or key visual elements.

(4) Saliency Detection Iteration: Enter the primary iterative loop, which includes the following steps:

(a) Use the improved saliency detection algorithm to compute the saliency map T_v. This step considers the specific requirements of the VR environment, such as a 360-degree view and depth information.

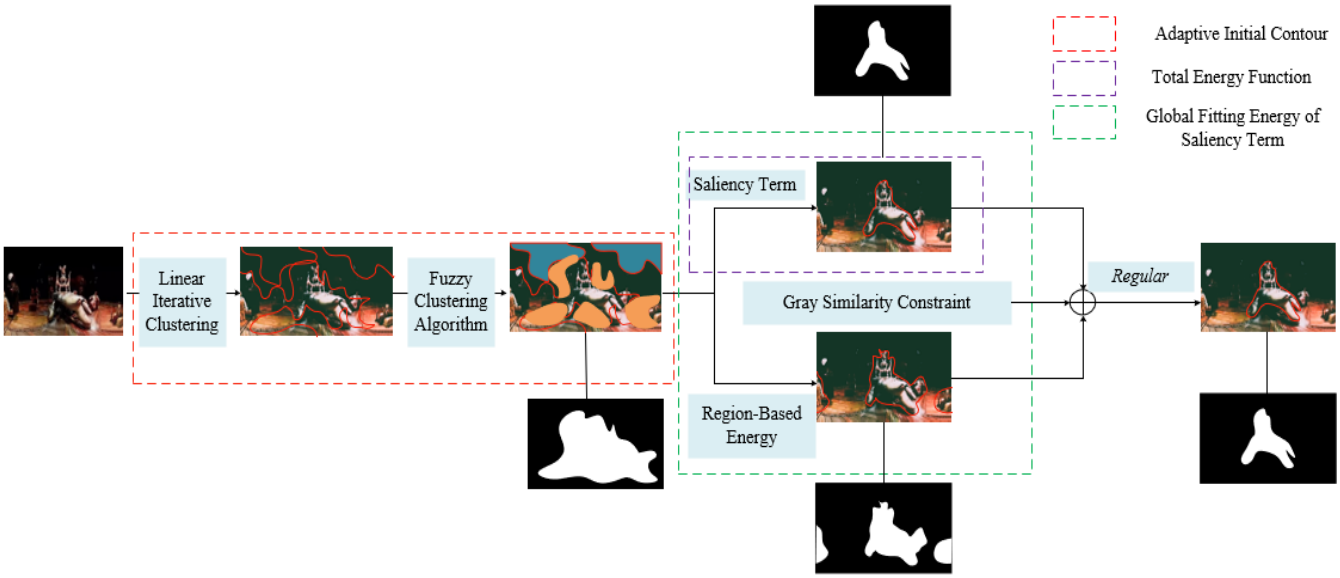

Figure 4. Framework of the saliency detection algorithm for interactive theater VR images

(b) Calculation of mz_u and t_u: These values are crucial for accurately identifying salient objects within the VR environment.

(c) Update the level set function, gradually refine the saliency detection results through iterative optimization.

(d) Periodic apply the gradient similarity constraints to improve the accuracy and stability of the detection.

(5) Result Optimization and Output: After completing the iterations, perform optimization on the final saliency detection results. This includes processes such as edge smoothing and small-region merging, which are essential to meet the needs of the VR environment. Finally, output the optimized saliency map.

5. EXPERIMENTAL RESULTS AND ANALYSIS

From the data in Figure 5, it can be seen that the proposed method performs well in terms of Boundary Recall under different numbers of superpixels. Specifically, as the number of superpixels increases from 250 to 1500, the Boundary Recall improves from 0.53 to 0.86. This result is competitive compared to other saliency detection models. For example, when the number of superpixels is 1000, the Boundary Recall of the proposed method is 0.78, slightly higher than the Itti-Koch Model (0.66) and Graph-Based Visual Saliency (0.71), but lower than the Deep Learning-Based Saliency (0.96) and Spectral Saliency Detection (0.96). Similarly, when the number of superpixels reaches 1500, the Boundary Recall of the proposed method is 0.86, which is comparable to methods such as the Saliency Map Model, Deep Saliency Model, and Visual Attention Model, and significantly better than the Itti-Koch Model (0.77) and Graph-Based Visual Saliency (0.82).

As shown by the data in Figure 6, the proposed method demonstrates competitive performance in terms of accuracy under different numbers of superpixels. Specifically, when the number of superpixels is 250, the accuracy of the proposed method is 0.66, which is higher than that of the Itti-Koch Model (0.56) and Graph-Based Visual Saliency (0.57), and relatively close to Deep Learning-Based Saliency (0.74) and Spectral Saliency Detection (0.7). As the number of superpixels increases, the accuracy of the proposed method gradually decreases, dropping from 0.66 with 250 superpixels to 0.55 with 1500 superpixels. However, this downward trend is common across models, such as Deep Learning-Based Saliency, which decreases from 0.74 to 0.6, and Spectral Saliency Detection, which decreases from 0.7 to 0.6. Notably, when the number of superpixels increases to 1000, the accuracy of the proposed method still maintains a level of 0.59, slightly higher than that of the Itti-Koch Model and Graph-Based Visual Saliency, both at 0.54.

As shown by the F-measure data in Figure 7, the proposed method demonstrates relatively stable and excellent performance across different numbers of superpixels. Specifically, when the number of superpixels is 250, the F-measure of the proposed method is 0.59, which is moderate; in comparison, the Itti-Koch Model is 0.475, and Graph-Based Visual Saliency is 0.5, both lower than the proposed method. As the number of superpixels increases, the F-measure of the proposed method gradually improves, maintaining a range of 0.63 to 0.635 within 500 to 1500 superpixels. This performance surpasses that of the Itti-Koch Model and Graph-Based Visual Saliency and is comparable to other traditional models such as the Saliency Map Model, Deep Saliency Model, and Visual Attention Model, all of which stabilize around 0.63 as the superpixel number increases. Notably, Deep Learning-Based Saliency and Spectral Saliency Detection exhibit higher F-measure values at 500 superpixels and above, reaching 0.76, though their complexity and resource requirements are also relatively higher.

From the data of under-segmentation error in Figure 8, there are significant differences in the performance of different algorithms at different superpixel numbers. The under-segmentation error of the proposed method is 0.3 when the number of superpixels is 250. As the number of superpixels increases, the under-segmentation error gradually decreases, eventually dropping to 0.13 at 1500 superpixels. In contrast, the under-segmentation error of the Itti-Koch Model is 0.45 at 250 superpixels, and it gradually decreases to 0.25 as the number of superpixels increases, still higher than that of the proposed method. The under-segmentation error of Graph-Based Visual Saliency is also higher than that of the proposed

method at all superpixel numbers, decreasing from 0.47 (250 superpixels) to 0.26 (1500 superpixels). Other traditional models, such as the Saliency Map Model and Deep Saliency Model, have under-segmentation errors that are higher than the proposed method across the entire range, especially at higher superpixel numbers, where the under-segmentation errors are 0.2 and 0.205, respectively. It is worth noting that the under-segmentation errors of Deep Learning-Based Saliency and Spectral Saliency Detection are high at all superpixel numbers, particularly the Deep Learning-Based Saliency, which has an under-segmentation error of 0.09 at 1500 superpixels.

From the data of achievable segmentation accuracy in Figure 9, there are significant differences in the performance of different algorithms at different superpixel numbers. The segmentation accuracy of the proposed method is 0.81 when the number of superpixels is 250. As the number of superpixels increases, the segmentation accuracy gradually improves, eventually reaching 0.91 at 1500 superpixels. In contrast, the segmentation accuracy of the Itti-Koch Model is 0.75 at 250 superpixels, and as the number of superpixels increases to 1500, its accuracy is only 0.86. Graph-Based Visual Saliency performs slightly better, with an initial accuracy of 0.74, eventually reaching 0.85. The segmentation accuracy of the Saliency Map Model and Deep Saliency Model is slightly higher across the entire superpixel range, but at 1500 superpixels, they only reach 0.88. Convolutional Neural Networks and Multi-Resolution Saliency have similar performance, with segmentation accuracies of 0.865 and 0.855 at 1500 superpixels, respectively. Deep Learning-Based Saliency and Spectral Saliency Detection have higher segmentation accuracy at all superpixel numbers, particularly Deep Learning-Based Saliency, which reaches 0.96 at 1500 superpixels, but its accuracy fluctuates significantly at lower superpixel numbers.

The aforementioned experimental results show that the proposed superpixel and contour-aware saliency detection algorithm significantly improves segmentation accuracy at different superpixel numbers and demonstrates very high stability and accuracy as the number of superpixels increases. Compared with other traditional and deep learning methods, the proposed method maintains a high segmentation accuracy across the range from 250 to 1500 superpixels, especially at higher superpixel numbers, where the segmentation accuracy is 0.91, significantly outperforming most traditional methods. Although Deep Learning-Based Saliency has the highest accuracy at high superpixel numbers, its accuracy fluctuates significantly at low superpixel numbers, making it less stable.

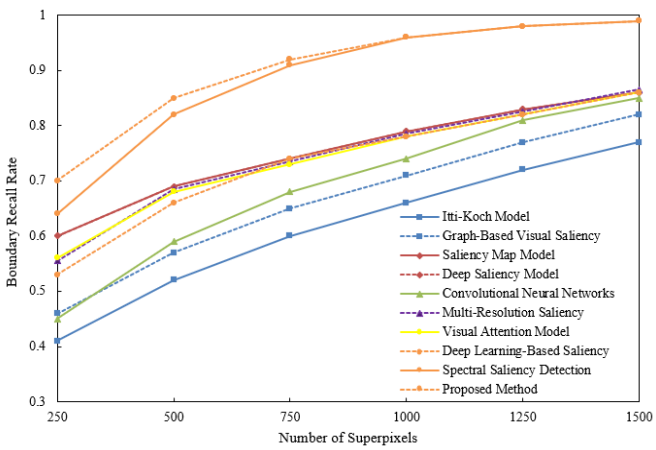

Figure 5. Boundary recall comparison

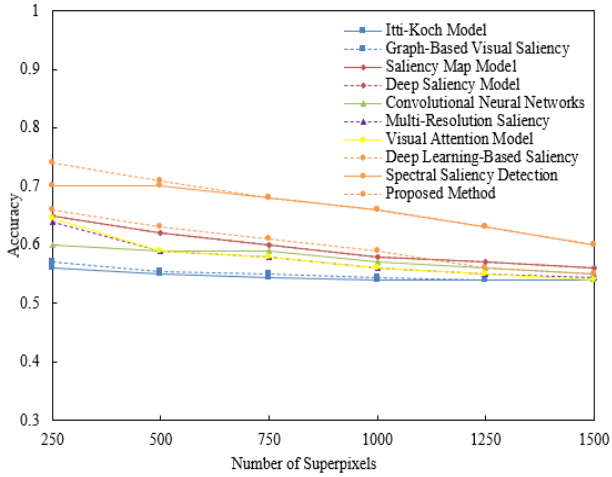

Figure 6. Accuracy comparison

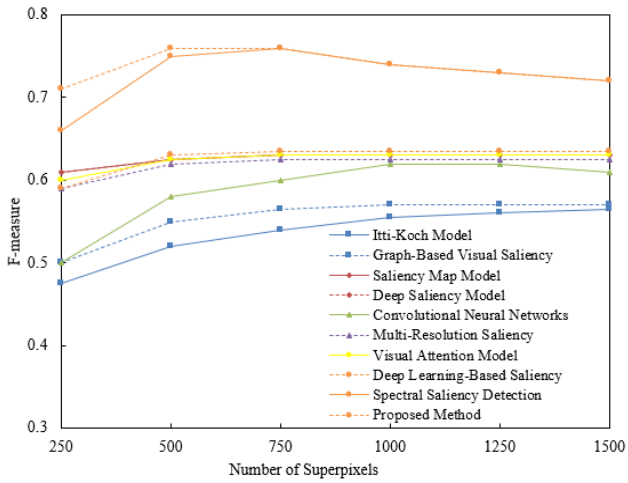

Figure 7. F-measure comparison

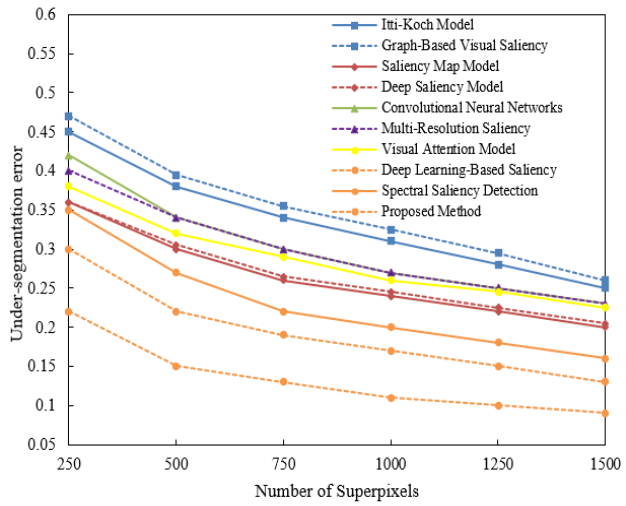

Figure 8. Under-segmentation error comparison

From Table 1, it can be seen that the proposed method performs relatively balanced in various metrics when the number of superpixels is 300. Specifically, the boundary recall is 0.514875, accuracy is 0.662315, F-measure is 0.578405, under-segmentation error is 0.298546, and the achievable segmentation accuracy reaches 0.823215. This performance outperforms many traditional saliency detection algorithms.

For example, the achievable segmentation accuracy of the Itti-Koch Model is only 0.732158, Graph-Based Visual Saliency is 0.721548, and the accuracy of the Saliency Map Model and Deep Saliency Model are 0.812256 and 0.789546, respectively. Although Deep Learning-Based Saliency has the highest segmentation accuracy at 0.874521, its under-segmentation error is 0.213256, indicating its shortcomings in model complexity and stability. While Spectral Saliency Detection shows a higher accuracy of 0.824512, its boundary recall and F-measure are slightly inferior to those of the proposed method.

Through the analysis of the above experimental data, it can be seen that the proposed superpixel and contour-aware saliency detection algorithm performs excellently in several key metrics, especially in terms of accuracy and segmentation performance. Its achievable segmentation accuracy reaches 0.823215, which is significantly higher than most traditional methods. This indicates that the proposed method can not only more accurately identify salient regions in images but also effectively control under-segmentation errors, thereby improving the overall quality of segmentation. For interactive drama experience design in VR environments, the application of this algorithm can significantly enhance user immersion and interaction. By more accurately identifying the visual content that the user focuses on and reducing unnecessary detail confusion, the proposed method provides more precise and efficient technical support for interactive drama VR image design, further optimizing personalized interactive experiences for users and enhancing the efficiency and accuracy of visual content recognition in VR.

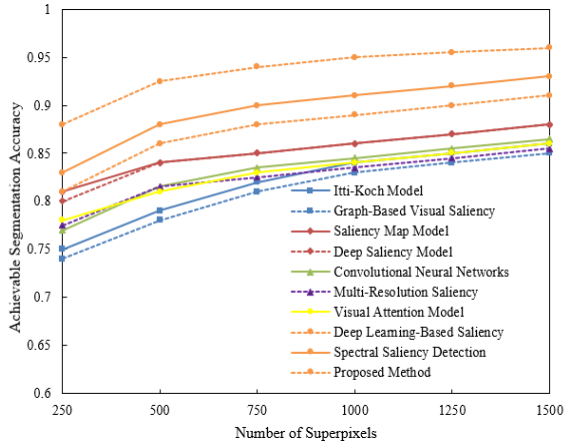

Figure 9. Achievable segmentation accuracy comparison

Table 1. Comparison of different metrics for algorithms at 300 superpixels

Metric / Algorithm	Boundary Recall	Accuracy	*F-measure*	Under-Segmentation Error	Achievable Segmentation Accuracy
Itti-Koch Model	0.421548	0.554152	0.465215	0.432152	0.732158
Graph-Based Visual Saliency	0.445268	0.562358	0.512325	0.456238	0.721548
Saliency Map Model	0.578495	0.623154	0.623154	0.332651	0.812256
Deep Saliency Model	0.589623	0.625866	0.623265	0.351248	0.789546
Convolutional Neural Networks	0.441526	0.612458	0.512148	0.421535	0.754126
Multi-Resolution Saliency	0.552369	0.612352	0.589516	0.389523	0.756233
Visual Attention Model	0.558795	0.623158	0.587415	0.362545	0.765412
Deep Learning-Based Saliency	0.681236	0.725894	0.721562	0.213256	0.874521
Spectral Saliency Detection	0.623514	0.712458	0.652545	0.332145	0.824512
Proposed Method	0.514875	0.662315	0.578405	0.298546	0.823215

6. CONCLUSION

This paper aims to address the issues of adaptability and saliency detection in interactive theater VR image design by proposing a new design solution. The paper explores the adaptive initial contours of interactive theater VR images. Based on image recognition technology, it analyzes user characteristics in real-time to achieve a personalized interactive experience. By dynamically adjusting the initial contours of the images to suit the needs and preferences of different users, it enhances user immersion and engagement. This paper proposes a superpixel and contour-aware saliency detection method to improve the visual content recognition efficiency and accuracy of interactive theater VR images. Through superpixel segmentation and contour-aware technology, it is more effective in detecting salient regions in the image, thus improving the accuracy and speed of image processing. This method helps to more quickly and accurately identify and process key content in the image, enhancing the overall visual effect and user experience.

The experimental results show a comparison of multiple algorithms and metrics, including boundary recall comparison curve, accuracy comparison curve, F-measure comparison curve, under-segmentation error comparison curve, achievable segmentation accuracy comparison curve, and a comparison of various metrics for different algorithms when the number of superpixels is 300. These results indicate that the proposed solution outperforms traditional methods in multiple metrics, especially in terms of boundary recall, accuracy, and F-measure.

Overall, the interactive theater VR image design solution proposed in this paper significantly enhances the user interactive experience and image processing efficiency through adaptive initial contours and saliency detection technology. The experimental results prove that this solution outperforms traditional methods in several key metrics, demonstrating higher boundary recall, accuracy, and F-measure. The research value of this paper mainly lies in: 1) improving the personalization and immersion of interactive theater VR images, enhancing the user experience, and 2) improving the accuracy and efficiency of saliency detection, which is especially important for real-time interactive scenarios.

However, the research also has certain limitations. The study mainly focuses on specific scenarios and user characteristics, and its effectiveness may vary in other types of interactive content. Real-time performance and computational resource consumption still need further optimization to be

applied in broader scenarios. In the future, the application scope of the study will be expanded to verify the suitability and effectiveness of the solution in different types of interactive content. Further optimization of the algorithm will be conducted to reduce computational resource consumption and enhance real-time processing capabilities.

ACKNOWLEDGMENT

This paper was supported by Jilin Province Higher Education Scientific Research Project (Research on the Promotion Strategy of "New Normal" for Online Education in Colleges and Universities, Grant No.: JGJX2022D44); The Third Phase of the Ministry of Education Supply and Demand Docking Employment Education Project (Research on Career Planning for Normal University Students in the Context of University-Enterprise Cooperation, Grant No.: 2023122621628); Ministry of Education Industry-University Cooperation Collaborative Education Project (Experimental Research on Multi-dimensional Journalism Communication, Grant No.: 230703579201612).

REFERENCES

[1] Näykki, P., Pyykkönen, S., Toivanen, T. (2024). Pre-service teachers' collaborative learning and role-based drama activity in a virtual reality environment. Journal of Computer Assisted Learning, pp. 1-14. https://doi.org/10.1111/jcal.13079

[2] Tang, X. (2022). Application and design of drama popular science education using augmented reality. Scientific Programming, 2022(1): 2097909. http://doi.org/10.1155/2022/2097909

[3] Wang, Y.H., Hu, X.B. (2020). Wuju opera cultural creative products and research on visual image under VR technology. IEEE Access, 8: 161862-161871. https://doi.org/10.1109/ACCESS.2020.3019458

[4] Uhm, J.P., Lee, H.W., Han, J.W. (2024). First-person experience in virtual reality sport advertisements: Transportation of embodied empathy. Presence-Virtual and Augmented Reality, 33: 269-286. https://doi.org/10.1162/pres_a_00426

[5] Balfour, M., Cattoni, J., Penton, J. (2022). Future stories: Co-designing virtual reality (VR) experiences with young people with a serious illness in hospital. RIDE-The Journal of Applied Theatre and Performance, 27(4): 458-474. https://doi.org/10.1080/13569783.2022.2034496

[6] O'Dwyer, N., Johnson, N. (2019). Exploring volumetric video and narrative through Samuel Beckett's Play. International Journal of Performance Arts and Digital Media, 15(1): 53-69. https://doi.org/10.1080/14794713.2019.1567243

[7] Muresan, A., McIntosh, J., Hornbæk, K. (2023). Using feedforward to reveal interaction possibilities in virtual reality. ACM Transactions on Computer-Human Interaction, 30(6): 1-47. https://doi.org/10.1145/3603623

[8] Ke, S.Q., Xiang, F., Zuo, Y. (2019). A enhanced interaction framework based on VR, AR and MR in digital twin. In 11th CIRP Conference on Industrial Product-Service Systems, 83: 753-758. https://doi.org/10.1016/j.procir.2019.04.103

[9] Zhang, H.X., Hu, Y., Li, W.L. (2022). A gaze-based interaction method for large-scale and large-space disaster scenes within mobile virtual reality. Transactions in GIS, 26(3): 1280-1298. https://doi.org/10.1111/tgis.12914

[10] Maheshwari, A., Thakur, A., Ahuja, L. (2023). VR tourism: A comprehensive solution with blockchain technology, AI-powered agents, and multi-user features. Journal of Information Assurance and Security, 18(5): 162-172. https://doi.org/10.1007/978-3-031-64650-8_27

[11] Kim, J., Ha, J. (2021). User experience in VR fashion product shopping: Focusing on tangible interactions. Applied Sciences-Basel, 11(13): 6170. https://doi.org/10.3390/app11136170

[12] Pathan, R., Rajendran, R., Murthy, S. (2020). Mechanism to capture learner's interaction in VR-based learning environment: Design and application. Smart Learning Environments, 7(1): 35. https://doi.org/10.1186/s40561-020-00143-6

[13] Liu, X.N., Deng, Y.S. (2021). Learning-based prediction, rendering and association optimization for MEC-enabled wireless virtual reality (VR) networks. IEEE Transactions on Wireless Communications, 20(10): 6356-6370. https://doi.org/10.1109/TWC.2021.3073623

[14] She, Y.Y., Wang, Q., Hu, B. (2023). An interaction design model for virtual reality mindfulness meditation using imagery-based transformation and positive feedback. Computer Animation and Virtual Worlds, 34(3-4): e2184. https://doi.org/10.1002/cav.2184

[15] Rettinger, M., Rigoll, G. (2023). Touching the future of training: Investigating tangible interaction in virtual reality. Frontiers in Virtual Reality, 4: 1187883. http://doi.org/10.3389/frvir.2023.1187883

[16] Matsuda, N., Wheelwright, B., Hegland, J., Lanman, D. (2021). VR social copresence with light field displays. ACM Transactions on Graphics, 40(6): 1-13. https://doi.org/10.1145/3478513.3480481

[17] Khundam, C., Vorachart, V., Preeyawongsakul, P., Hosap, W., Noël, F. (2021). A comparative study of interaction time and usability of using controllers and hand tracking in virtual reality training. Informatics-Basel, 8(3): 60. https://doi.org/10.3390/informatics8030060

[18] Bonfert, M., Muender, T., Steinicke, F., Bowman, D., Malaka, R., Döring, T. (2024). The interaction fidelity model: A taxonomy to communicate the different aspects of fidelity in virtual reality. International Journal of Human-Computer Interaction, 1-33. https://doi.org/10.1080/10447318.2024.2400377

[19] Li, K.Y., Li, X.X. (2022). AI driven human-computer interaction design framework of virtual environment based on comprehensive semantic data analysis with feature extraction. International Journal of Speech Technology, 25: 863-877. https://doi.org/10.1007/s10772-021-09954-5

[20] Taylor, V.J., Valladares, J.J., Siepser, C., Yantis, C. (2020). Interracial contact in virtual reality: Best practices. Policy Insights from the Behavioral and Brain Sciences, 7(2): 132-140. https://doi.org/10.1177/2372732220943638

[21] Fang, Y., Liu, Q., Xu, Y.W., Guo, Y.M., Zhao, T.S. (2023). Virtual reality interaction based on visual attention and kinesthetic information. Virtual Reality, 27: 2183-2193. https://doi.org/10.1007/s10055-023-00801-3