Hindawi
Advances in Mathematical Physics
Volume 2022, Article ID 4743456, 11 pages
https://doi.org/10.1155/2022/4743456

*Research Article*

# Application of Virtual Reality Technology and Unsupervised Video Object Segmentation Algorithm in 3D Model Modeling

**Hui Yang and Qiuming Liu**

*School of Software Engineering, Jiangxi University of Science and Technology, Nanchang 330013, China*

Correspondence should be addressed to Hui Yang; yanghui@jxust.edu.cn

Received 31 August 2022; Revised 23 September 2022; Accepted 29 September 2022; Published 13 October 2022

Academic Editor: Miaochao Chen

Copyright © 2022 Hui Yang and Qiuming Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

3D modeling is the most basic technology to realize VR (virtual reality). VOS (video object segmentation) is a pixel-level task, which aims to segment the moving objects in each frame of the video. Combining theory with practice, this paper studies the process of 3D virtual scene construction, and on this basis, researches the optimization methods of 3D modeling. In this paper, an unsupervised VOS algorithm is proposed, which initializes the target by combining the moving edge of the target image and the appearance edge of the target and assists the modeling of the VR 3D model, which has reference significance for the future construction of large-scale VR scenes. The results show that the segmentation accuracy of this algorithm can reach more than 94%, which is about 9% higher than that of the FASTSEG method. 3D modeling technology is the foundation of 3D virtual scene; so, it is of practical significance to study the application of 3D modeling technology. At the same time, it is of positive significance to use the unsupervised VOS algorithm to assist the VR 3D model modeling.

## 1. Introduction

VR refers to the artificial media space established by computer [1]. With VR technology, the formation of the concept of the complex or abstract system can be made possible by expressing the subcomponents of the system into symbols with exact meanings in some way [2, 3]. Among them, 3D modeling is the most basic technology to realize VR technology. VR simulates things in the real world in virtual digital space, and 3D modeling is to solve the problem of the representation of things in the real world in digital space [4]: how to use the computer to automatically analyze the 3D modeling data effectively and search the 3D modeling content efficiently, all of which bring great challenges to VR 3D modeling [5]. It is a new attempt to apply unsupervised VOS algorithm to the modeling process of the VR 3D model.

Modeling in virtual environment is the foundation of the whole VR system, and VR creates a virtual digital environment that is highly similar to the real environment in vision, hearing, and touch through the use of interactive computer technology as the core of science and technology. Users interact with objects in the virtual environment by using relevant professional equipment. It can also create an experience in the digital environment that is similar to the real environment and can span time and space. In VR, an interactive medium, users can perceive their positions and gestures in the virtual environment. It also causes a strong real sensory response, so that you can immerse yourself in the virtual world. Users rely on graphics and other technologies to feel the simulated objects and characters, so as to immerse their consciousness in the digital environment [6]. In order to create an immersive and realistic environment for users, one of the necessary conditions is to create a realistic virtual scene. When drawing such a complex model, it is often difficult to achieve real-time effect due to the restriction of machine performance, which is also difficult for people to accept [7]. Generally speaking, people need to take a compromise between the fineness of the model and the speed of rendering, which not only ensures a certain rendering quality but also does not cause the user's movement discomfort [8]. Because of the large amount of 3D modeling data and redundant information, and the general efficiency of the existing target segmentation algorithms is low, it is necessary to study and implement a fast target segmentation

algorithm. This paper proposes an unsupervised VOS algorithm, which initializes the target by combining the moving edge of the target image and the appearance edge of the target, assists the modeling of VR 3D model, and studies the integration and scheduling management of VR scene.

To develop a VR application system, we must first analyze the necessary tasks, clarify the purpose and performance index of the tasks, and then arrange appropriate hardware and software resources for the system [9]. The next step is to establish a virtual environment database and apply various physical features, motion constraints, audio, and interactive features to virtual objects and virtual scenes, including geometric modeling, motion modeling, physical modeling, audio modeling, and model segmentation.

The innovative contribution of this paper lies in the combination of the moving edge of the target image and the appearance edge of the target to initialize the target and assist in the 3D modeling of virtual reality. The application of the unsupervised VOS algorithm in 3D modeling of virtual reality technology is analyzed. This algorithm combines the moving edge of the target image and the appearance edge of the target to initialize the target and assist VR 3D model modeling. This paper introduces 3D information, that is, the depth difference between foreground objects and background areas, which can effectively improve the accuracy of object segmentation and make the segmented objects more detailed and complete. The development of virtual reality modeling technology is discussed, and the characteristics, main technical indexes, and basic contents of virtual reality modeling technology are systematically studied. It can be clearly seen that in most video frames, the segmentation results in this paper are better than other methods. In general, the algorithm has a certain practical value because of precision proofreading.

This article will be divided into five parts, and the specific contents are as follows:

The first section introduces the research background and significance and explains the organizational structure of this paper. The second section is related work. The third section analyzes the VR technology. The application of the unsupervised VOS algorithm in the 3D model modeling is discussed. In the fourth section, a lot of experimental analysis is carried out. The fifth section is summary and prospect.

## 2. Related Work

Soares Júnior et al. pointed out that 3D modeling is a core technology in VR [10]. It is written in 3DMax and VRML language, including pattern recognition technology and communication technology. Ko and Sim introduced the application and realization of 3D modeling technology in the joint station system from the aspects of system analysis and design, system 3D virtual scene construction, database technology, scene performance artistry, and 3D object motion simulation [11]. Zhang et al. and Yu et al. pointed out that detecting whether two polyhedra intersect can be done in linear time [12, 13]. If two point sets have disjoint convex hulls, then there must be a plane separating the two point sets. Zhuo et al. introduced the basic content of 3D technology and analyzed and studied its realization method [14]. Cao et al. used Multi Gen Creator and Vega software platform to develop a desktop virtual launch site simulation system [15]. Smith and Hamilton and Hung et al. proposed an object segmentation method using spectral clustering of layer features in images [16, 17]. On the basis of oversegmenting the image using the algorithm, the method extracts the middle-level features of each superpixel, which are edge features and color features, respectively, uses the superpixel as the basic node, fuses these two different features to construct a similarity matrix, and finally uses spectral clustering gets the final target segmentation result. Liu et al. proposed a segmentation algorithm based on finger touch. By fusing edge, regional texture, and locally collected geometric information of contact points into an appearance model, only one finger touch can identify the object of interest in the image [18]. Zhao and Kit designed a regularly sampled space-time bilateral grid to minimize long-term space-time connections between pixels [19]. Some methods segment moving objects in videos by building dense or sparse trajectories using probabilistic models. Liang et al. generated a fixed-size window with the current pixel as the center and extracted the lab color features within this window [20]. Then, this feature is compared locally with the features in other nearby windows to obtain the saliency calculation result, and the saliency value at multiple scales is combined to obtain the initialization result of the saliency target. Yilmaz et al. implemented a spatiotemporal video segmentation algorithm by combining long-term motion cues from past and future frames [21].

This paper studies the application of the unsupervised VOS algorithm in the 3D model modeling of VR technology. In this paper, an unsupervised VOS algorithm is proposed, which initializes the target by combining the moving edge of the target image and the appearance edge of the target and assists the modeling of VR 3D model, which has reference significance for the future construction of large-scale VR scenes.

## 3. Methodology

*3.1. 3D Model Modeling.* An important factor in VR system is the modeling of virtual world [22, 23]. The modeling process of VR is generally divided into the following steps: (1) describe the shape and appearance of virtual objects through geometric modeling. (2) Determine the position of 3D objects in the world coordinate system and their movement in the virtual world through motion modeling. (3) Physical modeling, which comprehensively reflects the physical characteristics of the object, including weight, inertia, and surface hardness. (4) For a large-scale simulation environment, it is necessary to model the behavior of some objects that users cannot control. The concrete process of 3D modeling operation includes data acquisition, data preprocessing, structure optimization, model creation, model optimization, scene optimization, scene integration, and scheduling management, and its application scope is wide [24]. Data preprocessing of 3D modeling is based on information collection. In the process of data collection, it is required to strictly

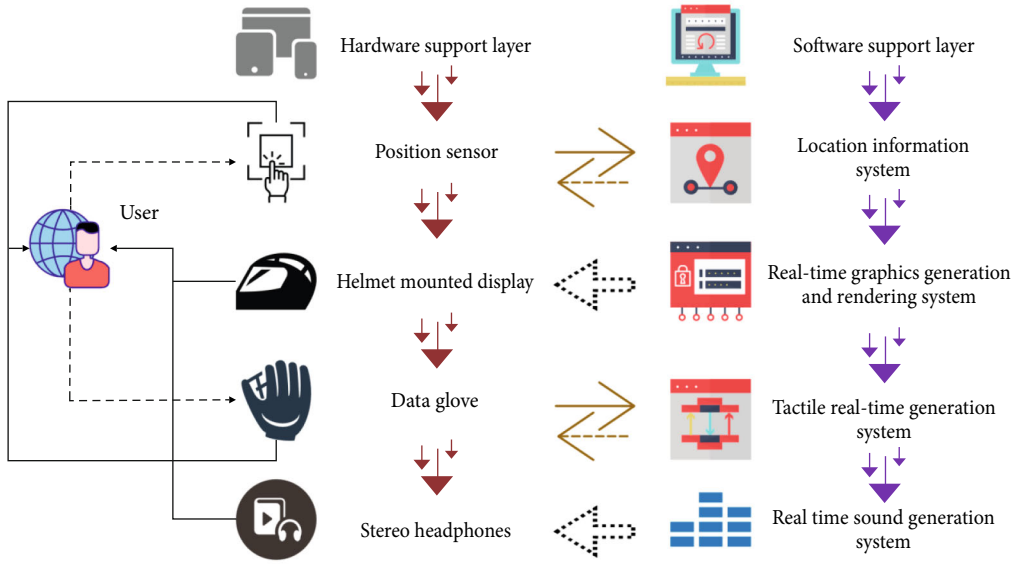


Figure 1: Architecture diagram of virtual environment.

follow the data collection specifications according to the operation process. The data acquisition process includes point control measurement, scanning station layout, spherical target layout, point cloud data scanning and acquisition, field data inspection and analysis, data export, and backup. Some filtering algorithms need to be used to filter out the point cloud data and discrete points of occlusion objects such as background during field operation and retain the main point cloud data of the object. The virtual environment architecture is shown in Figure 1.

The functions of the virtual environment framework are as follows: firstly, it supports the subscription and publication of object attributes and interactions; at the same time, it provides callback and event notification mechanisms and supports various time management strategies of HLA. Secondly, it provides services related to locating, creating and deleting various objects, and organizes and manages simulation entities with different functions and properties in a classified and unified way. Image-based modeling technology refers to the direct use of camera devices to collect discrete images of objects and other basic research materials for data processing; then, the panoramic image is generated by the combination and evolution of image processing software. Then, the panoramic image is further processed to the adaptive space model, and the VR real space is made. Because it is necessary to run 3D models in real time, its modeling method is very different from modeling-based modeling, and most of them use other techniques instead of increasing the complexity of geometric modeling to improve the fidelity [25]. There are three modeling methods for the VR system, which are mainly distinguished according to the construction methods of virtual scene: model-based rendering method, image-based rendering method, and mixed modeling method based on graphics and images. 3D graphics modeling technology mainly studies the generation and representation of 3D object information in the computer. Models describing 3D object information include geometric model, illumination model, and color model. In virtual reality hybrid modeling, users can enter the virtual scene in the form of virtual entity objects. Although the user avatar cannot interact with it, people can still obtain the depth information of the user avatar relative to the pure virtual object in the image by using binocular stereo vision technology and helmet mounted display. Because users expect that the scene objects that interact with them must be geometric model entities, hybrid modeling is required. In addition, in order to meet the visual reality, geometric model entities must be assigned with surface texture and material attributes. However, in hybrid modeling, it is difficult for user avatars to interact with virtual environment image objects established by the BMR method. The simulation requires the integration of geometric entity object and virtual environment image object, at least in vision. Although the distance or gap between virtual environment image objects can be perceived through the depth information of virtual environment image objects, however, how to smooth geometric entity objects into such space gaps remains to be solved. The mixed modeling technology based on graphics and images can integrate the advantages of both and make the best use of their strengths and avoid their weaknesses in application [26]. This not only increases the realism of the scene but also ensures real-time and interactivity and improves the immersion of users. In 3D modeling technology, there are often many problems that affect the authenticity of modeling. Therefore, in modeling technology, in order to improve the fidelity of display, the following methods are often used: blanking, shading model, and texture mapping.

*3.2. Application of Unsupervised VOS Algorithm in 3D Model Modeling.* Optimization technology is a crucial link in the process of 3D modeling. Usually, the basic plan is drawn by CAD using the position in the drawing, and then the plan is fully imported into the 3DMax construction model. In the process of making the model, the basic frame structure must

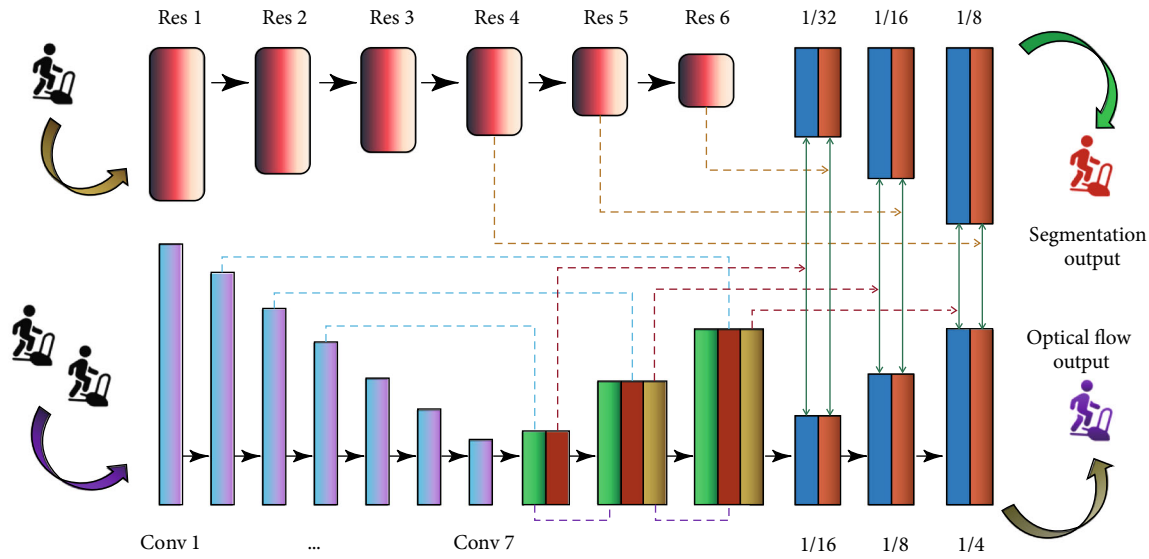


Figure 2: Network structure diagram.

Table 1: Comparative experiment of motion saliency segmentation network on DAVIS.

| Method | Mean |
|---|---|
| Mp-net-motion | 55.36 |
| Uovos-motion | 56.91 |
| Fseg-motion | 68.25 |
| Epo-motion | 76.58 |
| Methods of this paper | 80.36 |

be drawn first, and then the complete structure model can be drawn using the previous external contour. Then, each relevant model is effectively spliced, which will better optimize the overall structure of the model. Effectively map the structure after modeling. In the process of mapping, complete mapping must be carried out according to the specific structure size of the model. At the same time, different mapping scales are required for different precision models. We should deal with it effectively according to the real effect, so as to reflect the authenticity. This scene simulation system is an improvement of the traditional optimization technology, and the optimization technology used runs through the whole modeling process. The reality of an object's appearance mainly depends on its surface reflection and texture. Today's graphics hardware platform has the ability of real-time texture processing, which can enhance the sense of reality with a small amount of polygons and textures while maintaining the graphics speed. Texture can be generated by two methods, one is to interactively create, edit, and store texture bitmaps by image rendering software; the other is to take a picture of the required texture, then scan it, or take a picture directly with a digital camera. First of all, it is necessary to determine which space plane the surface patch projects on, which depends on the overall direction of the surface patch, and the plane with the smallest angle will be projected to which plane. Considering the convenience, when deciding the position of the target point, this paper calculates the error costs of two endpoints, respectively, takes out the one with smaller error, compresses it to the position of the other endpoint, and deletes the degraded triangle at the same time.

The object of segmentation is to detect the moving object. The simplest method of mask fusion is to calculate the intersection area of salient motion mask and general target mask, so as to satisfy the characteristics of motion and general target at the same time, but its accuracy is low. In order to make full use of the mask results of motion detection and target sampling, this paper adopts the method of deep learning and constructs a small fusion network to fuse the masks of the two. In unsupervised VOS, effective and full use of motion cues is crucial to segmentation performance. As a mainstream method of timing information modeling, optical flow can simulate the moving trend of the target according to the displacement changes of pixels in adjacent frames. The network structure composed of an appearance segmentation network and an optical flow prediction network is shown in Figure 2.

In the aspect of target image edge extraction, this paper obtains the moving edge of the target through the difference of the size of the motion and the direction of the motion. The specific description is as follows: first, calculate the optical flow vector value between two adjacent frames, and through formula (1), calculate the motion size $b_p^m$ of each pixel point $p$:

$$b_p^m = 1 - \exp\left(-\lambda^m \left\| \nabla \vec{f}_p \right\|\right). \tag{1}$$

In the formula, $b_p^m \in [0, 1]$ is the motion size of the pixel point $p$, $\vec{f}_p$ is the optical flow vector value of the pixel point

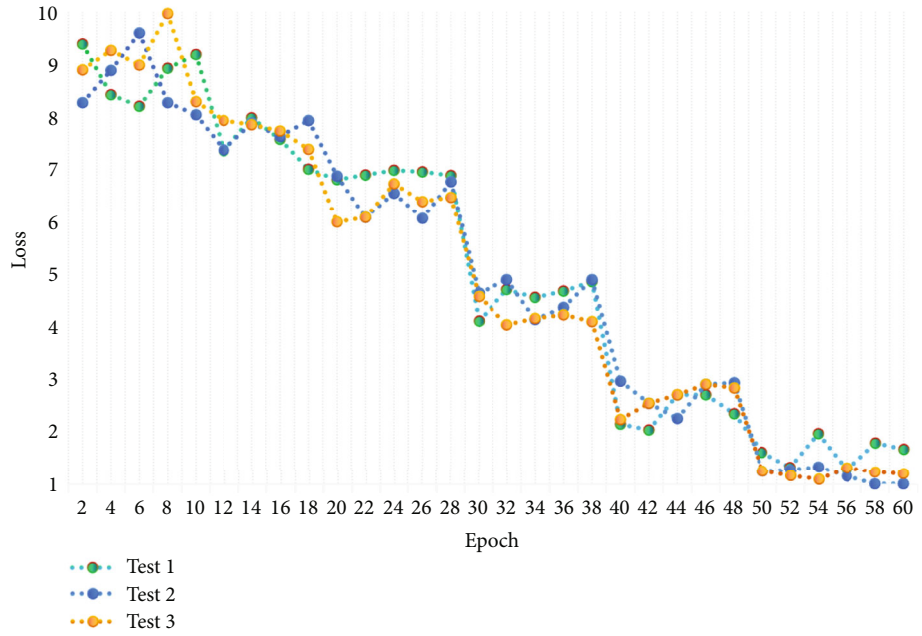


Figure 3: Training of the algorithm.

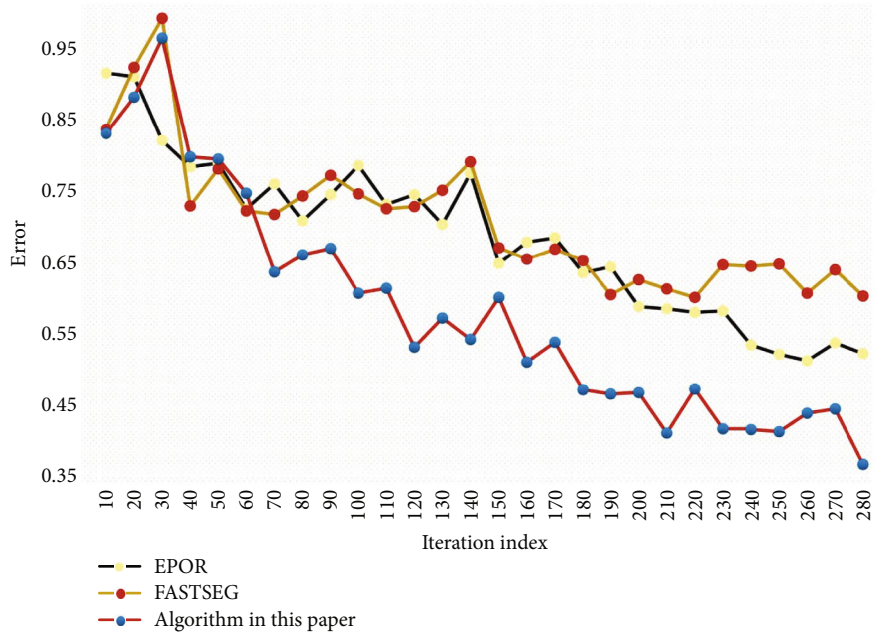


Figure 4: Error of the algorithm.

$p$, $\nabla$ is the gradient value, and $\lambda^m$ is the weight parameter. However, due to the shaking of the camera or following the target, the background will move violently. Therefore, this paper considers the use of the angle between the motion vectors to distinguish the target and the background, such as formula (2), to obtain the motion edge size $b_p^\theta$:

$$b_p^\theta = 1 - \exp\left(-\lambda^\theta \max_{q \in N}\left(\delta\theta_{p,q}^2\right)\right). \tag{2}$$

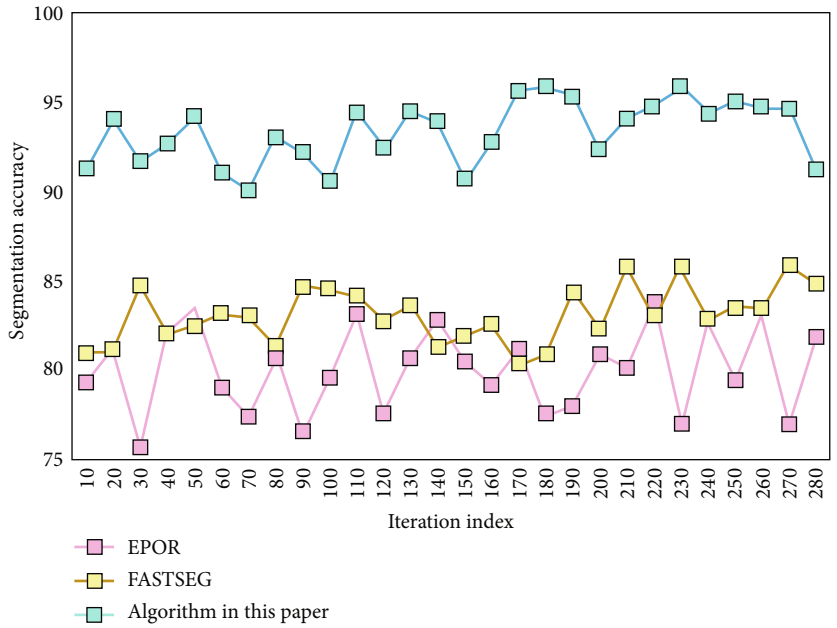


Figure 5: Segmentation accuracy of the algorithm.

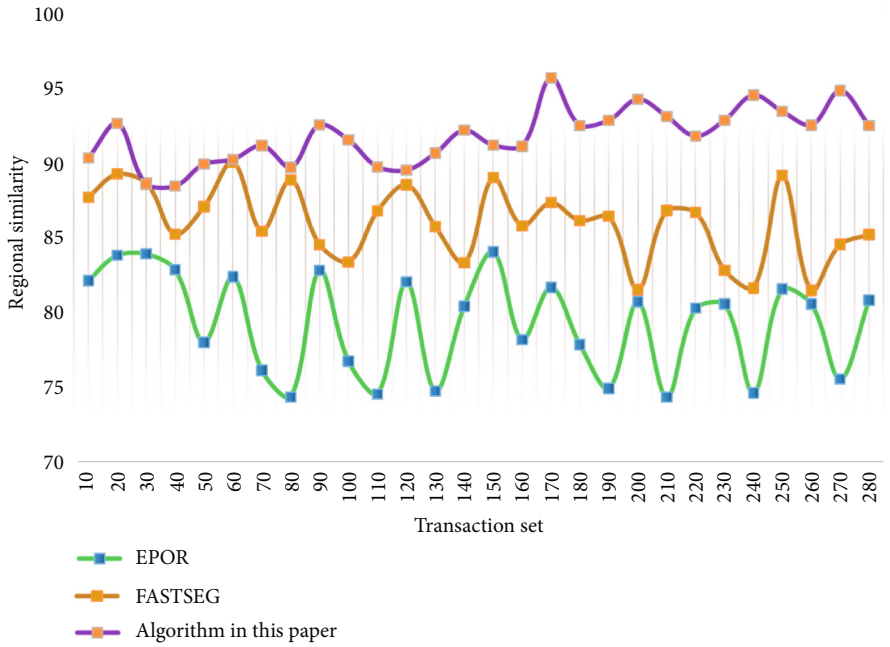


Figure 6: Experimental results of regional similarity.

In the formula, $b_p^\theta \in [0, 1]$ is the maximum angle distance between the pixel point $p$ and the surrounding pixels and $\delta\theta_{p,q}$ is the angle size of the motion vectors $\vec{f}_p$ and $\vec{f}_q$ of the pixel points dd $p$ and $q$. At the same time, the motion edge feature $b_p$ of the target is obtained by combining the motion size and direction of the pixel, and the motion edge of the target can be obtained. The formula is as follows:

$$b^p = \begin{cases} b_p^m, & \text{if } b_p^m > T, \\ b_p^m \cdot b_p^\theta, & \text{if } b_p^m > T. \end{cases} \tag{3}$$

In the formula, $T$ is the size of the threshold, and $b_p^m$ and

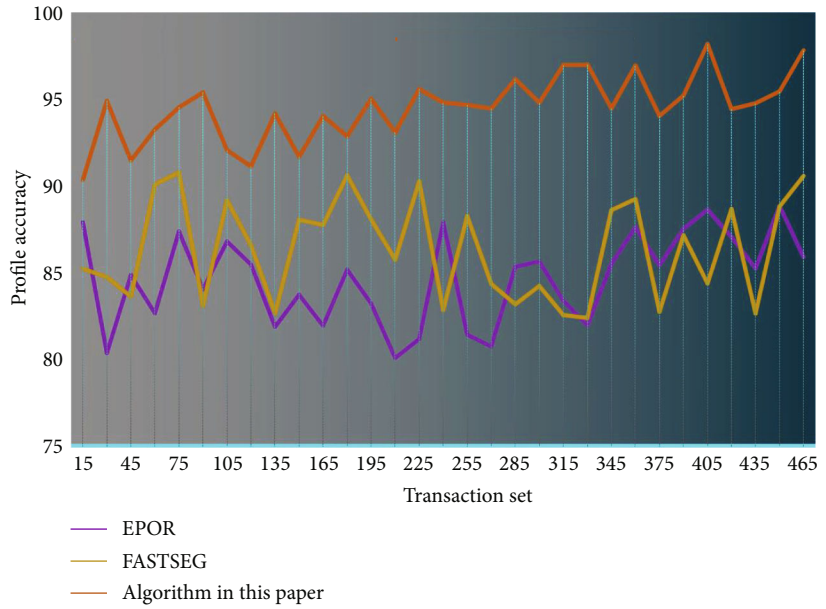


Figure 7: Experimental results of contour accuracy.

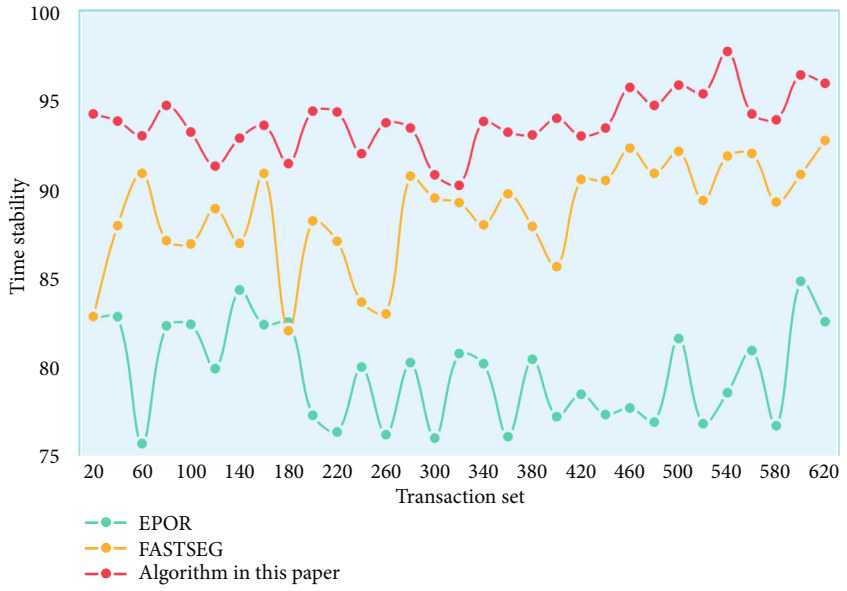


Figure 8: Experimental results of time stability.

$b_p^\theta$ represent the edge value obtained by the magnitude of the motion amplitude and the angle of the motion direction, respectively.

This paper associates each vertex with a set of planes near it, and the error of the vertex is expressed as the sum of the squares of the distances from this point to each of the planes in the set. When two vertices are compressed into a single point, the relevant plane group of the target point is the sum of the two groups of planes of the original point. Each plane can be written as the following equation:

$$n^T v + d = 0, \tag{4}$$

where $n = [n_x, n_y, n_z]^T$ is the normal vector of the plane, and $d$ is a constant. Then, the square of the distance from point $\mathrm{v} = [\mathrm{x}, \mathrm{y}, \mathrm{z}]^{\mathrm{T}}$ to the plane is

$$D^2 = \left(n^T v + d\right)^2 = \left(n^T v + d\right)\left(n^T v + d\right) = v^T\left(nn^T\right)v + 2dn^T v + d^2. \tag{5}$$

This is a quadratic, and let

$$Q = (A, B, C) = \left(nn^T, dn, d^2\right), \tag{6}$$

$$Q(v) = v^T A v + 2B^T v + C, \tag{7}$$

Table 2: Experimental results of indicators.

| Index | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Regional similarity | 0.959 | 0.951 | 1.971 | 0.968 |
| Accuracy of contour | 0.984 | 0.988 | 0.975 | 0.993 |
| Time stability | 0.989 | 0.991 | 0.987 | 0.983 |

Table 3: Accuracy comparison results of different segmentation methods on data sets.

| Method | Vehicle | Pedestrian | Horse | Plane |
|---|---|---|---|---|
| VOE | 5.2*E*+03 | 6.2*E*+03 | 4.2*E*+04 | 1.2*E*+03 |
| EPOR | 5.3*E*+03 | 5.0*E*+04 | 3.2*E*+04 | 7.2*E*+04 |
| RCC | 5.5*E*+03 | 5.5*E*+04 | 3.5*E*+04 | 4.5*E*+04 |
| VIBE | 6.2*E*+03 | 7.5*E*+03 | 3.2*E*+04 | 4.2*E*+04 |
| FASTSEG | 2.2*E*+04 | 6.1*E*+04 | 3.1*E*+04 | 4.6*E*+04 |
| Methods of this paper | 3.2*E*+03 | 5.8*E*+04 | 7.2*E*+03 | 1.2*E*+04 |

because

$$Q_1 + Q_2 = (A_1 + A_2, B_1 + B_2, C_1 + C_2). \tag{8}$$

So,

$$Q_1(v) + Q_2(v) = (Q_1 + Q_2)(v). \tag{9}$$

Therefore, to calculate the sum of the squares of the distances from a vertex to a set of planes, this article only needs to add all the quadratic formulas and finally gets a quadratic formula. After the two vertices are compressed into one vertex, the corresponding quadratic formula is also the sum of the quadratic formulas of the original two points. Therefore, the error of an edge-compression operation $(v_1, v_2) \longrightarrow v$ can be defined by the following formula:

$$Q(v) = Q_1(v) + Q_2(v) = (Q_1 + Q_2)(v). \tag{10}$$

Let $P^t = \{P_1^t, P_2^t, \cdots, P_N^t\}$ be the backward optical flow field between two frames $F^t$ and $F^{t-1}$. where each element $P_i^t = [u_i^t, v_i^t]$ is the optical flow vector of pixel $F_i^t$ in the horizontal and vertical directions; $N$ is the total number of pixels in the frame. Let $\tilde{S}^t$ be the salient motion map in the optical flow field $P^t$, and the global motion contrast $\tilde{S}^t{}_i$ can be expressed as the following formula:

$$\tilde{S}^t{}_i\left(P^t\right) = \sum_{\forall P_j^t \in P^t} d\left(P_i^t, P_j^t\right). \tag{11}$$

Among them, $\tilde{S}^t{}_i \in [0, 1]$ and $d(\cdot)$ are distance measures. Let $\phi$ be the binary segmentation function of the adaptive threshold method, an then the salient motion mask $S^t$ is shown in the following formula:

$$S^t = \phi\left(\tilde{S}^t\right), \tag{12}$$

where each element $S_i^t \in \{0, 1\}$ represents the binary foreground-background label of pixel $F_i^t$.

In this paper, a skip connection is used to connect the original features, which can preserve the predicted parts of the optical flow features in other directions without affecting the common salient regions. Similarly, this paper also performs similar operations on the backward optical flow feature and the saliency map generated by the forward optical flow. The operation of the entire structure is symmetrical. The overall process is described as follows:

$$f_{\text{bm}} = \sigma\left(\theta_1\left(f_f\right)\right), \tag{13}$$

$$f_{\text{fm}} = \sigma(\theta_2(f_b)), \tag{14}$$

$$F_{\text{forward}} = f_f \times f_{\text{bm}} + f_f, \tag{15}$$

$$F_{\text{backward}} = f_b \times f_{\text{fm}} + f_b, \tag{16}$$

$$f_m = \sigma(\theta_3(\text{concat}(F_{\text{forward}}, F_{\text{backward}}))). \tag{17}$$

The two processed features are connected, and the final optimized motion saliency map $f_m$ is obtained through a 3 $\times$ 3 convolution and Sigmoid function. The function of the sigmoid function is to compress the element value between $[0, 1]$ and generate a probability saliency map. The larger the value, the greater the saliency probability of the position.

Because of error estimation, each vertex in the vertex table needs a unit to store the compression error of that point in addition to its three coordinate values. In addition, each record in the side table not only records the serial numbers of the two vertices of the side but also records the compression error of the side. In this paper, combined with the actual hardware and software conditions, using modern advanced technology, the collected data should be preliminarily processed, and some incorrect or redundant data should be removed. At the same time, the collected data can keep a relatively high precision, which can meet the requirements of the system to the greatest extent. Combined with the collected data, the data is preprocessed. Considering that color, material, etc. should be processed in the development of the system, and 3DS is a very common data format, and the 3D graphics files saved in this format are also very rich, so this system uses 3DS data structure to convert data. As for the conversion of data format, the main core is to convert the data of model construction and form the list of model construction of 3DS, which can be used as the display list of OpenGL to reconstruct the model. Unfortunately, not all cases apply to parametric surfaces. There are some situations that require adjacent surfaces to fit together well (no cracks or T-joints) when an object is rendered into a polygon. Also, there are many jagged objects that cannot achieve good results using parametric surfaces, because the number of surfaces required may not be less than the number of polygons. The polygon-based face reduction method is generally more useful and can work on the current type of model.

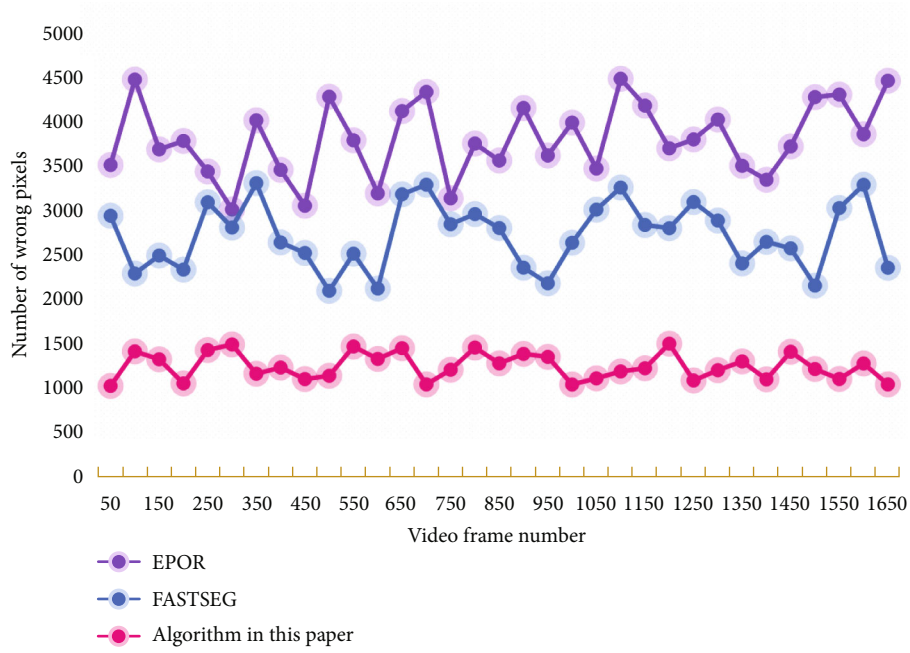


Figure 9: Comparison chart of the number of wrong pixels per frame.

## 4. Result Analysis and Discussion

In this section, the algorithm proposed in this paper is verified by experiments. Firstly, the data set used for the evaluation method and the corresponding evaluation indexes are introduced, and the corresponding modules proposed in this paper are tested on the specified data set. At the same time, the experimental results compared with other algorithms are introduced, and then the concrete discussion and analysis are carried out according to the experimental results. In order to further verify the effectiveness and practicability of this algorithm, several data sets in the experiment are YOUTUBE-OBJECTS public data set, DAVIS data set, and self-collected video set. Among them, a shot from the videos of airplane, horse, and motorcycle on DAVIS data set of YOUTUBE-OBJECTS public data set and the videos of pedestrians and two pedestrians in the surveillance scene taken by myself are selected. DAVIS data set is a large-scale video single-target segmentation data set, which contains 50 video sequences, including 30 video sequences in the training set and 3450 video frames in the test set, each with pixel-level labeling information. The dataset contains various challenges of target segmentation, such as scale change, fast motion, object occlusion, dynamic background, and motion blur. For the evaluation of experimental results, this section uses three evaluation indexes defined in DAVIS: regional similarity, contour accuracy, and time series stability.

Since salient object detection was introduced into the VOS field, many algorithms have directly applied salient object detection in the optical flow field and used the results of salient motion detection to perform the VOS task. Aiming at the motion saliency segmentation network, this paper mainly discusses the rationality of bidirectional optical flow and the effectiveness of the motion cue optimization module. Table 1 shows the comparative experimental results of motion saliency segmentation network on DAVIS.

It can be seen that the segmentation result of the dynamic cue optimization module on each video sequence should be completely superior to the segmentation result using only unidirectional optical flow. This fully proves the rationality and effectiveness of introducing bidirectional optical flow in this paper. The training of the algorithm is shown in Figure 3.

This paper uses the unsupervised VOS algorithm based on PyTorch to achieve target detection and segmentation. In order to verify the effectiveness of the unsupervised VOS algorithm proposed in this paper, its segmentation accuracy is compared and analyzed in the experiment, and the composition analysis is given. The experimental data in this experiment is YOUTUBE-OBJECTS, a public data set, and the collected video data. The comparison method includes the current mainstream target segmentation algorithms. Figure 4 shows the error of the algorithm. Figure 5 shows the segmentation accuracy of the algorithm.

It can be seen that the segmentation accuracy of this method is higher than that of the contrast algorithm. This is because there is often a large amount of background information in the low-level features, and the background information is further amplified by fusing the low-level features in the two branches. However, for VOS, too much background information is not conducive to the segmentation network's learning of the target area, and it will cause the segmentation network to misunderstand the background area and identify it as the foreground target area, thus greatly reducing the segmentation accuracy. In this paper, the effectiveness of deep semantic information fusion can effectively improve the segmentation accuracy of the target.

In the evaluation index, regional similarity is the intersection ratio between mask and true value. Contour accuracy divides the spatial range of the mask by treating the mask as a set of closed contours. Time stability is used to punish adverse effects such as boundary instability. In order to compare the effect of this algorithm with the current advanced algorithm, this paper uses the code and parameter settings provided by data set official website. The results of regional similarity experiment are shown in Figure 6. The results of contour experiment are shown in Figure 7. The experimental results of time stability are shown in Figure 8.

At present, most of the target segmentation algorithms use some underlying features of the target to initialize the target first and then accurately segment the target on this basis. Therefore, by fusing some underlying features such as the boundary features of the target, we can jointly model the target to further improve the segmentation accuracy and propose the corresponding fast solution algorithm to reduce the processing complexity. This paper conducted 20 experiments on each index and selected 4 of them to draw a table. The specific experimental results are shown in Table 2.

With the introduction of bidirectional optical flow in this paper, the result of motion segmentation is similar to that of truth mask, which can effectively suppress the nonsignificant regions and produce more accurate preestimation. In this paper, the performance improvement is attributed to the fact that the proposed motion cue optimization module can make full use of more motion information.

This paper collected five videos for experiments. It includes five videos: one pedestrian, two pedestrians, two people talking, running, and multiplayer football, and the target is manually marked. Table 3 shows the precision comparison results of different segmentation methods on data sets.

It can be seen that the segmentation result of this paper is obviously superior to other segmentation results.

In the algorithm, salient motion segmentation can segment the motion region, while target sampling can segment the target region. Significant motion segmentation and target sampling cannot separate moving targets in video frames; so, the purpose of fusion module is to remove potential noise, such as moving background and static targets in video. Only two-dimensional motion information cannot effectively solve the problem of motion blur, but this paper introduces 3D information, that is, the depth difference between foreground object and background area, which can effectively improve the accuracy of object segmentation and make the segmented object more detailed and complete. Figure 9 shows a comparison of the number of wrong pixels in each frame.

It can be clearly seen from the figure that the segmentation results of this paper are superior to those of other methods in most video frames. The segmentation accuracy of this algorithm can reach more than 94%, which is about 9% higher than that of FASTSEG method. On the whole, the accuracy of this algorithm exceeds that of other target segmentation algorithms.

## 5. Conclusions

VR modeling technology is developing rapidly, and it is welcomed by many users because of its ease of use, stability, and rapidity. At present, it has a wide application prospect in commerce, medicine, engineering design, art, entertainment, military, and so on. Based on this, this paper studies the process of 3D virtual scene construction by combining theory with practice, and on this basis, researches the optimization methods of 3D modeling. In this paper, an unsupervised VOS algorithm is proposed, which combines the moving edge of the target image and the appearance edge of the target to initialize the target and assist the VR 3D model modeling. The research shows that the segmentation accuracy of this algorithm can reach more than 94%, which is about 9% higher than that of the FASTSEG method. The segmentation results of this paper are superior to those of other methods in most video frames. At the same time, the accuracy of this algorithm exceeds that of other target segmentation algorithms. In this paper, it is of positive significance to use the unsupervised VOS algorithm to assist VR 3D model modeling. The next step will be to further improve the database management based on Web. Plan the system database reasonably and integrate different kinds and properties of data into the system database to the maximum extent. In addition, in order to make full use of the value of VR modeling, in the current social life, people from all walks of life should strengthen their own study and inquiry and strive to maximize the value of VR modeling in a reasonable system.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported by the Natural Science Foundation of Jiangxi Province: Research on Caching and Transmission Strategy in Content-Centered Wireless Networks with Non-Uniform Features (No. 20202BAB212003).

## References

[1] S. Piramanayagam, E. Saber, and N. D. Cahill, "Gradient-driven unsupervised video segmentation using deep learning techniques," *Journal of Electronic Imaging*, vol. 29, no. 1, p. 1, 2020.

[2] B. Martínez-González, J. M. Pardo, J. D. Echeverry-Correa, and R. San-Segundo, "Spatial features selection for unsupervised speaker segmentation and clustering," *Expert Systems with Applications*, vol. 73, no. 5, pp. 27–42, 2016.

[3] L. Li, W. Zhu, and H. Hu, "Multivisual animation character 3D model design method based on vr technology," *Complexity*, vol. 2021, Article ID 9988803, 12 pages, 2021.

[4] H. He, "Saliency and depth-based unsupervised object segmentation," *IET Image Processing*, vol. 10, no. 11, pp. 893–899, 2016.

[5] Y. M. Chen, I. V. Bajić, and P. Saeedi, "Moving region segmentation from compressed video using global motion estimation and Markov random Fields," *IEEE Transactions on Multimedia*, vol. 13, no. 3, pp. 421–431, 2011.

[6] J. E. Zosky, T. J. Vickery, K. A. Walter, and M. D. Dodd, "Object-based warping in three-dimensional environments," *Journal of Vision*, vol. 20, no. 6, p. 16, 2020.

[7] J. Dietlmeier, O. Ghita, H. Duessmann, J. H. M. Prehn, and P. F. Whelan, "Unsupervised mitochondria segmentation using recursive spectral clustering and adaptive similarity models," *Journal of Structural Biology*, vol. 184, no. 3, pp. 401–408, 2013.

[8] B. Krüger, A. Vögele, T. Willig, A. Yao, R. Klein, and A. Weber, "Efficient unsupervised temporal segmentation of motion data," *IEEE Transactions on Multimedia*, vol. 19, no. 4, pp. 797–812, 2017.

[9] H. Kim, J. Inoue, and T. Kasuya, "Unsupervised microstructure segmentation by mimicking metallurgists' approach to pattern recognition," *Scientific Reports*, vol. 10, no. 1, p. 17835, 2020.

[10] A. Soares Júnior, B. N. Moreno, V. C. Times, S. Matwin, and L. D. Cabral, "GRASP-UTS: an algorithm for unsupervised trajectory segmentation," *International Journal of Geographical Information Science*, vol. 29, no. 1, pp. 46–68, 2015.

[11] K. E. Ko and K. B. Sim, "Unsupervised stochastic segmentation of behaviour for learning by demonstration," *Electronics Letters*, vol. 52, no. 21, pp. 1767–1769, 2016.

[12] Q. Zhang, X. Song, X. Shao, R. Shibasaki, and H. Zhao, "Unsupervised skeleton extraction and motion capture from 3D deformable matching," *Neurocomputing*, vol. 100, no. 1, pp. 170–182, 2013.

[13] J. Yu, H. Di, and Z. Wei, "Unsupervised image segmentation via stacked Denoising auto-encoder and hierarchical patch indexing," *Signal Processing*, vol. 143, no. 2, pp. 346–353, 2017.

[14] T. Zhuo, Z. Cheng, P. Zhang, Y. Wong, and M. Kankanhalli, "Unsupervised online video object segmentation with motion property understanding," *IEEE Transactions on Image Processing*, vol. 29, no. 1, pp. 237–249, 2020.

[15] X. Cao, F. Wang, B. Zhang, H. Fu, and C. Li, "Unsupervised pixel-level video foreground object segmentation via shortest path algorithm," *Neurocomputing*, vol. 172, no. 1, pp. 235–243, 2016.

[16] P. C. Smith and B. K. Hamilton, "The effects of virtual reality simulation as a teaching strategy for skills preparation in nursing students," *Clinical Simulation in Nursing*, vol. 11, no. 1, pp. 52–58, 2015.

[17] M. H. Hung, C. H. Hsieh, C. M. Kuo, and J. S. Pan, "Generalized playfield segmentation of sport videos using color features," *Pattern Recognition Letters*, vol. 32, no. 7, pp. 987–1000, 2011.

[18] F. Liu, P. Chen, Y. Li et al., "Structural feature learning-based unsupervised semantic segmentation of synthetic aperture radar image," *Journal of Applied Remote Sensing*, vol. 13, no. 1, article 014501, 2019.

[19] H. Zhao and C. Kit, "Integrating unsupervised and supervised word segmentation: The role of goodness measures," *Information Sciences*, vol. 181, no. 1, pp. 163–183, 2011.

[20] D. Liang, B. Kang, X. Liu, P. Gao, X. Tan, and S.'. Kaneko, "Cross-scene foreground segmentation with supervised and unsupervised model communication," *Pattern Recognition*, vol. 117, no. 7, article 107995, 2021.

[21] R. M. Yilmaz, O. Baydas, T. Karakus, and Y. Goktas, "An examination of interactions in a three-dimensional virtual world," *Computers & Education*, vol. 88, no. 10, pp. 256–267, 2015.

[22] L. Zhi, L. Shen, and Z. Zhang, "Unsupervised image segmentation based on analysis of binary partition tree for salient object extraction," *Signal Processing*, vol. 91, no. 2, pp. 290–299, 2011.

[23] C. Chahine, C. Vachier-Lagorre, Y. Chenoune, R. el Berbari, Z. el Fawal, and E. Petit, "Information fusion for unsupervised image segmentation using stochastic watershed and Hessian matrix," *IET Image Processing*, vol. 12, no. 4, pp. 525–531, 2018.

[24] F. Tian, "Immersive 5G virtual reality visualization display system based on big-data digital city technology," *Mathematical Problems in Engineering*, vol. 2021, Article ID 6627631, 9 pages, 2021.

[25] F. Hao, Z. Jiang, and J. Shi, "Unsupervised texture segmentation based on latent topic assignment," *Journal of Electronic Imaging*, vol. 22, no. 1, p. 3026, 2013.

[26] J. Wang, H. Jiang, Y. Jia, X. S. Hua, C. Zhang, and L. Quan, "Regularized tree partitioning and its application to unsupervised image segmentation," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1909–1922, 2014.