

Research Article

The Path of Film and Television Animation Creation Using Virtual Reality Technology under the Artificial Intelligence

Xin Liu ¹ and Hua Pan ²

¹School of Art and Design, Lanzhou Jiaotong University, Lanzhou, China

²Lanzhou Echo Creative Culture Communication Co., Ltd., Lanzhou, China

Correspondence should be addressed to Xin Liu; marcliu@ljztu.edu.cn

Received 19 November 2021; Revised 9 December 2021; Accepted 27 December 2021; Published 13 January 2022

Academic Editor: Punit Gupta

Copyright © 2022 Xin Liu and Hua Pan. 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.

The purpose is to provide a more reliable human-computer interaction (HCI) guarantee for animation works under virtual reality (VR) technology. Inspired by artificial intelligence (AI) technology and based on the convolutional neural network—support vector machine (CNN-SVM), the differences between animation works under VR technology and traditional animation works are analyzed through a comprehensive analysis of VR technology. The CNN-SVM gesture recognition algorithm using the error correction strategy is designed based on HCI recognition. To have better recognition performance, the advantages of depth image and color image are combined, and the collected information is preprocessed including the relations between the times of image training iterations and the accuracy of different methods in the direction of the test set. After experiments, the maximum accuracy of the preprocessed image can reach 0.86 showing the necessity of image preprocessing. The recognition accuracy of the optimized CNN-SVM is compared with other algorithm models. Experiments show that the accuracy of the optimized CNN-SVM has an upward trend compared with the previous CNN-SVM, and the accuracy reaches 0.97. It proves that the designed algorithm can provide good technical support for VR animation, so that VR animation works can interact well with the audience. It is of great significance for the development of VR animation and the improvement of people's artistic life quality.

1. Introduction

Virtual reality (VR) technology is making continuous progress with the continuous development of science and technology, providing a new production method for animation creation [1]. The change of the new VR animation production mode leads to the corresponding change of the final work experience mode [2]. Among them, the most prominent is the audience's participation in the works, and the plot development of the works is closely related to the interaction of the audience. Human-computer interaction (HCI) technology under artificial intelligence (AI) needs to be further discussed to provide more possibilities for animation creation under VR technology [3].

In a VR environment, strong online perception ability and interactive feedback ability are needed, and gesture interaction is included in the abilities. Gesture interaction

generally includes static gesture recognition and dynamic gesture recognition. The static recognition has gradually changed from the artificial feature extraction method to the mainstream convolutional neural network (CNN) feature extraction method, which has a more efficient recognition ability. On this basis, scholars have proposed gesture recognition using a neural network as a classifier. This method is to use edge detection to obtain the gesture features and then recognize the gesture through the neural network, while its accuracy is not satisfactory [4, 5]. Therefore, scholars have introduced the contour descriptor based on the depth projection map after continuous exploration. It is generally used to obtain the hand shape and structure information in depth image. The recognition accuracy has been improved through support vector machine (SVM) classification. Gesture recognition is used more in interaction, so related research is very crucial. The existing gesture

recognition mostly uses CNN to build the relevant model of gesture recognition, which can greatly reduce the subjectivity and limitations caused by manual feature extraction. On this basis, the convolutional neural network—support vector machine (CNN-SVM) makes the model more robust [6]. The disadvantage is that the model has no relevant means to correct the wrong gesture when there is a recognition error.

In response to the shortcomings of previous algorithms, a new classification estimation error correction strategy is proposed based on the gesture recognition of CNN-SVM, and CNN-SVM is optimized to improve the final effect of the model. The innovation is to improve the recognition accuracy of similar gestures. The optimized CNN-SVM is designed to achieve the final effect of the optimization model. Later, the necessity of image preprocessing is discussed through experiments, and the recognition accuracy of the optimized CNN-SVM is compared, which proves that the accuracy of the designed gesture recognition algorithm is good enough. Thus, it provides reliable algorithm support for VR technology animation production and is of great significance to the development of art forms.

2. Materials and Methods

2.1. Differences between VR and Traditional Animation. VR technology is what people often call VR. This technology is a comprehensive new technology composed of various platforms established based on computer media [7]. Figure 1 shows its main technical basis.

The main function of VR is to create a simulated simulation environment to achieve very realistic effects like real life. In the process of realizing this environment [8], it is essential to build an image and sound in a three-dimensional (3D) space. Moreover, the simulated simulation environment also needs the ability of online perception and interactive feedback, such as vision, hearing, touch, and orientation. From the perspective of perception, it mainly records and analyzes people's relevant actions and other physical activity data, uses the computer to analyze the corresponding perception signals online, and transmits them to the perception equipment for people to perceive. The accuracy and timeliness of computer data processing are the most core in this process [9].

Making animation through VR technology is generally enriched at the perceptual level and added interactive mode. Animation has developed from hand drawing to computer drawing. Then, the 3D rendering of the image by computer has gradually matured, making the image have the level of 3D feeling and depth of field. The previous audience group of animation was relatively passive for the picture in animation, and there were almost no interaction and feedback with the relevant elements in the picture [10]. Present VR technology makes it possible for the groups watching animation to participate in the development of the plot. In daily life, there are various forms of interaction between people's subjective initiative and multiple elements in the real environment. People will also receive various forms of feedback in this process, that is, people's interaction in real life, which can

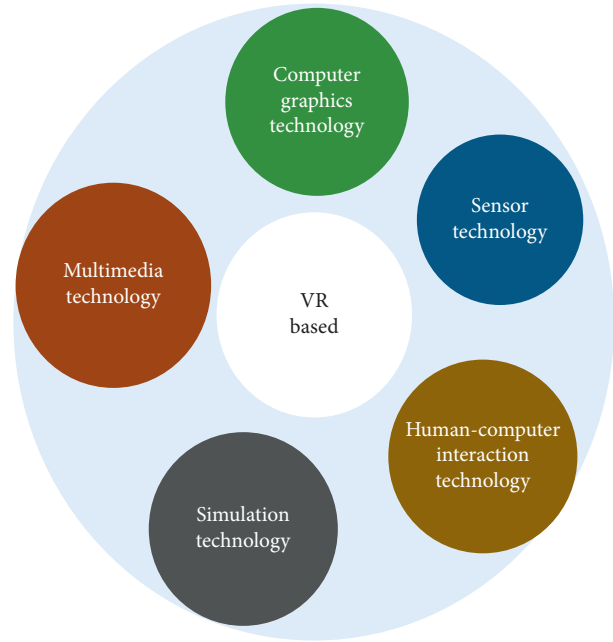


FIGURE 1: Technical basis of VR.

also exist in a VR environment. The previous animation had no interaction in any form, so there was no way for people's subjective behavior to change the plot in animation, and there was no feedback link in any form [11].

It should be noted that the emergence of VR technology does not mean that the traditional form of animation will disappear from people's vision. The form of animation will exist in diversified forms for a long time. There are two reasons for this. One is that traditional animation has established a relatively perfect theoretical system; the other is that the art shown by traditional animation has a unique beauty, which is difficult to disappear over time. Traditional animation may not have much freshness for viewers from the form of expression, but it still does not affect people's acceptance of it. More mature forms of expression and the addition of more high-quality content can still exert a great influence. A typical case is that many traditional animations launched in Japan have achieved strong influence in multiple countries. Different types of animation can meet people's different spiritual needs. The development research path of animation based on VR technology mainly includes the following three lines (Figure 2) [12].

The development path of VR animation is based on the above content. A brief analysis is given as follows:

- (1) The transformation from "plane" to "stereo" vision: the production of traditional animation is hand-drawn by relevant workers. It is to draw the motion track of the image very carefully, arrange it on the drawing paper in order, and then use the camera and other shooting tools to shoot in the corresponding order, followed by the rewashing work. Finally, the sample film is made. On this basis, fine editing work is conducted. In the animation production a long time ago, the animation workers first faced a piece of

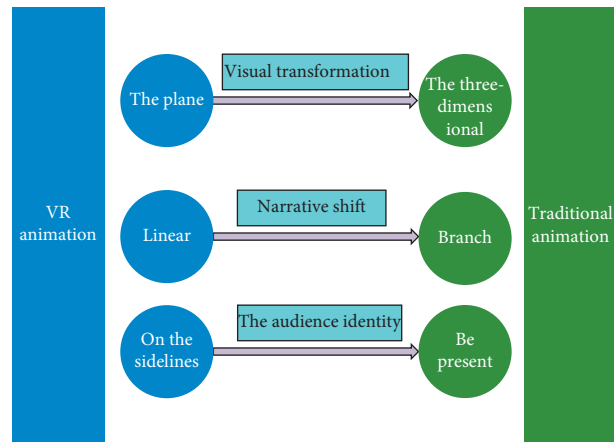


FIGURE 2: Research path of VR animation development.

white paper, and the core technology was the workers' painting skills. After the computer appeared, the computer monitor replaced the paper. The later 3D animation develops on this basis. However, it is still difficult to show a 3D feeling because it cannot get rid of the computer screen and cannot be said to be 3D. Later, this defect can be made up by wearing relevant equipment. 3D glasses are the most widely used equipment in daily life (Figure 3) [13].

However, the visual range is fixed, and the perceived "3D" has great limitations. The use of VR technology has greatly changed the previous creation methods, and the vision has changed from "plane" to "3D." The produced animation works are presented to the audience in a form without any dead corner, which gets rid of the previous screen and creates a 3D and realistic space. The whole animation production process is to use software to realize the whole process automation, greatly reducing the workload compared with the previous manual method [14].

- (2) Narrative transformation from "linear" to "branch": VR technology can reflect the feedback ability of animation, which is why it has "vitality." Figure 4 shows the change of narrative mode.

The continuous development and evolution of the whole plot of traditional animation are similar to the form of storytelling. It generally has a complete time-plot of the beginning, development, climax, and end of things, which is often referred to as the "linear" development structure. The audience is more passive to accept the whole story and has no impact on the development of the whole story. VR technology itself has the characteristics of interactivity, which has brought different narrative forms to the development of the whole story. A "branch" is added at a certain point in the story to make the structural change of a line become a "branch" structure. The audience chooses different branches according to their own preferences to make the story develop towards

different endings and exert an important impact on the development and change of the whole story [15].

- (3) The transformation from "watching" to "being present": in the past, when watching animation, the audience, as an independent individual, is a "bystander," just watches quietly, and has no relationship with any element in the animation. To make the audience feel immersive to the greatest extent, movie players will try their best to create a dark and quiet viewing environment during the screening. However, no matter how optimized the environment is, it cannot be denied that the audience is still a bystander [16]. VR technology provides a different viewing form from the past because the immersion characteristics of the technology itself will enable the audience to participate in the interaction in the animation from the first perspective, and interactive feedback is added in the animation production process, so that the audience can change from "watching" to "being present."

2.2. Principle of SVM Classifier. The lifelike effect of interactive feedback is very dependent on the development of HCI technology. Among them, AI technology provides crucial technical support. The gesture recognition method studied is the CNN-SVM hybrid model. In this hybrid model, SVM uses different kernel functions to transform the samples that cannot be divided into low-dimensional input space into high-dimensional feature space, so that it can be linearly divided. Its theoretical basis is to minimize institutional risk, form the best hyperplane in the feature space, and obtain the structured information of data distribution, so as to reduce the requirements for data scale and data distribution [16] and reduce the error of independent test set. The effect of the SVM classifier is affirmed by many people. SVM is evolved from the optimal classification surface in the separable state. The optimal classification surface needs to classify accurately and maximize the classification interval. If the training sample $(\vec{x}_1, y_1), \dots, (\vec{x}_i, y_i), \vec{x}_i \in R^m, i = \{1, 2, \dots, l\}$ needs a



FIGURE 3: 3D glasses based on the principle of the binocular angle difference.

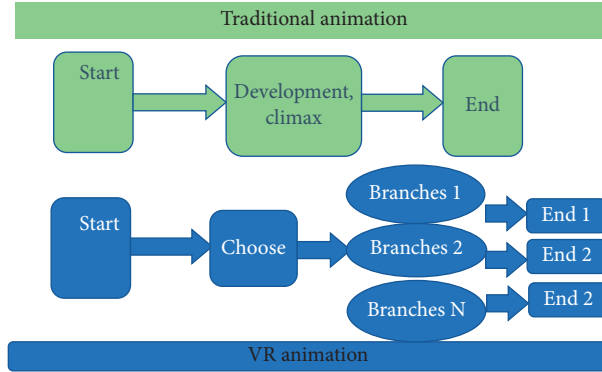


FIGURE 4: Change of VR animation narrative mode.

hyperplane with the largest interval so that the training set can be linearly separable (where \vec{x}_i is the eigenvector and y_i is the relevant label), the problem of finding a hyperplane can be transformed into the following problem:

$$\begin{aligned} \min : P(\vec{w}^T, b, \vec{\xi}) &= \frac{1}{2} \vec{w}^T \cdot C \sum_i \xi_i \\ \text{s.t} \left\{ \begin{array}{l} y_i (\vec{w}^T \phi(\vec{x}_i) + b) \geq 1 - \xi_i \\ \xi_i \geq 0, i = 1, 2, \dots, l, \end{array} \right. \end{aligned} \quad (1)$$

where \vec{w} is an m -dimensional vector, b is a scalar, and $\vec{\xi}$ is a relaxation variable. C is a penalty factor, which greatly affects the balance between edge maximization and classification error minimization. The training data \vec{x}_1 are mapped to a higher dimensional space by using function $\phi(\cdot)$ [17].

Chih-jen Lin developed a library for support vector machines (LIBSVM) that is used to build SVM. LIBSVM, as a software package for efficient classification and regression [18], can solve multiple classification problems. It uses a one-to-one method in the process of solving, which is to build

$k(k-1)/2$ classifiers. Each classifier uses two types selected from k -type data in the training set [19]. The problems to be solved read:

$$\begin{aligned} \min : P(\vec{w}^{ij}, b^{ij}, \vec{\xi}^{ij}) &= \frac{1}{2} (\vec{w}^{ij})^T \cdot \vec{w}^{ij} + C \sum_n \xi_n^{ij} \\ \text{st} \left\{ \begin{array}{l} (\vec{w}^{ij})^T \phi(\vec{x}_n) + b^{ij} \geq 1 - \xi_n^{ij} \quad y_n = i \\ (\vec{w}^{ij})^T \phi(\vec{x}_n) + b^{ij} \leq -1 + \xi_n^{ij} \quad y_n \neq i \\ \xi_n^{ij} \geq 0 \end{array} \right. \quad n = 1, \dots, k \frac{(k-1)}{2}, \end{aligned} \quad (2)$$

where i and j refer to type i training data and type j training data. When making classification decisions, LIBSVM uses the maximum winning algorithm. Each classifier will vote on the category it determines, and the final classification result will be qualified by the category with a higher number of votes. LIBSVM can classify the classification results and provide classification probability information for different test samples. SVM is used for classification results with

prediction probability during training. The probability of classification results will be used in gesture error correction to judge whether the classification results can be used [20].

2.3. Principle of CNN Classifier. CNN is a deep feedforward neural network, which generally has two parts: an automatic feature extractor and a trainable classifier [21]. It can make the deep CNN structure automatically obtain the high-level features of the image, so as to reduce the artificial design or selection of features. It transmits the obtained features to the classifier in the fully connected layer for classification processing. In this process, it will use supervised learning to optimize the weight between each layer in detail and later obtain a model with better robustness and accuracy [22].

Caffe framework is adopted to build CNN and identify and learn the model. Alex Krizhevsky network (AlexNet) is taken as the training network model. Figure 5 displays its network structure.

Figure 5 shows that the AlexNet network has 8 layers, including 5 convolution layers and 3 fully connected layers. The last fully connected layer outputs a 9-dimensional softmax to represent the prediction of 9 categories [23].

2.4. Error Correction Strategy Based on CNN-SVM Hybrid Model. CNN-SVM hybrid model is to replace the last output layer in CNN with SVM. The replacement process is as follows. First, the unprocessed image needs to be transmitted to the input layer for learning and training in CNN until convergence or the number of iterations is sufficient. Then, the images of the training samples are transmitted to the trained CNN training model, and 2048 dimensional training samples are obtained. The obtained sample feature vector is defined as the training set to train the SVM classifier, so as to obtain the CNN-SVM hybrid model [24].

In the following prediction results, LIBSVM estimates the probability that each sample is divided into a certain category. Finally, the most likely one will be selected as the classification conclusion. There is a problem of N classification in the decision-making process of LIBSVM. The similarity between the two classes is generally represented by distance. The absolute value of the prediction probability difference is used to represent the distance. The smaller the distance is, the smaller the gap between the two is. In the error correction strategy, the threshold M_{ij} represents the average distance between class i and class j . The threshold equation is as follows:

$$M_{i,j} = \frac{1}{s_{i,j}} \sum_{n=0}^{n=S_{i,j}} (P_n(i) - P_n(j)). \quad (3)$$

In equation (3), it should be noted that $P_n(i) < P_n(j)$. $P_n(i)$ is the probability of predicting that the n th test sample is class i , and it is the maximum value in the prediction results. $P_n(j)$ is the class j corresponding to the submaximum probability value; S_{ij} refers to the sample whose prediction result is class i , and it should meet that the submaximum value is class j . When the distance is lower than $M_{i,j}$, it shows

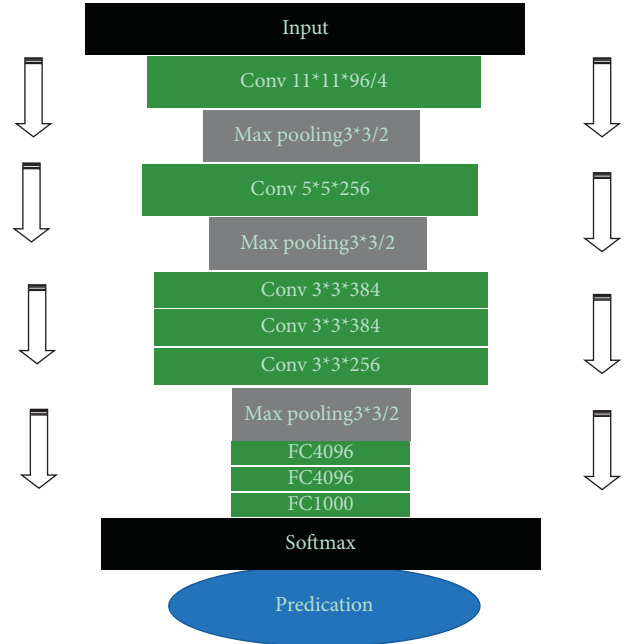


FIGURE 5: Model structure of AlexNet.

that classification error is probable between the two classes. When the classification prediction meets the following conditions, the category corresponding to the maximum value is changed to the category corresponding to the maximum value.

$$\text{s.t.} \begin{cases} W_n(i, j)M_{i,j} \\ \frac{1}{1 + e^{-p_{i,j}/w_n(i,j)}} \text{rand}(0, 1), \end{cases} \quad (4)$$

where $w_n(i, j)$ is the distance between the maximum value of possibility prediction and the submaximum value in the n th sample, that is, the value is $P_n(i) - P_n(j)$. $p_{i,j}$ is the probability that the predicted result is i , and the true rate is j . Small $w_n(i, j)$ and large $p_{i,j}$ indicate that the probability of errors in class i and class j will increase.

2.5. Data Preprocessing. Figure 6 shows the model running environment.

At present, multiple machine vision gesture databases are collected and obtained based on Kinect. Figure 7 shows the Kinect structure.

The data and images used in this experiment are the relevant gesture images of 400 college students' left hands 2 meters in front of Kinect, a total of 4000 depth images, and 4000 color images. The data need to be processed before use to improve the accuracy of the experiment.

Although the gestures made by people in color images can be recognized easily, it is still difficult to recognize them quite accurately. The reason is that gestures are affected by complex background conditions such as appearance and shape. The depth information in the depth image will not be disturbed by the environment such as light. Therefore, the

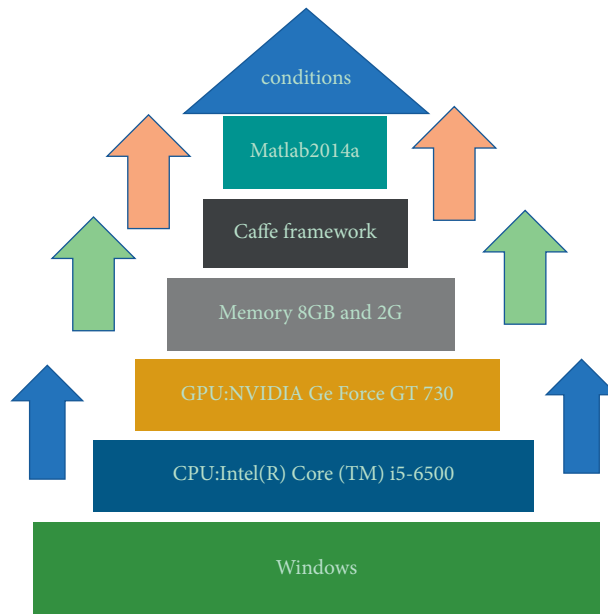


FIGURE 6: Model running environment.



FIGURE 7: Kinect structure.

depth image can well preserve the structural features of the human hand. The hand depth image is segmented, and then, the gesture range of the color image is segmented, so as to reduce the background interference of the color image. Figure 8 shows a specific preprocessing flow.

The main process of image preprocessing consists of three steps, which are explained in detail below. First, the collected depth image needs to be stored as a grayscale depth image with a pixel value of [0–255]. After the gray value 125 is used as the threshold, a binary image related to the gesture area will be obtained and then defined as a mask image. Then, the defined mask image and color image are calculated to obtain the gesture area image with low accuracy. The depth image and color image acquired by Kinect have the disadvantage of the inconsistent resolution, which will lead to the influence of other pixels near the acquired gesture area. Finally, the acquired gesture region needs skin color segmentation, and the final gesture region image is obtained through Bayesian skin color model. Figure 9 shows a depth gesture image after segmentation.

The interference of complex backgrounds and other environments is effectively eliminated. Later, the use of the Bayesian skin color model can also well preserve the useful information in the gesture area, so that the later recognition training has the support of data information.

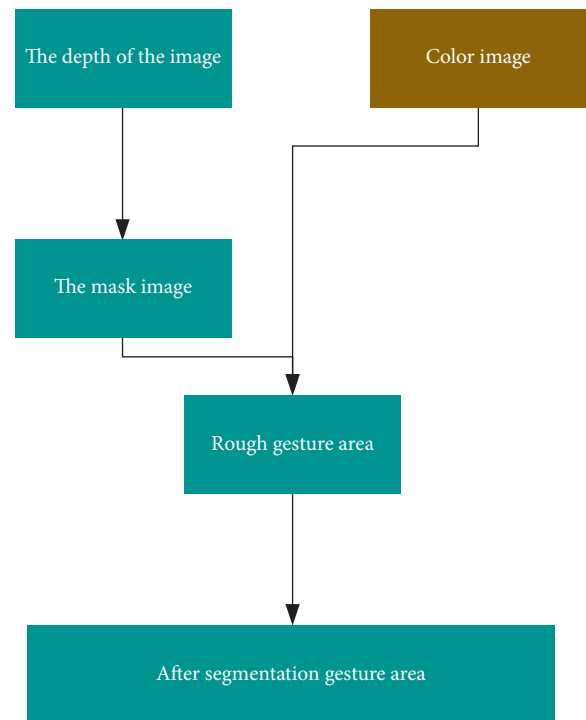


FIGURE 8: Preprocessing flow of image segmentation.

2.6. Case Analysis. In the following experimental link, the gesture image above will be segmented into 30000 images for the experimental dataset, among which 26000 images are used for model training and 4000 images are used for testing. On this basis, experiments will be conducted on the relationship between iteration times and the accuracy of different images. Besides, the accuracy of the designed optimized CNN-SVM recognition method, AlexNet, and CNN-SVM is tested, so as to evaluate their performance. The

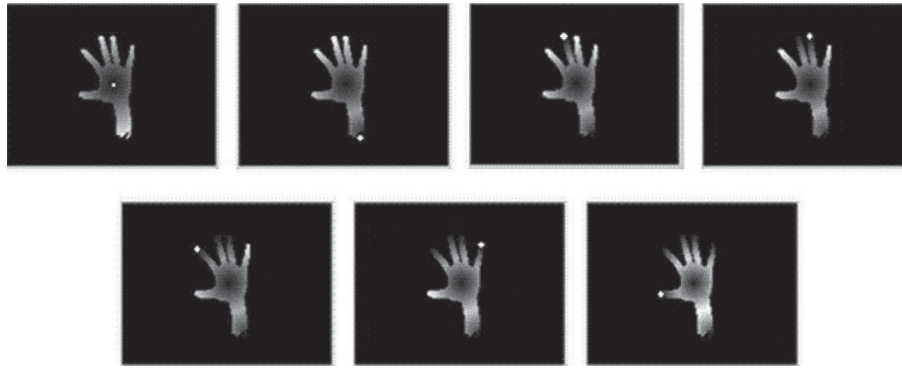


FIGURE 9: Depth gesture image after segmentation.

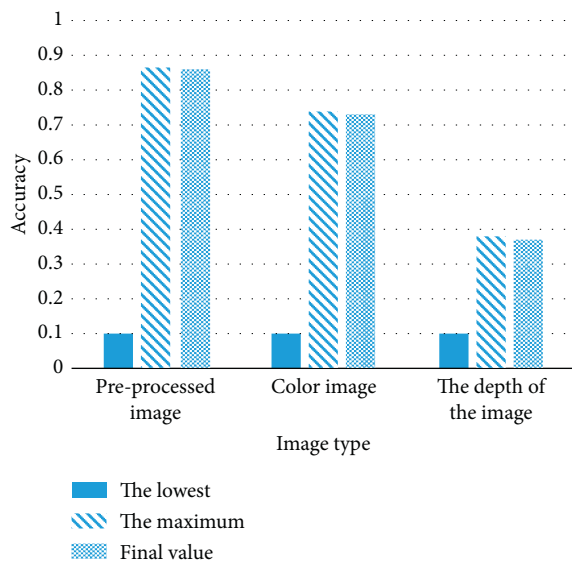


FIGURE 10: Accuracy of different images.

accuracy and performance of the algorithm are plotted by using Excel 2020 software.

3. Results

3.1. Relationship between Iteration Times and Accuracy of Different Image Training. The relationship between the number of iterations and accuracy of different images is obtained through unilateral CNN network training on 26000 processed images (the number of iterations is specified as 3000), as shown in Figure 10.

Figure 10 shows that when the number of iterations is 3000, the highest accuracy of the color image is only 0.37, the highest accuracy of the depth image is 0.73, and the highest accuracy of the preprocessed image is 0.86. Therefore, the gesture after preprocessing and segmentation can effectively reduce the influence of complex background and other interference factors, so that CNN network learning can obtain richer and more accurate features.

3.2. Accuracy of Different Methods on the Test Set. The accuracy of the optimized CNN-SVM, AlexNet, and CNN-

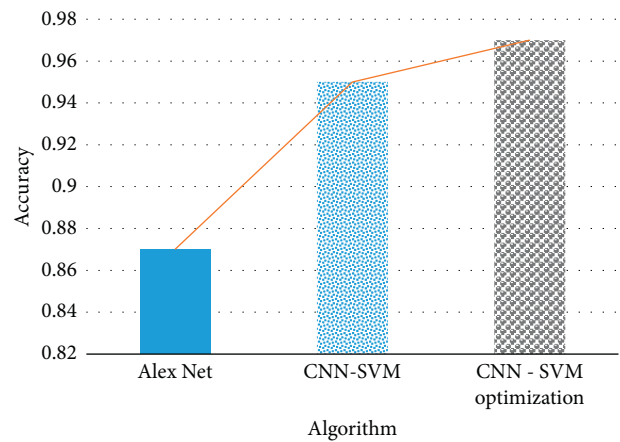


FIGURE 11: Accuracy of different methods on the test set.

SVM recognition methods is tested on the test set of 4000 images. Figure 11 shows the test results.

Figure 11 shows that the accuracy of the optimized CNN-SVM is higher than that of the original CNN-SVM. From the numerical comparison, the optimized CNN-SVM has a recognition accuracy of 0.97, which is better than the other two algorithms. It provides excellent algorithm support for HCI and provides more reliable technical conditions for VR animation creation.

4. Conclusion

With the continuous development of science and technology, VR is also making continuous progress. It provides a new production method for animation creation. The animation works based on VR need the support of a reliable human-computer interaction technology when they are in use. Inspired by AI, a CNN-SVM gesture recognition algorithm with the error correction strategy is proposed. Based on CNN-SVM, the differences between animation works under virtual reality technology and traditional animation works are analyzed based on VR, and the CNN-SVM gesture recognition algorithm based on the error correction strategy is discussed from the perspective of human-computer interaction. The advantages of depth images and color images are combined, and the collected information is preprocessed to make the algorithm have

better recognition performance. After experiments, the image preprocessing steps are summarized, and the recognition accuracy of the optimized CNN-SVM is compared with that of other algorithms. The comparison results show that the optimized algorithm is superior to others. It plays an important role in improving the ability of interactive feedback of VR and enhancing the interactive ability of films and television animation works. However, the size of the data is small, and it will be expanded in the follow-up study, making the research conclusions more convincing. The study promotes the development of VR and improves people's living standards.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] L.-H. Ho, H. Sun, and T.-H. Tsai, "Research on 3D painting in virtual reality to improve students' motivation of 3D animation learning," *Sustainability*, vol. 11, no. 6, p. 1605, 2019.
- [2] B. Gan, C. Zhang, Y. Chen, and Y. C. Chen, "Research on role modeling and behavior control of virtual reality animation interactive system in Internet of Things," *Journal of Real-Time Image Processing*, vol. 18, no. 4, pp. 1069–1083, 2021.
- [3] Q. Wang, C. Li, Z. Xie et al., "The development and application of virtual reality animation simulation technology: take gastroscopy simulation system as an example," *Pathology and Oncology Research*, vol. 26, no. 2, pp. 765–769, 2020.
- [4] A. Cannavò, C. Demartini, L. Morra, and F. Lamberti, "Immersive virtual reality-based interfaces for character animation," *IEEE Access*, vol. 7, Article ID 125463, 2019.
- [5] H. Liang, S. Deng, J. Chang, and J. J. Zhang, "Semantic framework for interactive animation generation and its application in virtual shadow play performance," *Virtual Reality*, vol. 22, no. 2, pp. 149–165, 2018.
- [6] L. Zhang, "Application research of automatic generation technology for 3D animation based on UE4 engine in marine animation," *Journal of Coastal Research*, vol. 93, no. SI, pp. 652–658, 2019.
- [7] W. Chujitarom and P. Piriyaawong, "Animation augmented reality book model (AAR book model) to enhance teamwork," *International Education Studies*, vol. 10, no. 7, pp. 59–64, 2017.
- [8] J. Lee, "A Study on effective directive technique of 3D animation in Virtual Reality-Focus on Interactive short using 3D Animation making of Unreal Engine," *Cartoon and Animation Studies*, vol. 6, pp. 1–29, 2017.
- [9] M. J. Blanchard, M. Reisch, and Y. V. Mejia, "Swing: 2D and 3D animation in virtual reality work-in-progress," *Frame*, vol. 2, no. 1, p. 12, 2020.
- [10] J. Qi, G. Jiang, G. Li, Y. Sun, and B. Tao, "Intelligent human-computer interaction based on surface EMG gesture recognition," *Ieee Access*, vol. 7, Article ID 61378, 2019.
- [11] B. K. Chakraborty, D. Sarma, M. K. Bhuyan, and K. F. MacDorman, "Review of constraints on vision-based gesture recognition for human-computer interaction," *IET Computer Vision*, vol. 12, no. 1, pp. 3–15, 2018.
- [12] A. Haria, A. Subramanian, N. Asokkumar, S. Poddar, and J. S. Nayaka, "Hand gesture recognition for human computer interaction," *Procedia Computer Science*, vol. 115, pp. 367–374, 2017.
- [13] J. Singha, A. Roy, and R. H. Laskar, "Dynamic hand gesture recognition using vision-based approach for human-computer interaction," *Neural Computing & Applications*, vol. 29, no. 4, pp. 1129–1141, 2018.
- [14] G. Li, H. Wu, G. Jiang, S. Xu, and H. Liu, "Dynamic gesture recognition in the internet of things," *IEEE Access*, vol. 7, Article ID 23713, 2018.
- [15] T. H. Tsai, C. C. Huang, and K. L. Zhang, "Design of hand gesture recognition system for human-computer interaction," *Multimedia Tools and Applications*, vol. 79, no. 9, pp. 5989–6007, 2020.
- [16] C. Linqin, C. Shuangjie, X. Min, and Y. Jimin, "Dynamic hand gesture recognition using RGB-D data for natural human-computer interaction," *Journal of Intelligent and Fuzzy Systems*, vol. 32, no. 5, pp. 3495–3507, 2017.
- [17] Y. Xu and Y. Dai, "Review of hand gesture recognition study and application," *Contemporary Engineering Sciences*, vol. 10, no. 8, pp. 375–384, 2017.
- [18] W. Lv, "Gesture recognition in somatosensory game via Kinect sensor," *Internet Technology Letters*, p. e311, 2021.
- [19] P. S. Negi, R. Pawar, and R. Lal, "Vision-based real-time human-computer interaction on hand gesture recognition," in *Micro-Electronics and Telecommunication Engineering, Lecture Notes in Networks and Systems*, D. K. Sharma, V. E. Balas, L. H. Son, R. Sharma, and K. Cengiz, Eds., vol. 106, Singapore, Springer, 2020.
- [20] S. Hazra and A. Santra, "Robust gesture recognition using millimetric-wave radar system," *IEEE sensors letters*, vol. 2, no. 4, pp. 1–4, 2018.
- [21] S. Ahmed, K. D. Kallu, S. Ahmed, and S. Cho, "Hand gestures recognition using radar sensors for human-computer-interaction: a review," *Remote Sensing*, vol. 13, no. 3, p. 527, 2021.
- [22] J. Li, J. Wang, and Z. Ju, "A novel hand gesture recognition based on high-level features," *International Journal of Humanoid Robotics*, vol. 15, no. 2, Article ID 1750022, 2018.
- [23] X. Yu and Y. Yuan, "Hand gesture recognition based on faster-RCNN deep learning," *Journal of Computers*, vol. 14, no. 2, pp. 101–110, 2019.
- [24] K. A. Smith, C. Csech, D. Murdoch, and G. Shaker, "Gesture recognition using mm-wave sensor for human-car interface," *IEEE sensors letters*, vol. 2, no. 2, pp. 1–4, 2018.