Application of VR Technology Based on Gesture Recognition in Animation-form Capture

Jing Yang[1], Hao Zhang[2]
Department of Information Engineering, Anhui Industry Polytechnic, Tongling, 244000, China[1]
School of Mathematics and Computer, Tongling University, Tongling, 244000, China[2]

***Abstract*—To accurately capture the posture of animation characters in virtual vision and optimize the user's experience when wearing virtual vision equipment, the hybrid Gaussian model has gained wide attention. However, various types of animation show an exponential growth trend, and the hybrid Gaussian model is prone to low-dimensional explosion when processing these single frames. Based on the mixed Gaussian model, this study conducts animation character gesture recognition experiments on the Disert data set to solve these problems. Meanwhile, it is improved by frame rate reduction method to generate fusion algorithm. In this paper, the video is first grayened and filtered, and the model feature points of the image are marked. Then the weight learning rate is introduced and added to the set of pixels, and then the peak signal-to-noise ratio of Wronsky function is adjusted by changing the parameters. Then similar image sets are extracted and the structure elements are opened and closed. Finally, the proposed algorithm is applied to Disert data set. Meanwhile, the prediction accuracy of PSO is tested and compared with fusion algorithm. A total of 400 experiments were conducted, and the prediction accuracy of the fusion algorithm reached 392 times, with an accuracy of 98.0%. The accuracy of PSO is close to that of fusion algorithm (88.2%). It is verified that the suggested model can identify the four common gestures of cartoon characters well, and users will get a good viewing experience.**

Keywords*—Frame rate reduction method; model feature points; Wronsky function; mixed Gaussian model; weight learning rate

I. Introduction

In a society with the rapid increase of animation works, the pose recognition model of animation characters has attracted wide attention, and the requirements of users have become higher and higher [1,2]. Character shape dynamic monitoring technology is to intercept the target from the played video, and then analyze the force status of the character's body in the frame, and then judge the current posture of the character. However, animation works have high requirements for clarity, that is, the frame rate contained in each video is rapidly increasing. The ability of the current algorithm to process fast-passing images is gradually falling behind, and the captured images cannot describe the video completely. In recent years, Hybrid Gaussian model (HG) has attracted the attention of many scholars due to its fast frame detection ability [3]. However, the suitable environment for HG processing pixels is low dimension, and the number of pixels that can be carried by the same dimension has an upper limit. HG is easy to cause data accumulation and even low-dimensional disaster in the working process. When a low-dimensional disaster occurs, it will not only destroy the two-dimensional configuration of the image, but also cause difficulties for subsequent information analysis. Most of the existing methods improve the ability of image frame detection, or make too much effort at the level of image clarity, thus ignoring the problem of information protection when analyzing information. In order to improve this situation, in this study, HG is optimized based on the Frame Rate Reduction (FRR), and the two are creatively fused to generate a fusion algorithm (HG-RPP). The algorithm builds a shared hyperplane, which can map linearly uncorrelated pixels to a higher plane and reduce the burden of low dimension. The steps of this study are divided into four parts. The first part mainly analyzes and summarizes the application effect of the current frame rate reduction model. The second part introduces the influencing factors of shape judgment and the construction of HG-FRR capture model. In the third part, the simulation test is carried out on the Disert data set. Finally, the performance of the optimization model and the traditional model is analyzed and compared, and the shortcomings of this study are pointed out. The aim of this study is to improve HG-RPP's tendency to fall into the low-dimensional dilemma, which not only has a significant effect on user happiness, but also on the sale of Virtual Reality (VR) products for merchants.

II. Related Works

A very important branch of animation character shape capture is animation character pose recognition, which plays a very important role in computer vision, users and businesses [4]. Liao et al. designed two terms based on supervised subspace learning and fragment sections to reduce the difference terms of in-class differences and to promote the block diagonal regularization terms of samples represented by in-class samples. The model maps the original samples to the learned subspace and classifies the mapped samples using a representation classifier with non-local constraints. Experiments have verified the good performance of the algorithm [5]. Yang et al. proposed a novel algorithm that can track multiple face gestures in a single frame or video. To create the first estimate of posture, the convolutional neural network model is first integrated with face identification and average shape learning. Then, a two-objective optimization strategy is used to iterate and apply algebraic filtering. Finally, experimental validation shows the benefits of the suggested approach [6]. Jain et al. proposed using deep learning to identify yoga poses and constructed a dataset with 10 positions from yoga. To capture the video, the study used a

smartphone camera. To recognize yoga poses in real time, a 3D convolutional neural network structure is designed and a supplementary layer, batch normalized average pool, is introduced. The experiment achieved a test recognition accuracy of 91.15% on the constructed data set. On publicly available data sets, competitive test recognition reached 99.39% [7]. Gunawan et al. proposed a deep learning architecture model that enables equivariant mapping of deep residuals to improve the attitude robustness of lightweight models as a solution to the underlying problem. The investigation assessed and contrasted the precision of several posture data sets, and the results showed that the suggested model improved accuracy by 0.07% and reduced verification time by 0.17 milliseconds [8].

With the development of the animation field, the clarity of the animation is gradually improved, and the number of attached pixels will also increase. The existing methods cannot continuously refine the pixel, so the algorithm based on pixel analysis is gradually studied. Zhang et al. suggested a deep learning model based on adversarial networks to synthesize faces and recognize faces with constant pose through the shape geometry of face images. The model generates a face that maintains identity by guiding the target pose. And the shape conveyed by the face separates identity from expression change. On controlled and field benchmark datasets, the proposed method performs well compared to the most advanced algorithms [9]. Yoganand et al. proposed a method to identify different gestures from video sequences. In this model, the video sequence is segmented into frames by the lens fragmentation process, and then from each shot, faces are recognised for subsequent processing. Facial features are extracted after the face is detected and the best frame in the video series is chosen using these characteristics. This approach is based on supervised learning and a bat algorithm-improved artificial neural network. Finally, a feature database is applied to recognise the facial picture [10]. Li et al. proposed an evaluation algorithm based on Demps Ter-Shafer Theory (DS). They believe that face detection can be achieved with low pixel monitoring. Therefore, DS is used to integrate the state of students' concentration, and the curve of students' concentration over time is obtained to describe the state of classroom concentration. On the data set provided by the computer camera, experiments have verified the efficiency and viability of the algorithm's design, and the accuracy rate of the algorithm is about 85.3% [11]. Chen et al. proposed a dimensionality reduction algorithm for principal component analysis. The algorithm uses the feature point screening function to filter the feature points extracted in advance by other algorithms, and projects the high-dimensional data into the low-dimensional space. Meanwhile, the redundant feature points are removed and the generation method of feature descriptors is changed to achieve the effect of dimensionality reduction. Experiments on open ORL face library have verified the superiority of the proposed model [12]. Zhu et al. found the importance of automated embedded systems, which work well in applications like private security and surveillance. Research suggests that smart door locks are prone to damage, increasing the risk. In order to provide users with effective open source software, the research proposes attitude tracking algorithms designed to ensure the security of modern keylock systems. Experimental results show that the established system has higher efficiency [13]. Maehama et al. developed an attitude recognition system that allows a robot to evaluate the content of aggressive words and determine whether the words are serious or joking. The study created a dataset of 16 participants. Different attitudes towards robots were observed in field experiments, which were applied to speech quality features of breath and pressed tone analysis, and combined with traditional prosodic features. Finally, the proposed method is compared with the speech standard. The final experimental results show that the combination of speech quality is superior to traditional neural networks in this task, with an accuracy equivalent to human guessing [14].

Through many international studies, it is found that the algorithm based on capturing the gesture of animated characters is very popular in the world, but the research on fusion algorithm is very few. In this study, we innovatively introduce feature point marking and bilateral filtering, and on this basis, take into account the impact of peak signal-to-noise ratio and shared hyperplane, and finally generate a fusion algorithm (HG-FRR). The purpose of this research is to improve the problem that HG-RPP is prone to fall into the low-dimensional dilemma, so as to enhance the user's sense of experience, and provide reference value for the relevant merchants' VR product sales strategy.

III. Modeling with the Mixed Gaussian Model of Wronsky Function Method

With the development of Virtual Reality (VR), more and more users are beginning to use VR technology. To cater to the preferences of users, the research on gesture recognition has gradually become a research hotspot, and the animation character shape capture technology is the most important part [15]. However, animation generally has objective factors such as high frame rate and complex character action, which increases the difficulty of shape capture equipment. In view of this, we combined the Frame Rate Reduction (FRR) method with the Hybrid Gaussian model (HG). This study first introduces the model based on FRR, and then describes the fusion method of the two in detail.

A. Establishment of Human Pose Feature Extraction Model based on Frame Rate Reduction Method

Before using VR to capture the shape of animation characters, the video image must be pre-processed. These include the method of turning gray to reduce the number of colors, and the method of bilateral filtering for video noise reduction. When the image color turns gray, the weighted average method is used in the study, as shown in Equation (1).

$$I = 0.3R + 0.5G + 0.11B \tag{1}$$

In Equation (1), I represents the color type after gray transfer treatment. The red, green and blue components are denoted as R, G, B respectively. Bilateral filtering is a nonlinear filtering method, which can use fixed templates to process adjacent pixels, and it can accurately distinguish spatial intensity differences [16]. The calculation of the bilateral filtering method is shown in Equation (2) below.

$$\begin{cases} g(x,y)=\dfrac{\sum_{k,l} f(k,l)w(i,j,k,l)}{\sum_{k,l} w(i,j,k,l)} \\ d(i,j,k,l)=\exp\left(-\dfrac{(i-k)^2+(j-l)^2}{2\sigma_d^2}\right) \\ r(i,j,k,l)=\exp\left(-\dfrac{\|f(i,j)-f(k,l)\|^2}{2\sigma_r^2}\right) \\ w(i,j,k,l)=d(i,j,k,l)-r(i,j,k,l) \end{cases} \tag{2}$$

In Equation (2), i, j, k, l, x, y represents pixels of the image. $f(k,l)$ represents the gray degree of the image before bilateral filtering, and is represented by $g(x,y)$ after filtering. The domain of pixel is denoted as $d(i,j,k,l)$. The range is represented by $r(i,j,k,l)$. The difference between the two is the weight coefficient of the image, denoted as $w(i,j,k,l)$. After graying and bilateral filtering, the image information will be missing, which is described by Equation (3).

$$\begin{cases} NMSE=\dfrac{\sum_{i=0}^{N}\sum_{j=0}^{N}\left[g(x,y)-f(k,l)\right]^2}{\sum_{i=0}^{N}\sum_{j=0}^{N} f(k,l)^2} \\ PSNR=10*\log\dfrac{MAX^2*M*N}{\sum_{i=0}^{N}\sum_{j=0}^{N}\left[g(x,y)-f(k,l)\right]^2} \end{cases} \tag{3}$$

In Equation (3), $NMSE$ is used to calculate the root-mean-square error before and after pixel conversion. Its signal-to-noise ratio is denoted as $PSNR$. Their values are positively correlated with algorithm performance. MAX is A constant, and the value is 255, which represents the gray value of the image. The gray method and bilateral filtering method are applied to all pixels, and the neighborhood window is finally obtained, as shown in Fig. 1.

In Fig. 1, in the panoramic region set I composed of several pixels, take any pixel point with coordinate (x,y), and scale the coordinate of the point in format $(x\pm1, y\pm1)$ to obtain nine neighborhood points, which together form the neighborhood window [17]. When the video sequence is clipped, it needs to be divided into several time nodes, and then the target image is obtained through the continuous changes between them, and then the shape capture is carried out, as shown in Equation (4).

$$d_t(x,y)=\left|F_t(x,y)-F_{t+nt}(x,y)\right| \tag{4}$$

In Equation (4), the image fragments of two consecutive time points are denoted as $F_t(x,y), F_{t+nt}(x,y)$. The difference between pixel points in frame rates is represented by $d_t(x,y)$. All image segmentation points have threshold values, and the segmentation of threshold values is as follows (5).

$$b_t(x,y)=\begin{cases} 255 & d_t(x,y)\geq T \\ 0 & otherwise \end{cases} \tag{5}$$

In Equation (5), the image capture effect of the character is denoted as $b_t(x,y)$. T represents the time interval for pre-selection. After using FRR to reduce the frame rate of the image, it can describe the outline of the object better. As a result, the structural elements can measure the image size faster, thus reducing the difficulty of video data. The structural elements can echo the corresponding form of the picture to make it compact, as shown in Equation (6).

$$\begin{cases} X\oplus B=\bigcup_{b\in B}(X)_B=\left\{Y:Y=x+b,x\in B,b\in B\right\} \\ X\otimes B=\bigcap_{b\in B}(X)_B=\left\{Y:(Y+b)\in B,b\in B\right\} \end{cases} \tag{6}$$

In Equation (6), the structural element is denoted as B, and there are two operations of expansion and corrosion between it and image set X. The expansion operation is represented by $\oplus$, which is a complement to the incomplete part of the image. The corrosion operation is denoted as $\otimes$, and the marginal parts of the image can be clipped. Because the human body has a certain degree of softness, the posture presented is complex and changeable, so the ability of the algorithm has certain requirements. Up to now, the accuracy of monocular detection for 3D feature description is still lacking, far less than that of 2D. The two-dimensional description of human features is divided into two modules, including pre-processing module and feature extraction module, as shown in Fig. 2.

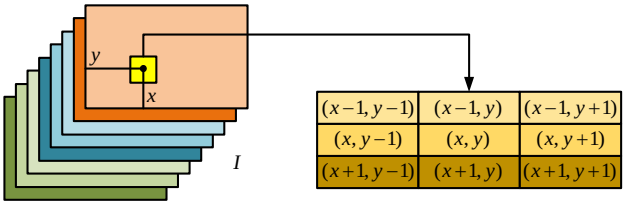

Fig. 1. The neighborhood window for pixels.

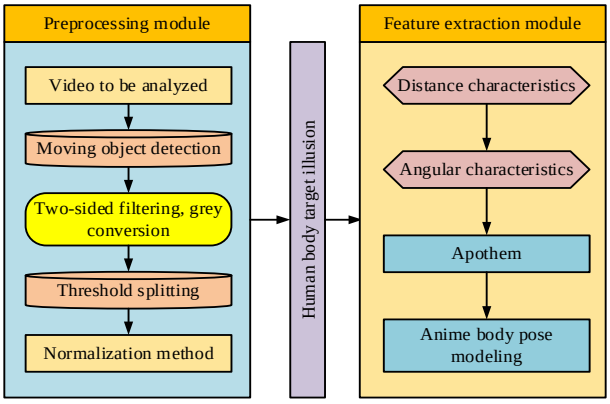

Fig. 2. Two-dimensional flow chart describing human characteristics.

In Fig. 2, the video to be tested is first monitored for moving targets in the pre-processing module, and noise reduction is carried out at the same time. Then the data set is segmented by bilateral filtering threshold, and the human body image containing the target is obtained after normalization, and input into the feature extraction module. Then the features such as distance and Angle in the image are extracted and their eccentricity is calculated. Finally, the model of human body posture can be output, and the Equation (7) can be established according to the feature points of the model.

$$Feature = \{D, A, e\} \tag{7}$$

In Equation (7), the distance between a model feature and the human body's centre of gravity is denoted as D. The line defined by each feature point and the geometric center is called A. If all lines are connected to form A closed pattern, their eccentricity is denoted as e. There are countless lines passing through the center of mass of the human body, and the two lines that are parallel and perpendicular to the horizontal line are extracted. Using the law of Cosine, it is possible to determine the angle between the remaining line and the horizontal line, as illustrated in Equation (8).

$$\alpha_i = \arccos\left[\frac{\beta - \chi}{\delta}\right] \bullet \frac{180}{\pi} \tag{8}$$

In Equation (8), β represents the abscissa of any feature point. The horizontal coordinate of the geometric center point of the human body is denoted by δ, and the vector distance between two points is denoted by δ. However, when the FRR hears the noise, it is greatly affected, which will lead to incomplete data collection, and further affect the accuracy of the results.

B. Research on Fusion Algorithm based on Hybrid Gaussian Model

When the FRR is used in practice to capture the posture characteristics of the human body in animation, there will be various errors. To avoid these errors, HG and its generative fusion algorithm (HG-FRR) are introduced. HG can apply multiple single Gaussian random distributions so that each pixel is included. And it can only correspond to the corresponding model, which is conducive to fast work, as shown in Fig. 3.

The HG-FRR workflow shown in Fig. 3 includes four main processes [18]. First, the predefined video is imported, and if the Gaussian model has been initialized, the Gaussian model background is established, and the Gaussian model is updated, and then the current frame is cut. Then determine whether the Langsky Function (LF) needs to be operated. After LF calculation, the level of Gaussian model can be obtained, and then the result can be output after human morphology processing or direct calculation. The frame rate segment without LF is processed by the moving target, and then the initial video is judged to be over. If it has finished, the final result is printed. The unfinished video continues the LF operation. The weighted summation method of HG is shown in Equation (9).

$$\varepsilon(a_t) = \sum_{t=1}^{K} w_{i,t} * \phi(a_t, u_{it}) \tag{9}$$

In Equation (9), the moment of video interception is t. The pixel value at that time is denoted as a_t. i is the total number of models. Its weight and mean value are called $w_{i,t}, u_{it}$ respectively, and their calculation equations are as follows (10).

$$\begin{cases} w_{i,t} = (1-\varphi) * w_{i,t-1} + \varphi \\ u_{it} = (1-\gamma) * u_{it-1} + \gamma * a_t \end{cases} \tag{10}$$

In the above Equation (10), $w_{i,t-1}, u_{it-1}$ is the weight and mean value of the previous moment respectively. The mean learning rate is denoted as γ, whose value is constant (0.001 in this study). The weight learning rate is represented by φ, and the value range is $[0.001, 0.010]$. LF can relate the same coordinates of nearby frame rate units together and is a vector type function, as shown in Equation (11).

$$|\iota| = n^{-1} \bullet \sum_{i=1}^{n} \left[\left(\frac{\kappa_t(x,y)_i}{\kappa_{t-1}(x,y)_i} \right)^2 - \frac{\kappa_t(x,y)_i}{\kappa_{t-1}(x,y)_i} \right] \tag{11}$$

In formula (11), the gray difference of the same pixel at adjacent moments is denoted as $\kappa_t(x,y)_i, \kappa_{t-1}(x,y)_i$. The total number of pixels in the picture is denoted as n. $|\iota|$ is a physical quantity that describes the correlation between pixels, and the greater the distance from 0, the more uncorrelated the two pixels are. When $|\iota| \to 0$, the image set and structural elements can be switched on, that is, deepened on Equation (6), as shown in Equation (12).

$$\begin{cases} X \langle B = (X \oplus B) \otimes B \\ X \rangle B = (X \otimes B) \oplus B \end{cases} \tag{12}$$

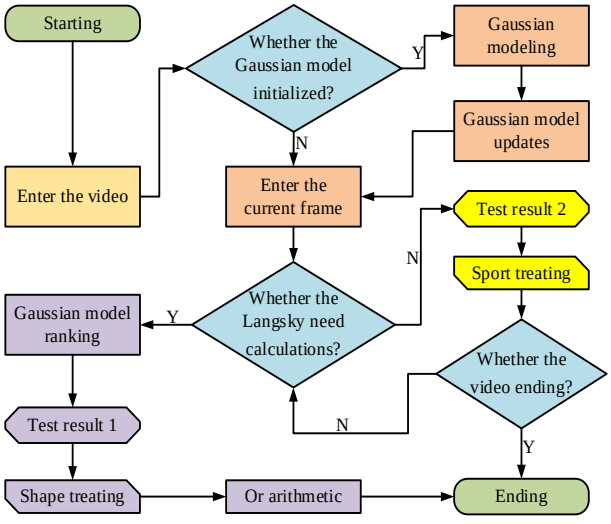

Fig. 3. Flowchart of HG working characteristics.

In Equation (12), $\langle$ represents the open operation, which has the ability to fill image gaps. The closed operation is denoted as $\rangle$, which removes redundant parts of the image. Things in the real world come with regularities, which are often found by learning machines. The learning machine requires lower empirical risk and higher fitness, and its learning process is shown in Equation (13).

$$\mu(w)=\int L(v,o(x,w))d\varpi(x,y) \tag{13}$$

In Equation (13), the classical parameters are denoted as w. The loss function is represented by $L(v,o(x,w))$. $\varpi(x,y)$ is a directivity function for positioning pixels. The coordinate set of all pixels is denoted as $o(x,w)$ [19]. A Support Vector Machine (SVM) can process uncorrelated pixels in an image due to its kernel technique. The working state of SVM control confidence range is shown in Fig. 4.

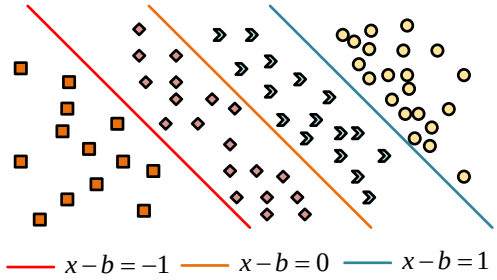

Fig. 4. Schematic diagram of SVM working state.

In Fig. 4, $x-b=0$ represents the optimal hyperplane, and different shapes represent different types of pixels. Fig. 4 shows the best working state of SVM, which can completely distinguish pixels according to whether the linearity is correlated. The kernel function, when dealing with linearly uncorrelated pixels, can map them to a high-dimensional space, thus avoiding a low-dimensional spatial disaster, as shown in Equation (14).

$$\theta(x)=\sum_{i=1}^{n}\vartheta_i\varsigma_i\omega(x_i,x)+b \tag{14}$$

In Equation (14), ($A\,\omega(x_i,x)$) represents the kernel function of the SVM. $\theta(x)$ is a classification function of the kernel function. The original spatial dimension is denoted as ϑ_i. The mapped spatial dimension and pixel abscissa are called ς_i, x_i. b represents the vector difference between dimensions. Even in the face of multi-pixel data sets, it can be trained one-to-one by Equation (14). To compare the similarity between images, the direct square matrix normalization method is introduced, and its operating equation is shown in Equation (15) [20].

$$sim(Fig_1,Fig_2)=1-\varphi(\xi_1-\xi_2) \tag{15}$$

In Equation (15), the randomly extracted images are denoted as Fig_1, Fig_2. Their gray coefficients are indicated by ξ_1, ξ_2.

IV. Model Experiment of Animation Character Shape Capture Based on HG

To verify the effect of HG-FRR algorithm in practice, the research builds HG-FRR model based on personalized recommendation, and iterates and verifies its accuracy. Finally, simulation experiments are carried out on Disert dataset using HG-FRR model.

A. HG-FRR System Development Environment and Model Parameter Determination

In this study, self-collected Disert data set was selected, including four kinds of animation character postures: upright, walking, running and jumping, with a total of 2417 frames of action images. Considering the limited types of data, the training set and test set are split up into the data set in a 2:3 ratio. The research equipment and software used in the experiment are shown in Table I.

TABLE I. Experimental Parameters

Data Set	Development Language	Code	Internal Storage
Disert	Python 11.2	Open CV	512 G
Operating system	Display card	Database	Processor
128Ubnutu 22.01.20	36.0 GHz	Mysql 5.20.2023	Intel Core i8
Web development framework	Language	Operator	Model
Django1.22.3	Easy Chinese	Standing, walking...	F2.8LII-USM

The collected data sets need to be further processed to enable the research algorithm to learn. For the processing of data sets, HG-FRR was used for iterative optimization. In order to verify its accuracy, traditional K-means algorithm, Random Forest algorithm (RF) and Particle Swarm Optimization algorithm (PSO) are compared with it. The results of accuracy and error rate in the training set are shown in Fig. 5.

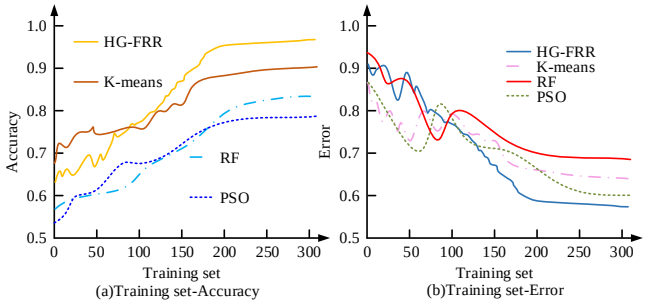

Fig. 5. Comparison of accuracy-training set image and error rate-training set image.

From Fig. 5, HG-FRR has a slightly lower accuracy rate and a higher error rate than K-means and PSO before 100 training sessions. However, the accuracy of HG-FRR is higher than the two algorithms when the number of iterations reaches more than 100, and tends to be stable when the number of iterations reaches 190, and is higher than the other three algorithms. Although the increase of iterations will reduce the operating efficiency of the model, the accuracy weight of the

model is higher after comprehensive consideration. Therefore, the proposed HG-FRR algorithm has better performance. After learning the HG-FRR algorithm, the parameter determination in the test should also be taken into account, as shown in Fig. 6.

The parameter of this study is weight learning rate $\varphi \in [0.001, 0.010]$. It can be seen from Fig. 6 that the error rate is lowest (0.042) when the weight learning rate is 0.005 and there are 150 iterations. Therefore, it is finally determined that the number of iterations is 150 times, and the regularization term takes a value of 0.005.

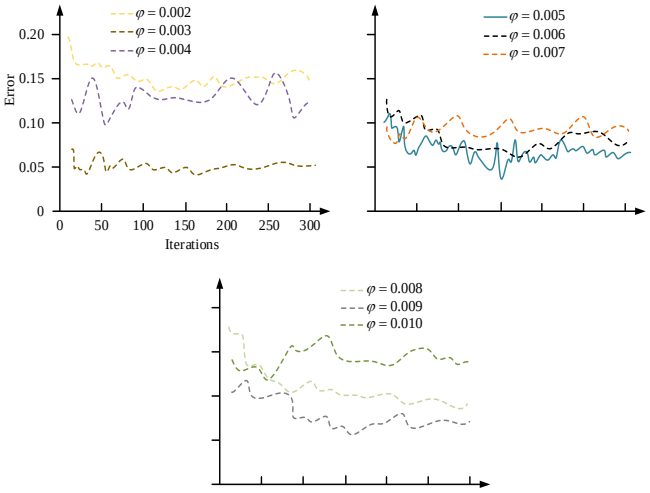

Fig. 6. Error-training times image of regularization term.

B. Experimental Verification of Animation Character Pose Recognition based on HG-FRR

To verify the accuracy of HG-FRR model in recognizing animation characters' pose, simulation experiments were conducted. By observing the task image discrimination ability of HG-FRR algorithm, the practicability of HG-FRR algorithm is judged. First, the HG-FRR algorithm is initialized, and then the video to be detected is input in the data preprocessing module. Finally, the weight learning rate is set to 0.005, and the attitude judgment records of animation characters within 60 seconds are collected, and the images are drawn after calculating errors, as shown in Fig. 7.

Fig. 7 shows the error comparison of the four algorithms in the experiment. From Fig. 7, we see that after 38s, the total error of HG-FRR has approached 0, while the other three algorithms have not stabilized. In particular, the PSO algorithm reached the highest error value of the four algorithms at 3s, which was -1.48%. The total error variation range of HG-FRR, K-means, RF and PSO is significantly compared, and the algorithm performance can be easily compared. However, it is not objective enough to rely only on the total error analysis, so the research separately analyzed the attitude misjudgment caused by the gender difference of the four cartoon characters and the motion amplitude, and drew the image as shown in Fig. 8.

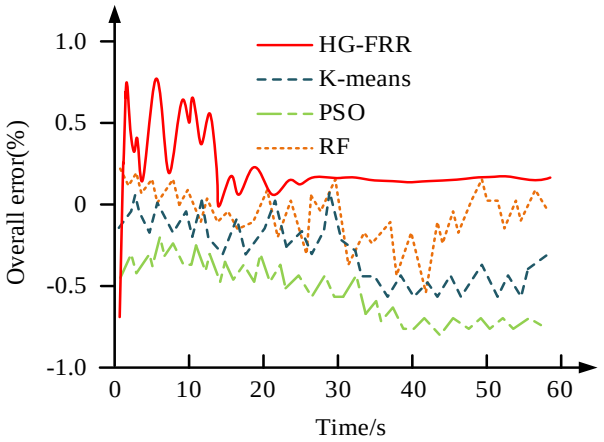

Fig. 7. Total error-time image of four algorithms.

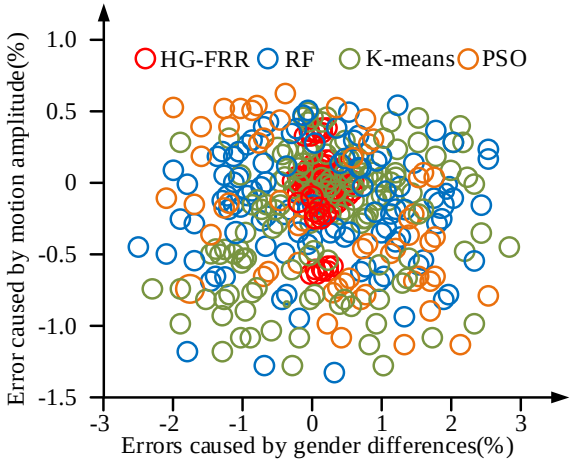

Fig. 8. The types and genre errors of the four algorithms.

From Fig. 8, the experimental results of the HG-FRR model are concentrated in the range of total error 0. The error range caused by gender difference of cartoon characters is [-0.3%, 0.4%]. The error range due to the motion amplitude is [-1.0%, 0.5%]. The errors of the remaining three algorithms are widely distributed, and the larger errors are sparsely distributed. In order to more intuitively distinguish the error correction ability of the four algorithms, the research conducted 400 experimental data records and drew the image as shown in Fig. 9.

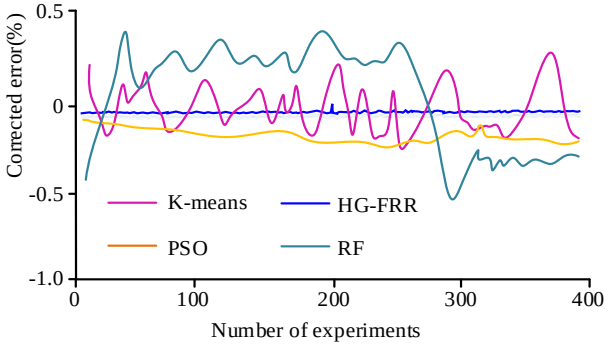

Fig. 9. Error changes of four algorithms in four hundred calibration experiments.

From Fig. 9, RF has the largest error variation range among 400 error test experiments. That is, the frequency of errors was the highest, recorded as [-0.8%, 0.6%]. The second is the K-means algorithm, which is between [-0.2%, 0.3%]. The error range of PSO is close to that of HG-FRR, with values above [-0.13%, -0.03%]. The error curve of HG-FRR fluctuates between -0.02% and 0.04%, with the smallest fluctuation range. The RF and K-means algorithms with the top two errors were excluded, and only HG-FRR algorithm and PSO were compared to capture the correct experimental results, and the error fan chart was drawn to obtain the image as shown in Fig. 10.

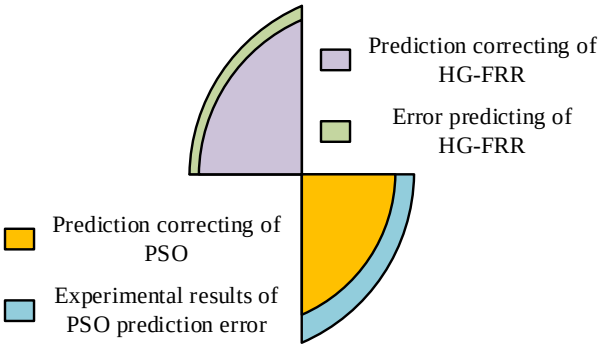

Fig. 10. Error matrix of HG-FRR algorithm and PSO algorithm.

Fig. 10 identifies four pose types of upright, walking, running and jumping based on the characteristics of cartoon characters and characters, and shows the experimental results of accurate prediction of HG-FRR and PSO. HG-FRR was accurate 392 times, with an accuracy of 98.0%, and PSO was accurate 88.2%. In order to observe the experimental results of HG-FRR and PSO more directly, linear fitting graphs of the two algorithms were drawn based on matrix, and the predicted values of the two algorithms were compared with the real values, as shown in Fig. 11.

Fig. 11 shows the comparison between the predicted and true values of the two algorithms. In Fig. 11, the linear fit degree (R^2) of HG-FRR algorithm is 0.9904, and that of PSO is 0.9546, indicating that there is no underfitting of the model. To sum up, it can be concluded that the HG-FRR algorithm model can well capture and identify the common gestures of animation characters, so as to give users a good VR experience.

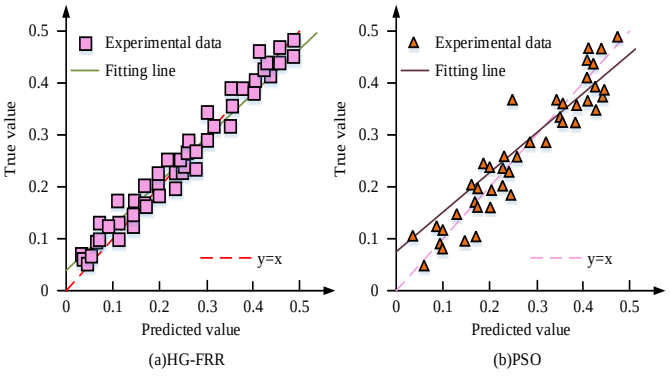

Fig. 11. Linear fitting diagram of HG-FRR and PSO.

V. Result and Discussion

It can be seen from the experimental results that the total error of HG-FRR has approached zero after 38s, while the other three algorithms have not stabilized, which shows that the experimental effect of the fusion algorithm is better than the other three algorithms in the stability test of the system. In the total error of HG-FRR, the error range caused by the gender difference of cartoon characters is [-0.3%, 0.4%], and the error range caused by the action range is [-1.0%, 0.5%], while the error distribution of the other three algorithms is irregular and sparse, and only the total error of HG-FRR can be judged to be the lowest. It shows that the proposed fusion model has the best prediction effect under the same experimental environment. For the persuasiveness of the experiment, the study conducted 400 experiments. In 400 error test experiments, the error range of RF is the largest, which is recorded as [-0.8%, 0.6%]. K-means algorithm is between [-0.2%, 0.3%]. The error range of PSO is close to HG-FRR, and the value is above [-0.13%, -0.03%]. The error curve of HG-FRR fluctuates between-0.02% and 0.04%, and the fluctuation range is the smallest. It shows that HG-FRR is still the best in the extensive error test. In the experimental results of accurate prediction of HG-FRR and PSO, the prediction accuracy of HG-FRR is 98.0%, and that of PSO is 88.2%. The linear fitting degrees of them are 0.9904 and 0.9546, respectively. It shows that HG-FRR is extensive in the experiment of predicting cartoon characters. Through many experiments, it is proved that the fusion algorithm proposed in this study has the lowest total error and the lowest distribution error among the algorithms for judging the image of anime characters. Because this error is very sensitive to the influence of cartoon character image judgment, so the small error in the experiment cannot be ignored. However, most animation works rely on freehand drawing and post-processing, which will bring a lot of noise to the prediction of animation characters. Under this background, the proposed fusion algorithm can still keep the error between-0.02% and 0.04%, which shows that the algorithm is highly adaptable to the judgment of animation character image and is suitable for capturing the character form in animation.

VI. Conclusion

With the development of the Internet industry, it is becoming more and more important to identify the posture of animation characters when watching animation, such as increasing the play experience for users and improving the visibility of animation merchants. In this study, a fusion algorithm (HG-FRR) was constructed based on Frame Rate Reduction (FRR) and Hybrid Gaussian model (HG). In this study, both bilateral filter denoising and Lonsky function are taken into account, on the Disert data set, simulation tests are run, and comparison is made with K-means and other three algorithms. 40% of Disert data set was extracted and trained on HG-FRR model through experiments of shared hyperplane. Finally, the number of iterations is determined to be 150 times, and the weight learning rate is 0.005. In the error analysis experiment, a total of 400 experiments were conducted. K - means and RF error of the two algorithms in [- 0.2%, 0.3%], [- 0.8%, 0.6%] within the scope of volatility. The error curve of HG-FRR fluctuates between -0.02% and 0.04%, with the

smallest fluctuation range. The variation range of PSO is close to that of HG-RPP, between [-0.13%, -0.03%]. In 400 experiments, the prediction accuracy of HG-FRR is 98.0% and that of PSO is 88.2%. The linear fitting graph of the two algorithms based on matrix is studied. R^2 of HG-FRR is 0.9904, indicating excellent linear fitting. PSO's R^2 is 0.9546. To sum up, it can be concluded that the HG-FRR algorithm model can accurately capture the posture of animation characters, which can not only improve the user's viewing happiness, but also make the animation produced by merchants more well-known. However, the HG-RPP model is only applicable to the analysis of animation with blurred background. For works with complex background, the character characteristics are not obvious, and the model will mark them as noise. This is because the commercial value of animation works belongs to private information, and the dataset analyzed in this study contains few types. With more volunteers, it is believed that future studies can be improved.

ACKNOWLEDGMENT

The research is supported by: Key Research Project of Humanities and Social Sciences in Colleges and Universities in Anhui Province, Research on the Application of Digital Media Art in Animation Design, (No. SK2020A0930), Research on the Creation of Regional Brand Culture and Art Derivatives based on the Design Art in the Consumption Era, (No. SK2021A1030). Outstanding Top-notch Talent Cultivation Funding Project in Anhui Province, Domestic Visit Study (No. gxgnfx2021108). Key Research Project of Natural Science in Colleges and Universities in Anhui Province, Research on Case-based Reasoning Decision Support Technology of Traffic Congestion Alleviation, (No. KJ2021ZD0126), "Double Base" Teaching Demonstration Course in Anhui, Photoshop, (No. 2020SJJXSFK0451).

REFERENCES

[1] Guo Y, Mustafaoglu Z, & Koundal D. Spam Detection Using Bidirectional Transformers and Machine Learning Classifier Algorithms. Journal of Computational and Cognitive Engineering, 2022, 2(1), 5–9.

[2] Kumar Y, Verma S K, Sharma S. Multi-pose facial expression recognition using hybrid deep learning model with improved variant of gravitational search algorithm. Int. Arab J. Inf. Technol., 2022, 19(2): 281-287.

[3] Dougherty J J, Bray N N, Vanier C H. Attitudes toward osteopathic recognition under the single GME accreditation system: a survey of deans at colleges of osteopathic medicine and chairs of osteopathic manipulative medicine departments. Journal of Osteopathic Medicine, 2020, 120(2): 81-89.

[4] Wang C, Zhang Z, Xi Z. A human body based on sift-neural network algorithm attitude recognition method. Journal of Medical Imaging and Health Informatics, 2020, 10(1): 129-133.

[5] Liao M, Wang C, Gu X. Algorithm using supervised subspace learning and non-local representation for pose variation recognition. IET Computer Vision, 2020, 14(7): 528-537.

[6] Yang X, Jia X, Yuan M, D Yan. Real-time facial pose estimation and tracking by coarse to fine iterative optimization. Tsinghua Science and Technology, 2020, 25(5): 690-700.

[7] Jain S, Rustagi A, Saurav S. Three-dimensional CNN-inspired deep learning architecture for Yoga pose recognition in the real-world environment. Neural Computing and Applications, 2021, 33: 6427-6441.

[8] Gunawan K W, Halimawan N. Lightweight end to end pose-robust face recognition system with deep residual equivariant mapping. Procedia Computer Science, 2021, 179: 648-655.

[9] Zhang F, Zhang T, Mao Q, Xu. Geometry guided pose-invariant facial expression recognition. IEEE Transactions on Image Processing, 2020, 29: 4445-4460.

[10] Yoganand A V, Kavida A C, Devi D R. Pose and occlusion invariant face recognition system for video surveillance using extensive feature set. International Journal of Biomedical Engineering and Technology, 2020, 33(3): 222-239.

[11] Li S, Dai Y, Hirota K, Z Zuo. A Students' Concentration Evaluation Algorithm Based on Facial Attitude Recognition via Classroom Surveillance Video. Journal of Advanced Computational Intelligence and Intelligent Informatics, 2020, 24(7): 891-899.

[12] Chen Z, Xie W. Infrared Image Face Recognition Method Based on Signal Interference Technology Advanced Hybrid Information Processing: EAI International Conference, ADHIP, Proceedings, Part I. Cham: Springer Nature Switzerland, 2022, 28(3):280-287.

[13] Zhu Z, Cheng Y. Application of attitude tracking algorithm for face recognition based on OpenCV in the intelligent door lock. Computer Communications, 2020, 154: 390-397.

[14] Maehama K, Even J, Ishi C T, Kanda T. Enabling Robots to Distinguish Between Aggressive and Joking Attitudes. IEEE Robotics and Automation Letters, 2021, 6(4): 8037-8044.

[15] Ünver M, Olgun M, Türkarslan E. Cosine and cotangent similarity measures based on Choquet integral for Spherical fuzzy sets and applications to pattern recognition. Journal of Computational and Cognitive Engineering, 2022, 1(1): 21-31.

[16] Zhou L, Xue F. Show products or show people: An eye-tracking study of visual branding strategy on Instagram. Journal of Research in Interactive Marketing, 2021, 15(4): 729-749.

[17] Bassi A, Fabbri A. Under pressure: Evolution of the social economy institutional recognition in the EU. Annals of Public and Cooperative Economics, 2020, 91(3): 411-433.

[18] Taherpour F, Ghiasvand E, Namian M. The effect of fatigue on safety attitude, hazard recognition and safety risk perception among construction workers. Amirkabir Journal of Civil Engineering, 2021, 53(8): 3299-3316.

[19] Oslund S, Washington C, So A, Chen, T, &Ji, H. Multiview Robust Adversarial Stickers for Arbitrary Objects in the Physical World. Journal of Computational and Cognitive Engineering, 2022, 1(4): 152-158.

[20] Choi H J, Kwon Y C. Factors Influencing Dementia Attitude and Recognition of Dementia Policy of Nursing Students. The Journal of the Convergence on Culture Technology, 2020, 6(2): 161-168.