

Research Article

3D Face Modeling Algorithm for Film and Television Animation Based on Lightweight Convolutional Neural Network

Cheng Di ^{1,2}, Jing Peng,¹ Yihua Di,³ and Siwei Wu ^{1,2}

¹School of Arts and Communication, China University of Geosciences, Wuhan 430074, China

²Faculty of Fine and Applied Arts, Burapha University, Tambon Saensuk, Chonburi 20131, Thailand

³College of Mechanical and Electrical Engineering, Wuhan Textile University, Wuhan 430074, China

Correspondence should be addressed to Siwei Wu; sewell@cug.edu.cn

Received 16 April 2021; Revised 12 May 2021; Accepted 15 May 2021; Published 25 May 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Cheng Di et al. 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.

Through the analysis of facial feature extraction technology, this paper designs a lightweight convolutional neural network (LW-CNN). The LW-CNN model adopts a separable convolution structure, which can propose more accurate features with fewer parameters and can extract 3D feature points of a human face. In order to enhance the accuracy of feature extraction, a face detection method based on the inverted triangle structure is used to detect the face frame of the images in the training set before the model extracts the features. Aiming at the problem that the feature extraction algorithm based on the difference criterion cannot effectively extract the discriminative information, the Generalized Multiple Maximum Dispersion Difference Criterion (GMMSD) and the corresponding feature extraction algorithm are proposed. The algorithm uses the difference criterion instead of the entropy criterion to avoid the “small sample” problem, and the use of QR decomposition can extract more effective discriminative features for facial recognition, while also reducing the computational complexity of feature extraction. Compared with traditional feature extraction methods, GMMSD avoids the problem of “small samples” and does not require preprocessing steps on the samples; it uses QR decomposition to extract features from the original samples and retains the distribution characteristics of the original samples. According to different change matrices, GMMSD can evolve into different feature extraction algorithms, which shows the generalized characteristics of GMMSD. Experiments show that GMMSD can effectively extract facial identification features and improve the accuracy of facial recognition.

1. Introduction

Film and television animation, in a broad sense, is to turn some originally inactive things into moving images through film production and projection, which is called film and television animation. Movie animation is the collective term for movie animation, TV animation, and animation in movies. It is an active, virtual image created by various technical means. It is a comprehensive art that integrates movies, literature, painting, and music. Film and television animation is the crystallization of audiovisual art and technology. Its inclusiveness and interdisciplinarity are unmatched by any other art form; its complex and pluralistic nature cannot be replaced by other art forms. The production of film and television animation is a very complicated process. It is labor-intensive labor that requires a team

to complete it. It is the crystallization of collective wisdom. Any complete film and television animation must have three major components: story, character, and scene [1]. The story is the content of the plot; the role is the actor who develops the story and the conflicts; the scene is all the scenery that has a relationship with the role along with the development of the story, the living place, the natural environment, the social environment, and the historical environment [2]. In film and television animation, these three parts are equally important, inseparable, and indispensable.

The film and television animation scene is a typical space environment for the unfolding of the plot and role activities, and it is the main creative link in the formation of animation style. For the modeling performance of the animation scene, it can show the animation style, light, color, and other aspects, which play a very important role in the expansion of

the plot and the formation of the character's character [3]. Film and television animation scenes have a strong uniqueness in artistic creation, and research on it is of great significance for improving the level of animation production [4]. Facial transplantation is not too problematic due to the precise feature positioning technology. In order to improve the authenticity of the transplantation result, eye movement and head posture rotation are also very critical factors. Because of the temporal correlation between the eye expression and the head, that is, the action of the previous frame has a direct influence on the action of the next frame, while the face has no temporal correlation [5]. How to deal with the transplantation of the face, eyes, and head at the same time is the biggest problem [6].

This article builds a lightweight 3D facial feature extraction model. Since there are currently fewer 3D face data sets publicly available, this article annotates the 3D face data sets required for training and testing the model and makes corresponding predictions for the data sets. In order to narrow the search range of the feature extraction model, before training the model, face frame detection is performed on the training samples. The detection algorithm is a face detector with an inverted triangle structure, which combines the fast speed of the Adaboost algorithm based on LAB features. The high accuracy of detection and the MLP classifier based on SURF features, from coarse to fine step by step, fully considers the speed and accuracy. In the 3D face feature extraction stage, considering the light weight of the depth separable convolution structure, this article uses it as the main structure of the model to design a more accurate feature point extraction with lower computational complexity lightweight convolutional neural network model. This paper proposes the GMMSD criterion. Based on this criterion, QR decomposition is applied to solve the feature matrix to extract feature vectors that are more discriminative for recognition, which is used for facial recognition and facial recognition. The criterion uses the idea of interclass and intraclass scatter matrix difference to avoid the "small sample" problem that may appear in the Fisher criterion. At the same time, the transformation matrix is extracted from the whitening sample data, which effectively reduces the training time of the algorithm and maintains the effectiveness of the algorithm. Due to the generalized characteristics of GMMSD, choosing different transformation matrices can evolve into different feature extraction algorithms. The experimental comparison results of Jaffe and RML face libraries, AR, FERET, and Yale face libraries show that the GMMSD algorithm can not only improve the recognition accuracy but also has lower training complexity than other algorithms.

2. Related Work

Because of the wide application prospects of face recognition technology, it has attracted a large number of researchers to research and improve it theoretically [7]. Therefore, face recognition technology has been greatly developed in recent years. Many software research and development units have also invested a lot of manpower

and material resources to develop face recognition systems that can be used in commercial applications [8]. However, the most current face recognition system adopts two-dimensional face recognition technology. Although the accuracy of the two-dimensional face recognition technology has been greatly improved with the further development of the technology, the accuracy of the two-dimensional face recognition technology has been greatly improved. In the process of face recognition, the recognition defects caused by the changes in illumination, posture, face, etc., can only be compensated by three-dimensional face recognition technology. Therefore, three-dimensional face recognition has become a relatively new research direction in the current image processing research [9]. The recognition method based on three-dimensional face data is mainly used to solve the difficult facial pose caused by the two-dimensional face recognition method.

At present, the main channel for obtaining 3D face image information is the 3D face data information obtained by the laser scanner method, but the method of obtaining data through this channel is expensive, and the noise of the obtained data is relatively large, which cannot be directly applied to 3D face recognition; reconstruction of the realistic three-dimensional face through the front and side color images, due to the complexity of face surface information, is easy to cause the loss of a lot of depth information [10, 11]. The amount of calculation and data required is large and the speed is very slow. These methods for obtaining 3D information have various defects, but the biggest obstacle to the application of 3D face recognition in practice is that a large amount of data during use often consumes a lot of time, and it is difficult to meet the real-time needs in actual applications [12, 13].

Relevant scholars obtain multiple samples of a single-sample face based on the Singular Value Decomposition (SVD) perturbation algorithm and apply the traditional LDA algorithm for feature extraction [14]. Although these methods can alleviate the recognition accuracy of the single-sample face recognition problem to a certain extent, there is a common problem. That is, the virtual face samples have a very high correlation, and these samples cannot be regarded as independent. Therefore, the distinguishable feature subspaces learned in the feature extraction stage have redundancy [15]. Although there is only a single sample for each type of face in the database, the sample expansion method can generate virtual multiple samples of these faces from the original sample image, expanding the number of training samples for each type [16]. Through the sample expansion method, each person in the training face image database no longer has a single sample, which becomes a general multiple face recognition problem. Researchers proposed a general learning framework, and based on this framework, they proposed several distinguishable feature extraction algorithms based on the entire image [17]. Relevant scholars have proposed a transfer subspace learning method, which can perform single-sample face recognition by learning from a general training sample set to obtain information that distinguishes different people in the transfer subspace model [18].

The regression model based on cascading shapes first creates a rough face shape and then gradually optimizes the rough shape to the most suitable point by a multistage regression. Because it can be used to train large-scale data, it is easy to operate and it is more convenient to replace different features, so the algorithm is popular as soon as it is proposed [19]. But it also faces some drawbacks, such as slow positioning speed and low positioning accuracy when the posture is complex. In this regard, scholars proposed a robust cascaded shape regression model [20–22]. The model also labels the occluded part of the face in the training set and sets a value for the visibility of each area. The occlusion problem in face alignment is solved. In addition, the introduction of shape index features improves the accuracy of face positioning in many situations, but the model is very complicated and the running speed will still be affected [23, 24]. In recent years, many improved algorithms have appeared, but the cascaded regression structure always needs to construct the initial face shape, and the possible deviation is still inevitable [25, 26].

Relevant scholars proposed a denoising autoencoder on the basis of a stacked autoencoder [27]. By adding noise to the signal as a training set, the reconstructed signal is compared with the signal without added noise as the reconstruction error. The device has the ability to resist noise. Since the deep belief network ignores the two-dimensional structure of the image, for detecting the weight of a given feature, each position must be learned separately, which undoubtedly increases the amount of calculation. On this basis, related scholars combined deep belief networks and convolutional neural networks [28]. By sharing weights at all positions in an image, they proposed convolutional deep belief networks and applied them to voice recognition and face recognition. Convolutional deep belief networks are based on convolution restricted Boltzmann machines. CRBM is similar to RBM, but all positions in a graph share the weights between the hidden layer and the visible layer. It is found that the receptive field of small (usually the smallest) convolutional neurons produces a large amount of network depth, which is roughly similar to the sparsely connected neural layer between the retina and visual cortex of mammals, and only activated neurons are affected. According to this, related scholars unite multiple deep network models to do averaging processing and achieve comparable competitiveness with human performance in some recognition tasks [29].

3. Facial Feature Point Cloud Data Processing Method Based on 3D Scanning Engineering Modeling

3.1. Point Cloud Data Flattening Method. The splicing of point cloud data is an important part of preprocessing, and there are many forms of splicing. For the splicing of point cloud data with obvious characteristics, the coordinate values of the point cloud data can be determined by means of points, lines, and surfaces for splicing. Most products are composed of free-form surfaces, there are no obvious data features on the model, the position of the measured parts is

arbitrarily placed during the measurement, and the connection between the obtained point cloud data coordinate system is not available. Therefore, it is particularly important to combine the data.

The ICP algorithm is one of the most used matching algorithms in the point cloud data registration process. It optimizes the matrix through iteration. In each iteration process, each point on the target point set must be found in the reference point set to find the closest point, calculate the corresponding rotation matrix and translation vector by using the closest point found in this way, apply these parameters to the target point set to obtain a new target point set and proceed to the next iteration process, and finally get an excellent conversion matrix.

The basic principle of the ICP algorithm is to give a data point set and a model, establish a point-to-point correspondence by continuously searching each data point in the data point set to the closest point of the model, and then iteratively calculate the sum of the data point set and the model. Taking the alignment of two curves as an example, the basic principle of the ICP algorithm can be described as follows: ideally, when the relationship between the corresponding points on the two curves is determined, it is obvious that the correct rotation matrix and translation matrix can be found to align the two curves.

Given two point sets to be spliced P and Q , we set Q as the reference point set (fixed point set), and P as the target point set (it needs to be transformed to the coordinate system based on Q). In order to make P and Q can be spliced together, first, we find a point in Q that is the closest to it for each point in the point set P , establish the point-to-point mapping relationship, and then calculate an optimal coordinate by the least square method, that is, rotation matrix R and translation vector T , and set $P = RP + T$, and then iterative solution until the accuracy is met, and the final rotation matrix and translation vector can be obtained.

We solve the rotation matrix R and the translation vector T according to the rigid body transformation so that the sum of the squares of the distances of all the closest points (p_i, q_i) is minimized, and the error function is established:

$$E(R, T) = e^{\prod_{i=0}^{n-1} \|R p_i - T - q_i\|^2}. \quad (1)$$

One transformation does not minimize the error function. It takes multiple iteration calculations to find the optimal rotation matrix and translation vector. Suppose that the rotation matrix obtained in the k th iteration is R_k and the translation vector is T_k ; let

$$P_{k+1} = R_k \cdot P_{k-1} - T_k \cdot T_{k-1}. \quad (2)$$

The conditions for determining the number of iteration terminations N are as follows. Given a threshold γ , if the results of two adjacent iterations satisfy the following formula, the iteration ends.

$$E(R_k, T_k) - E(R_{k-1}, T_{k-1}) \leq \gamma. \quad (3)$$

We calculate the centroids of the point sets $\{p_i\}$ and $\{q_i\}$ respectively:

$$\begin{aligned}\vec{u}_p &= e^{1/(n-1)} \sum_{i=0}^{n-1} p_i, \\ \vec{u}_q &= e^{1/(n-1)} \sum_{i=0}^{n-1} q_i.\end{aligned}\quad (4)$$

According to the rotation matrix R and the centroids u_p and u_q of the two point sets, the translation vector can be obtained:

$$T = \vec{u}_q + R\vec{u}_p. \quad (5)$$

3.2. Modeling Method of Facial Feature Point Cloud Data. Nonuniform rational B-spline (NURBS) can be used to represent analytical geometric shapes and can be used to represent curves and surfaces. At present, it has become the standard for curve and surface representation in the current CAD system. NURBS surface can represent the sum of rational polynomials in segments, and the form of expression is

$$S(u, v) = e^{\prod_{i=0}^n \prod_{j=0}^m W_{i,j} P_{i,j} B_{j,d} B_{i,k} / \prod_{i=0}^n \prod_{j=0}^m W_{i,j} B_{j,d} B_{i,k}}. \quad (6)$$

Triangulation is regarded as the most basic kind of grid, it can adapt to the regular and irregular distribution of data, and it has unique advantages in shape performance. The Delaunay triangulation is a collection of adjacent and nonoverlapping triangles. The circumscribed circle of each triangle does not contain other vertices. The Delaunay triangle is formed by connecting three adjacent points. The Voronoi polygon corresponding to these three adjacent points has a common vertex.

Among various triangulations, only Delaunay triangulations perform best in shape fitting. Delaunay triangulation has a “circle rule” or “maximum and minimum angle rule”; that is, the circumcircle of any triangle does not include other nodes, so as to avoid the appearance of sharp internal angles as much as possible.

4. 3D Facial Rapid Modeling Based on Lightweight Convolutional Neural Network

4.1. Face Detection. Before feature extraction, choosing a suitable algorithm for face detection is a very important step, which will help improve the results of feature extraction. The existing face detector is used to detect a single face with a clear front, but there are some uncontrollable factors in the face to be detected in practical applications, such as the following: the head deflection angle is too large, the face is exaggerated, and the face is too strong or too strong. These factors are a test of the robustness of the face detector. In order to solve the multiview face problem, the most direct way is to use multiple detection models to form a parallel structure to detect various angles of the face. However, this method requires each model to classify the candidate window, resulting in computational cost, and the false alarm rate is relatively high.

As shown in Figure 1(a), the first step is to classify the sample set S1 to obtain a subclassifier C1. In order to

enhance the processing of classified wrong faces, the wrong faces are weighted to obtain a new sample set S2.

The second step is to classify S2 and train a subclassifier C2, as shown in Figure 1(b). The wrongly classified faces are weighted to obtain a new sample set S3. As shown in Figure 1(c), the third step is to classify S3 to obtain the subclassifier C3. Finally, the subclassifiers of the first three steps are combined. Although there are certain errors in each of the previous steps of the classification process, the three weak classifiers can be combined in a specific way to achieve extremely high classification accuracy.

Therefore, the top layer uses multiple Adaboost cascade classifiers to extract local binary Haar features and quickly and roughly estimate the face range. The second and third layers mainly verify whether the candidate window output by the upper layer is a facial area and combine the multilayer perceptual classifier with SURF features to classify finer areas while ensuring speed. After two-layer rough detection, the last layer uses a unified multilayer perceptron combined with shape index features to integrate the output windows of the previous layer. The complexity of the classifier gradually increases from top to bottom, and the classification accuracy is gradually improved. In addition, the number of classifiers gradually decreases from top to bottom in the structural design, and the features of the top and second and third layers are selected in consideration of speed, so the final detection speed is also guaranteed.

4.2. Deeply Separable Convolution Structure. The depth separable convolution structure is a change from the traditional convolution structure. In the traditional convolution structure, the channel of each convolution kernel is convolved with the data of the corresponding channel of the input data, and the calculation results of all channels are superimposed after the convolution is completed. Assuming a convolutional layer of a * a, the input channel is M and the output channel is N, then N convolution kernels are needed to complete the convolution, and each convolution kernel needs a * a * M parameters, because the convolution kernel needs to perform a convolution with all input data channels, so the total number of parameters required is a * a * M * N, and the computational complexity is very high. The depth separable convolution structure is shown in Figure 2.

We compare the number of parameters required by the traditional convolution structure and the depth separable convolution structure. Compared with the traditional convolution, the number of parameters of the depth separable convolution is greatly reduced, and the amount of calculation is also reduced accordingly. In addition, the traditional convolution structure often considers the two factors of channel and region at the same time, while the depth separable convolution structure separates the two. The region is considered first and then the channel is considered. Experiments have proved that the building is built in this way. The accuracy of the final model is better than the model built by traditional convolution, and the running speed is correspondingly improved. For applications targeting mobile devices, this type of model has strong applicability.

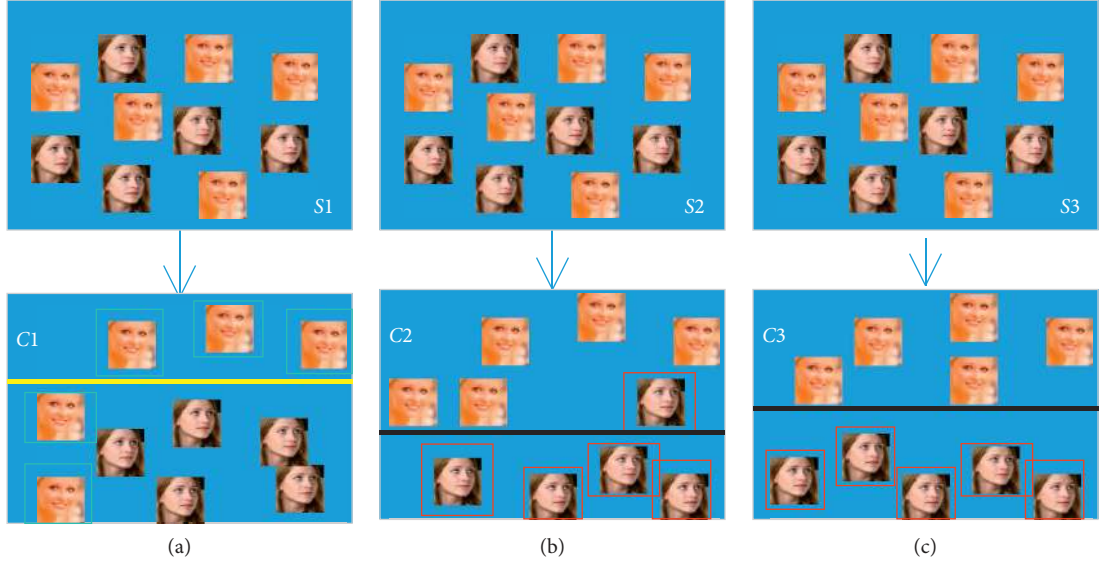


FIGURE 1: Schematic diagram of Adaboost algorithm. (a) The first step of classification. (b) The second step of classification. (c) The third step of classification.

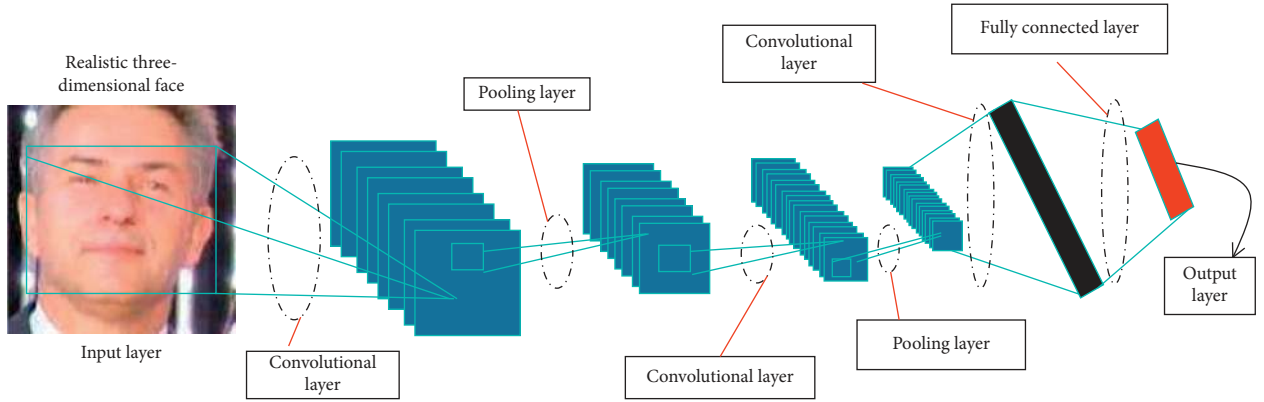


FIGURE 2: Depth separable convolution structure.

In order to evaluate the error between the output coordinates of the model and the real coordinates, this paper uses the Smooth $L1$ function as the loss function, and the formula is as follows:

$$\text{loss}(x_i, y_i) = \begin{cases} \frac{2}{3} \ln(x_i - y_i), & |x_i - y_i| < 1, \\ \ln\left(x_i - y_i - \frac{1}{2}\right), & |x_i - y_i| \geq 1. \end{cases} \quad (7)$$

The error of this function is the square loss in the $(-1, 1)$ interval and the $L1$ loss in other cases. This method cleverly avoids the problem of gradient explosion.

4.3. 3D Facial Rapid Modeling. Delaunay triangle mesh algorithm is mainly divided into point-by-point insertion method and region growing method. The region growing method is mainly for the processing of point cloud data, while the point-by-point interpolation method is more

suitable for grid reconstruction based on feature points. The region growth law first constructs an initial triangle surface, adds three edges to the boundary set, adds the rest to the surface set, and then grows from the boundary to generate a new triangle surface area until a complete triangle mesh is generated. The point-by-point insertion method is to insert all the discrete points in sequence, find the triangles that contain new points in the circumscribed circle, and connect the points with these triangles to realize the insertion of the new points. This paper uses point-by-point interpolation to reconstruct the grid. The insertion process is shown in Figure 3, and the specific details can be seen in Figure 4.

5. Experimental Results and Analysis

5.1. Comparison Experiment of GMMSD versus MMSD. Since both GMMSD and MMSD use the method of difference criterion to extract discriminative features, this section compares the recognition performance of GMMSD and MMSD based on different numbers of training samples. P

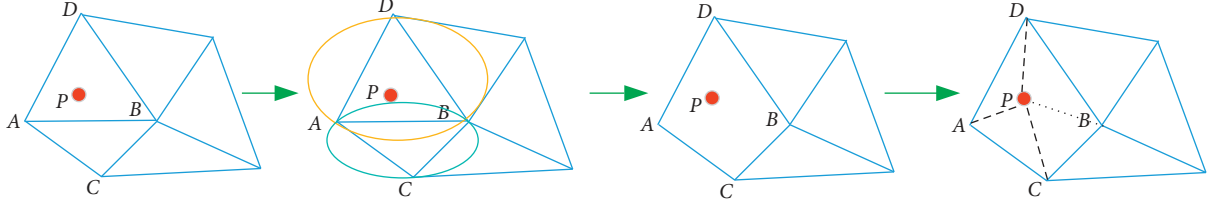


FIGURE 3: Triangular mesh reconstruction algorithm.

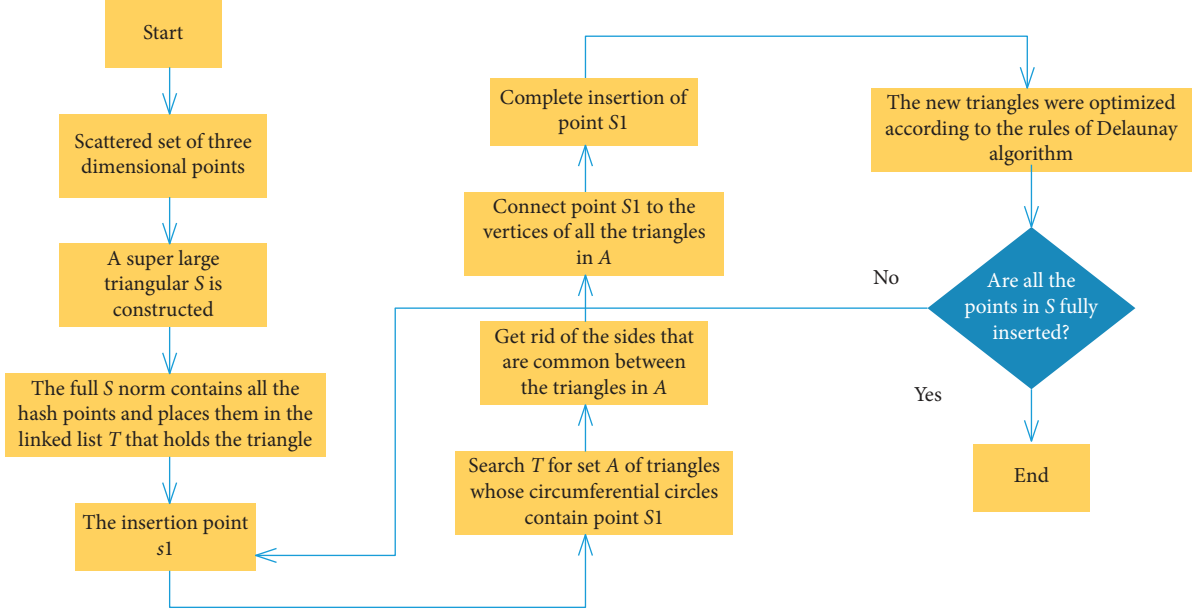


FIGURE 4: Delaunay point-by-point insertion method.

samples of each face (or each person) participate in training, and the remaining samples are used as tests. Figure 5 shows the recognition rates of different numbers of training samples corresponding to different numbers of feature vectors.

It can be seen from the above experimental results that the number of identification feature vectors is different, and the recognition rate is also different. When the number of training samples increases, the recognition rates of GMMSD and MMSD both increase significantly, as shown in Figure 5(a). It is worth noting that, on the one hand, the recognition performance of GMMSD is comparable to that of MMSD, and in most cases, the recognition performance of GMMSD is better. On the other hand, compared with MMSD, GMMSD has the advantage of low computational complexity. Similar results were obtained on the FERET library, as shown in Figure 5(b).

5.2. Feature Extraction Performance Analysis of Each Algorithm in the Face Database. In order to further evaluate the classification performance of GMMSD, simulation and comparison of five traditional feature extraction algorithms are PCA + LDA, R-LDA, N-LDA, MSD, and MMSD. Similarly, the nearest neighbor classification method is used in

classification and recognition. Figures 6(a) and 6(b), respectively, show the recognition performance of five feature extraction algorithms in two face databases. It can be observed from Figure 6 that the recognition results of GMMSD and MMSD are higher than other algorithms. In addition, the recognition results of GMMSD and MMSD are similar. This is because they all extract discriminative features from the dewhitened sample matrix. However, GMMSD has lower computational complexity than MMSD.

The recognition performance of R-LDA is better than PCA + LDA, and the recognition performance of N-LDA is the worst. The reason may be that R-LDA is closer to Fisher's linear criterion, just to avoid the singularity of the intraclass scatter matrix, a disturbance matrix is added to it, and N-LDA loses a lot of discrimination information in the feature extraction process.

5.3. Comparison of Feature Extraction and Recognition Performance between GMMSD and Other Algorithms. This section attempts to give a comparison of recognition rates between GMMSD and several other popular algorithms. In particular, it should be pointed out that due to the different experimental environment and parameters, such as the

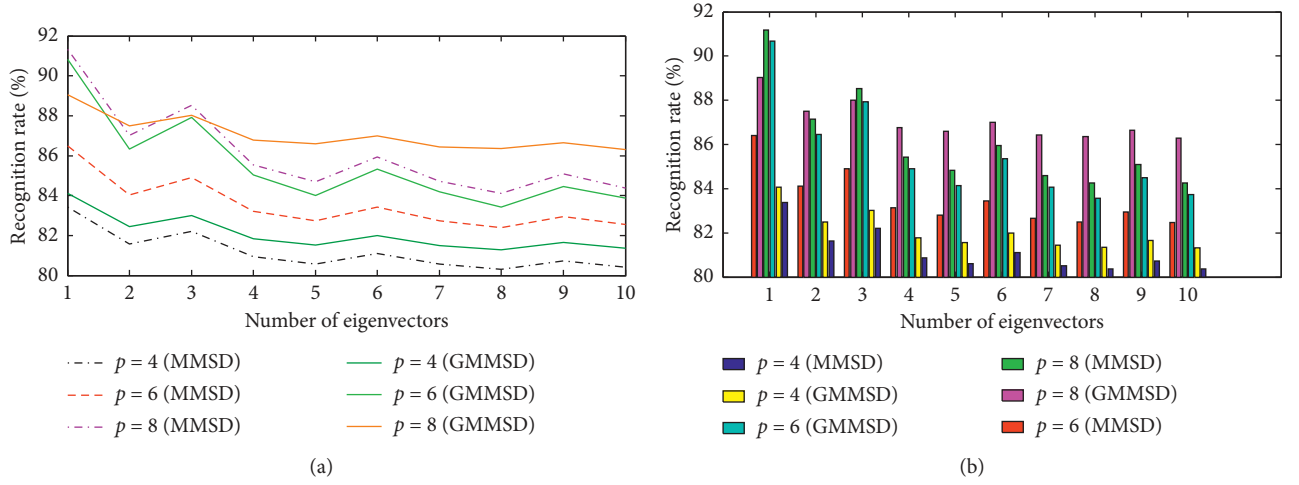


FIGURE 5: GMMSD versus MMSD recognition rate. (a) RML face library. (b) FERET face database.

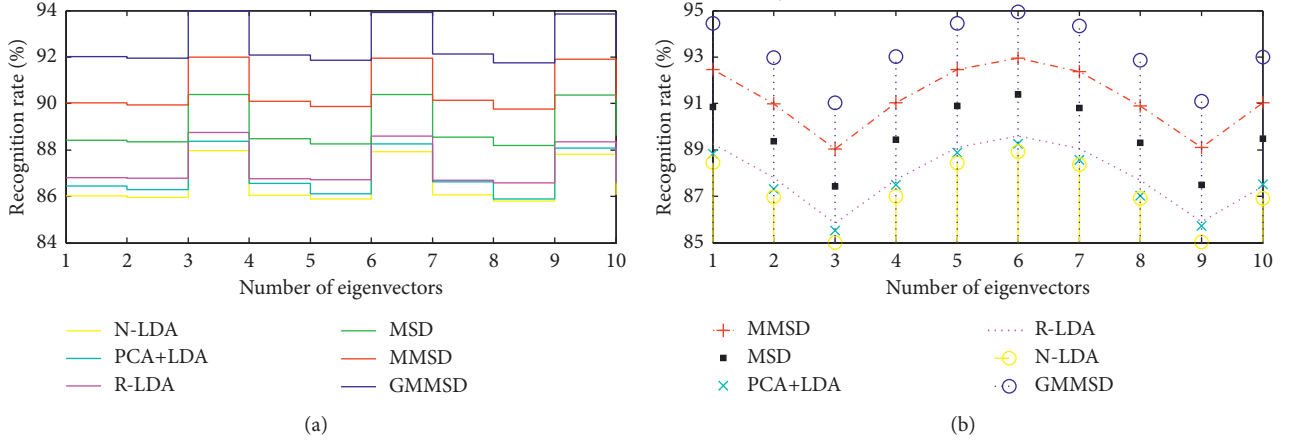


FIGURE 6: Average recognition rate of faces. (a) Jaffe face library. (b) RML face library.

number of samples, face types, and other factors, the results of different algorithms may not be directly compared and studied, but the identification results can still reflect the discriminating ability of these algorithms. Figure 7 shows the comparison and recognition results of several algorithms. The first seven uses the entire face as a sample, while the latter three use key areas of the face as a sample. It can be seen from Figure 7 that GMMSD achieves the best recognition effect.

5.4. Comparison and Analysis of Computational Complexity. In order to verify the advantages of the GMMSD algorithm in terms of computational complexity, this section simulates and compares the computational complexity of GMMSD,

involving algorithms such as MMSD, MSD, PCA + LDA, N-LDA, and R-LDA. The experiment was run on a personal laptop computer with the main configuration: Intel Celeron CPU and 4 GB RAM. Randomly, we combine 10 samples of each person (each face) into the training set. In order to obtain statistically significant test results, the experiment of each algorithm is repeated 10 times. Figure 8 shows the average time required for each algorithm to extract features. It can be seen from Figure 8 that the feature extraction time of GMMSD is less than 30 s and is significantly shorter than that of MMSD and MSD. The training efficiency of each algorithm is also simulated and compared on the FERET library and the Yale library. As shown in Figures 8(b) and 8(c), it can also be seen that the computational complexity of GMMSD is much lower than other algorithms.

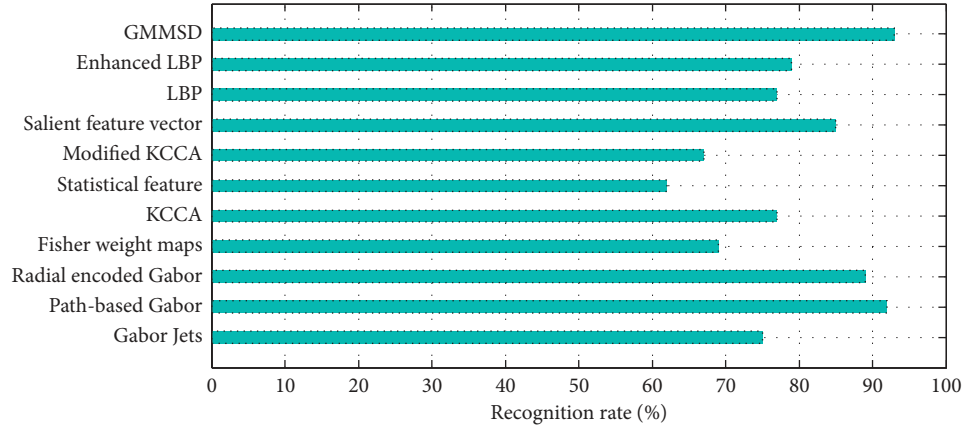


FIGURE 7: The recognition rate of different algorithms in the Jaffe face library.

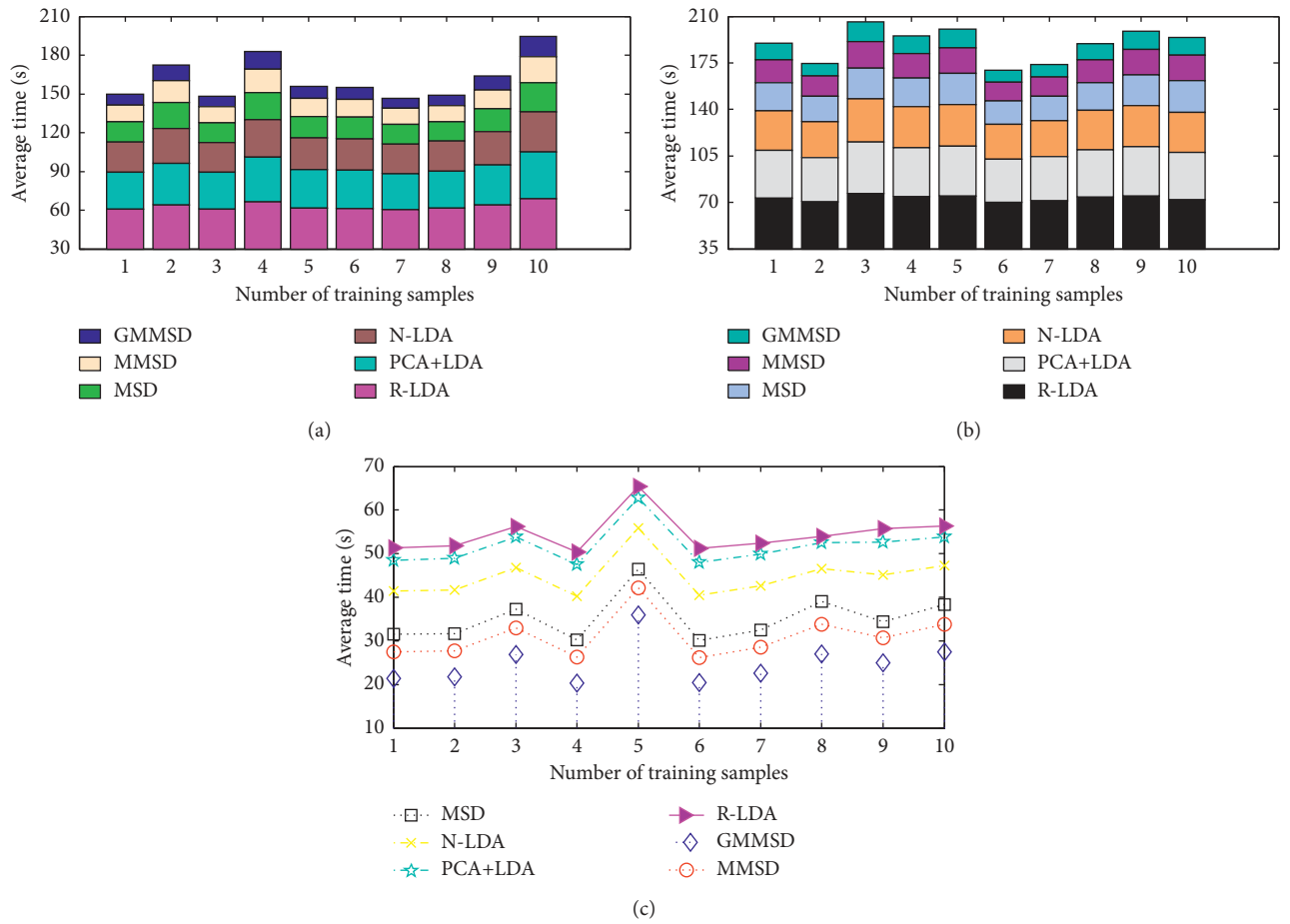


FIGURE 8: Comparison of computational complexity. (a) RML library. (b) FERET library. (c) Yale library.

6. Conclusion

This paper proposes a lightweight convolutional neural network (LW-CNN) based on a separable convolution structure, which integrates the advantages of a separable convolution structure to extract more accurate feature points with fewer parameters. Considering that the feature extraction method based on the difference criterion cannot extract effective discriminative features, the Generalized Maximum Dispersion Difference Criterion (GMMSD) and the corresponding feature extraction algorithm are proposed. Using QR decomposition can extract effective discriminative features for facial feature extraction, while also reducing the complexity of feature extraction. GMMSD avoids the “small sample” problem, there is no need to perform feature representation on the original face sample, and there is no need to perform preprocessing steps on the sample. It uses QR decomposition to extract the features of the original sample, and the original identification information is not lost after the sample is reduced in dimension, and the distribution characteristics of the original sample are well preserved. According to different change matrices, GMMSD can evolve into different feature extraction algorithms, indicating the generalized characteristics of the GMMSD algorithm. Experimental simulations on RML and Jaffe face libraries, as well as Yale, FERET, and AR face libraries, show that the GMMSD algorithm can effectively extract facial and facial discriminative features, and the recognition results obtained are higher than traditional subspace analysis algorithms and popular ones. The mapping algorithm in this article uses a facial transplant technology based on feature point difference vectors. This technology does not require high requirements on the acquisition equipment and can be implemented only by ordinary cameras, which can be applied to portable machines. However, this also directly affects the mapping effect and it is difficult to achieve the transplantation of facial effects such as wrinkles. In this regard, if you want to improve the application field of transplantation technology, how to solve the mapping effect of subtle faces is a point to be improved.

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 conflicts of interest.

Acknowledgments

This work was supported by the National Social Science Foundation Art Project: Comparative Study on Cultural Value Orientation and Communication Effect of Contemporary Tibetan Film and Television Works Domestic and Overseas (17CC188).

References

- [1] Z. Gao, Y. Li, and S. Wan, “Exploring deep learning for view-based 3D model retrieval,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 16, no. 1, pp. 1–21, 2020.
- [2] W. Zhang, C. Zhao, and Y. Li, “A novel counterfeit feature extraction technique for exposing face-swap images based on deep learning and error level analysis,” *Entropy*, vol. 22, no. 2, p. 249, 2020.
- [3] D. Mehta, S. Sridhar, O. Sotnychenko et al., “VNect,” *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1–14, 2017.
- [4] A. Woodward, Y. H. Chan, R. Gong et al., “A low cost framework for real-time marker based 3-D human expression modeling,” *Journal of Applied Research and Technology*, vol. 15, no. 1, pp. 61–77, 2017.
- [5] X. Ribes, “Is the YouTube Animation algorithm-friendly? How YouTube’s algorithm influences the evolution of animation production on the internet,” *Animation*, vol. 15, no. 3, pp. 229–245, 2020.
- [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. sp1, pp. 652–658, 2019.
- [7] A. Ojeda, K. Kreutz-Delgado, and T. Mullen, “Fast and robust Block-Sparse Bayesian learning for EEG source imaging,” *NeuroImage*, vol. 174, pp. 449–462, 2018.
- [8] S. Zhang, S. Zhang, T. Huang et al., “Learning affective features with a hybrid deep model for audio-visual emotion recognition[J],” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 10, pp. 3030–3043, 2017.
- [9] Z. Liu, Z. Lin, X. Wei et al., “A new model-based method for multi-view human body tracking and its application to view transfer in image-based rendering,” *IEEE Transactions on Multimedia*, vol. 20, no. 6, pp. 1321–1334, 2017.
- [10] T. Fukiage, T. Kawabe, and S. Y. Nishida, “Perceptually based adaptive motion retargeting to animate real objects by light projection,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 5, pp. 2061–2071, 2019.
- [11] B. Takacs, Z. Vincze, H. Fassold et al., “Hyper 360-towards a unified tool set supporting next generation VR film and TV productions,” *Journal of Software Engineering and Applications*, vol. 12, no. 05, pp. 127–148, 2019.
- [12] K. Ruhland, M. Prasad, and R. McDonnell, “Data-Driven approach to synthesizing facial animation using motion capture,” *IEEE Computer Graphics and Applications*, vol. 37, no. 4, pp. 30–41, 2017.
- [13] Y. Su, W. Sun, J. Liu, G. Zhai, and P. Jing, “Photo-realistic image bit-depth enhancement via residual transposed convolutional neural network,” *Neurocomputing*, vol. 347, pp. 200–211, 2019.
- [14] A. S. Al-Shamayleh, R. Ahmad, M. A. M. Abushariah, K. A. Alam, and N. Jomhari, “A systematic literature review on vision based gesture recognition techniques,” *Multimedia Tools and Applications*, vol. 77, no. 21, pp. 28121–28184, 2018.
- [15] K. Somandepalli, N. Kumar, T. Guha et al., “Unsupervised discovery of character dictionaries in animation movies,” *IEEE Transactions on Multimedia*, vol. 20, no. 3, pp. 539–551, 2017.
- [16] Y. Gui, Y. Tian, D. J. Zeng et al., “Reliable and dynamic appearance modeling and label consistency enforcing for fast and coherent video object segmentation with the bilateral grid,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 12, pp. 4781–4795, 2019.
- [17] S. J. Park, “A bio-edutainment system to virus-vaccine discovery based on collaborative molecular in real-time with VR,” *Journal of The Korea Society of Computer and Information*, vol. 25, no. 6, pp. 109–117, 2020.

- [18] Q. Abbas, M. E. A. Ibrahim, and M. A. Jaffar, "A comprehensive review of recent advances on deep vision systems," *Artificial Intelligence Review*, vol. 52, no. 1, pp. 39–76, 2019.
- [19] C. Koniaris, M. Kosek, D. Sinclair et al., "Compressed animated light fields with real-time view-dependent reconstruction," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 4, pp. 1666–1680, 2018.
- [20] Y. Liu, D. Liu, and G. Yue, "BGMM: a body gauss-markov based mobility model for body area networks," *Tsinghua Science and Technology*, vol. 23, no. 3, pp. 277–287, 2018.
- [21] S. Park, S. Yu, B. Moon, S. Ko, and J. Paik, "Low-light image enhancement using variational optimization-based retinex model," *IEEE Transactions on Consumer Electronics*, vol. 63, no. 2, pp. 178–184, 2017.
- [22] T. Lucas and R. Abd Rahim, "The similarities and nuances of explicit design characteristics of well-received online instructional animations," *Animation*, vol. 12, no. 1, pp. 80–99, 2017.
- [23] M. Rojc, I. Mlakar, and Z. Kačič, "The TTS-driven affective embodied conversational agent EVA, based on a novel conversational-behavior generation algorithm," *Engineering Applications of Artificial Intelligence*, vol. 57, pp. 80–104, 2017.
- [24] B. Yang, Z. Li, Y. Sun, and E. Cao, "EM-FEE: an efficient multitask scheme for facial expression estimation," *Interacting with Computers*, vol. 32, no. 2, pp. 142–152, 2020.
- [25] J. Iskander, M. Hossny, and S. Nahavandi, "A review on ocular biomechanic models for assessing visual fatigue in virtual reality," *IEEE Access*, vol. 6, pp. 19345–19361, 2018.
- [26] F. Lamberti, G. Paravati, V. Gatteschi et al., "Virtual character animation based on affordable motion capture and reconfigurable tangible interfaces," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 5, pp. 1742–1755, 2017.
- [27] B. Ria, C. Fda, D. Bma et al., "Big data analytics and computational intelligence for cyber-physical systems: recent trends and state of the art applications," *Future Generation Computer Systems*, vol. 105, pp. 766–778, 2020.
- [28] J. Wen, J. Yang, B. Jiang et al., "Big data driven marine environment information forecasting: a time series prediction network," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 1, pp. 4–18, 2020.
- [29] A. Zg, L. A. Xin, A. Nw et al., "Towards secure and efficient energy trading in IIoT-enabled energy internet: a blockchain approach," *Future Generation Computer Systems*, vol. 110, pp. 686–695, 2020.