Computer Communications 199 (2023) 30–41

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

3D face recognition algorithm based on deep Laplacian pyramid under the normalization of epidemic control

Weiyi Kong [a], Zhisheng You [a,b], Xuebin Lv [b,*]

[a] National Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University, Chengdu, 610065, PR China
[b] School of Computer Science, Sichuan University, Chengdu, 610064, PR China

ARTICLE INFO

Keywords:
3D face recognition
Multimodal fusion
Face reconstruction
Deep learning
Epidemic control

ABSTRACT

Under the normalization of epidemic control in COVID-19, it is essential to realize fast and high-precision face recognition without feeling for epidemic prevention and control. This paper proposes an innovative Laplacian pyramid algorithm for deep 3D face recognition, which can be used in public. Through multi-mode fusion, dense 3D alignment and multi-scale residual fusion are ensured. Firstly, the 2D to 3D structure representation method is used to fully correlate the information of crucial points, and dense alignment modeling is carried out. Then, based on the 3D critical point model, a five-layer Laplacian depth network is constructed. High-precision recognition can be achieved by multi-scale and multi-modal mapping and reconstruction of 3D face depth images. Finally, in the training process, the multi-scale residual weight is embedded into the loss function to improve the network's performance. In addition, to achieve high real-time performance, our network is designed in an end-to-end cascade. While ensuring the accuracy of identification, it guarantees personnel screening under the normalization of epidemic control. This ensures fast and high-precision face recognition and establishes a 3D face database. This method is adaptable and robust in harsh, low light, and noise environments. Moreover, it can complete face reconstruction and recognize various skin colors and postures.

1. Introduction

Under the background of normalization of epidemic prevention and control, it is the focus of epidemic prevention work to do a good job of monitoring personnel flow. Some countries have introduced the policy of removing masks. Therefore, it is particularly critical to realize the non-contact, efficient and safe management of people or people in public places. Among them, there are two most serious problems. First, how to use the method of quickly recognizing faces in public places, you can confirm your identity, and get information about your health and itinerary. Second, how to use scientific and technological means to obtain information about people closely linked in time and space, and timely screen to avoid a large-scale outbreak of the epidemic [1]. To solve the above problems, face recognition and reconstruction algorithms are listed as important research directions by researchers. Because three-dimensional face recognition and three-dimensional reconstruction have many advantages, such as non-contact, non-sensory recognition, good range, and safety. Therefore, it can provide an effective method for the normalization of epidemic prevention and control, and effectively avoid the risk of cross-infection of bacteria and viruses caused by physical contact [2]. Concerning user privacy, first of all, all data sets used in this algorithm research are public data sets for research. Secondly, when the algorithm is applied to the 3D vision face recognition module, the collected face images only record the face feature values. Face image data will not be saved, so users do not have to worry about privacy leakage at all. Therefore, under the background of the normalization of epidemic prevention and control, it is essential to realize face recognition of different skin colors and angles quickly, efficiently, and accurately in the face of many scenes without masks. This paper proposes an innovative three-dimensional face recognition algorithm based on the deep Laplacian pyramid. It can be used for rapid face recognition, and reconstruction in the normalization of epidemic prevention and control and is helpful to obtain their identity and health information efficiently. Because this paper focuses on the high-precision recognition and reconstruction of three-dimensional faces, the research in the field of face recognition is mainly analyzed below.

Face recognition is a biometric identification technology based on facial feature information for identity authentication. The development of this technology mainly relies on deep learning, 3D face recognition, and ultra-low resolution face [1–3]. Generally, face recognition is divided into two-dimensional face recognition and three-dimensional face recognition two categories. 3D face recognition is relative to 2D face recognition. The data used in two-dimensional face recognition is a two-dimensional image, which is essentially a projection of a three-dimensional object in a two-dimensional plane. Because the face itself

* Corresponding author.
E-mail address: lvxb@scu.edu.cn (X. Lv).

https://doi.org/10.1016/j.comcom.2022.12.011
Received 22 June 2022; Received in revised form 2 December 2022; Accepted 6 December 2022
Available online 13 December 2022

has three-dimensional attributes, so the use of three-dimensional face data for face recognition has more advantages [4–6]. 3D face data can be estimated from 2D color images or obtained directly by 3D imaging equipment. That is to say, face recognition based on 3D is the mainstream research direction [7,8]. The main research of this paper is also based on 3D face recognition.

However, previous studies on 3D face recognition and 2D face recognition are often independent. Traditional two-dimensional face recognition mainly uses mathematical methods to extract corresponding features from the image matrix, which generally scales invariant features. The commonly used algorithms include SURF, SIFT, HARRIS, GFTT, and so on. 3D face recognition processing is 3D data, such as point clouds, voxels, etc. These data are complete, three-dimensional, and can express the facial features of objects from various angles. The processing methods and processes are similar in two and three dimensions. The difference is that the data to be processed is different. Considering the limitations of two-dimensional face recognition and the difficulty of three-dimensional face recognition. In this paper, the data fusion strategy is innovatively adopted to combine two-dimensional data processing and three-dimensional data analysis. Using the idea of residual iteration, 2D and 3D features of the human face are fused without loss of network speed. The feature information of different levels of the image is highly correlated. In the network model design, the Laplace function is skillfully used to construct a pyramid structure, which is densely connected by residual modules. Each level is no longer a single independent distribution. It is a compact architecture. Based on this, our algorithm has higher recognition results. The generalization works better. Moreover, it can tolerate more postures, expressions, and skin tone changes.

Facial recognition outcomes are often analyzed at the data level. But this approach is not suitable for subjective visualization. Some researchers [6,9] have proposed that facial recognition and reconstruction should not be completely independent of one another. Both perform matching calculations on features such as pixel and texture, so the recognition and reconstruction fields are also similar in principle. Inspired by the above, we use facial recognition and recon visualization for multimodal fusion. It aims at improving recon performance and reconstruction visualization. Furthermore, the traditional research of three-dimensional face recognition and face reconstruction is often independent of each other. Resulting in poor face pose diversity recognition. Face reconstruction research has its limitations. In addition, light noise and different facial skin colors and expressions have significant interference with 3D recognition, resulting in poor accuracy.

Under the background of normalization of epidemic prevention and control, in order to monitor the flow of people in real time. As well as the timely tracking of people infected in time and space in case of an outbreak. This paper proposes a five-layer Laplacian pyramid network structure to solve the above problems. This method is a 3D face reconstruction method based on depth residual matching, and has high accuracy in 3D face recognition. The problem of effective reconstruction of 3D face data under occlusion and high reflection conditions is solved. It solves the problem of multi-angle and multi-pose face reconstruction. Multi-mode end-to-end reconstruction strategy based on disparity matching. Through multi-dimensional matching and feature fusion, the problem of 3D face reconstruction without image feature information is effectively solved, and it has high robustness to the environment. Thanks to the real-time performance of non-inductive recognition and reconstruction of 3D algorithm. To a large extent, we guarantee the freedom of personnel, and the algorithms and research can be used to monitor the epidemic prevention and control, so as to better maintain social stability.

Under the background that COVID-19 needs to wear a mask, the existing face recognition algorithms need to locate the features of different local information of the face, or need to learn the global face as a whole, so they generally cannot recognize the face accurately. Based on this problem, our proposed algorithm not only pays attention to global face recognition, but also introduces learning factors with different resolutions. More attention is paid to the local feature recognition of the face, and the complete face reconstruction can be carried out, which makes a new direction for the high-precision face recognition and reconstruction work under the condition of wearing a mask.

The main contributions of the algorithm in this paper include:

1. Innovatively proposed multi-pose stereo face detection and dense alignment. The three-dimensional graph structure models of the front and side key points are established to provide a high-precision model for the following 3D face reconstruction algorithm.
2. Innovative proposed a lightweight end-to-end Laplacian 3D face recognition reconstruction network. It integrates 3D face recognition and reconstruction visualization into the entire network architecture. The loss function of residual design is introduced based on Laplace network architecture. The joint dense fusion strategy is used to reconstruct the disparity map of face structure image and face texture image, which improves the accuracy and speed of the algorithm to a certain extent.
3. This paper combines the relationship between face recognition and expression reconstruction to form a complete algorithm. It has universal applicability. It can be applied to face reconstruction and scene recognition in low light and low texture. It has stable and efficient performance for three-dimensional face reconstruction and recognition with different skin colors, expressions, and postures. It can be applied to practical projects.

2. Related work

As mentioned above, the main problems that our method can solve are 3D face recognition and visualization representation. We discuss the work closely related to these tasks in the following sections.

2.1. Traditional 3D face recognition

According to the source of 3D face data, the 3D face recognition method is divided into the 3D face recognition method based on a color image, the face recognition method based on high-quality 3D scan data, face recognition method based on a low-quality RGB-D image [10,11]. Among them, the 3D face recognition method is based on color images and included in the use of 3D face model parameters for face recognition, using a 3D face model to synthesize a new human face image recognition [9]. Blanz [12] uses the model's 3D geometric distortion for facial recognition. This method from a single image estimates the geometry and 3D texture parameters. Second, the Euclid distance of these parameters is used to determine if they belong to the same person. This method uses the benefits of the 3D facial model and gets a better recognition effect under specific conditions. However, the drawback is that it is highly affected by lighting and the calculation is very large [13].

A multi-pose 3D facial recognition method based on the combination of 3D geometry and local analysis is proposed [10,11,14]. To improve the recognition accuracy, this method could be divided into the different local 3D faces. Reconstruction by the method of 3D geometric face locally, different parts of the geometric parameters, and texture combination as recognition characteristics. According to the identification of each component to determine its weight in the overall classification, achieve the result that improves recognition constructed a statistical model for 3D face modeling and recognition [15]. By matching BFM with color images, the corresponding attitude and illumination settings can be obtained. At the level of recognition, the facial identity coefficients of the different models can be directly compared. However, BFM's limitation is that it can only model neutral faces [16,17], therefore it is not suitable for facial images with expressions. For the face recognition method based on high-quality 3D scanning data,

which respectively includes the global feature method and local feature method in the global method, the face recognition method based on 3DMM and the face recognition method based on the curve are very classic. Papath [18] represented the face as a four-dimensional point set, and the elements in the set were composed of the three-dimensional coordinates of a point, and the gray level of the corresponding points in the two-dimensional image. ICP was suitable for matching between rigid surfaces, but the face was not a rigid surface, so some methods only used the areas less affected by expressions for recognition. Chang et al. [19] selected the nose region for ICP registration and recognition. Faltemier et al. [20] divided the face into 28 overlapping regions. Mohammadzade et al. [21] first found the nose tip of a three-dimensional face and intercepted a certain range of face regions according to the nose tip. Then, according to a reference face model, the iterative nearest neighbor normal point method is used to find the nearest neighbor point set corresponding to the reference model for each 3D face data. Finally, the normal vector of these points is used as the feature for face recognition ICP face registration, and the average distance of matching points is generally used as the similarity measure of two faces. Since the average distance is greatly affected by noise points, Hausdorff distance is used as the similarity measure of the two-point sets [22]. To reduce the influence of noise points and 3D point cloud sampling differences on ICP registration, also some researchers [23,24] used sparse ICP combined with a resampling method for registration and achieved a good recognition effect.

Curvilinear method Drira et al. [25] used the radial curve based on the nose tip to represent the whole face surface. In this method, the nose tip is first located, and then the face surface is segmented by a plane passing through the nose tip at every certain Angle. The plane and face surface intersection line is a radial curve. Some radial curves are discontinuous or too short due to occlusion or missing surface data, and these radial curves are discarded. The rest of the curve is used for face recognition. The distance between curves was obtained by elastic shape analysis. Lei et al. [26] also defined a curve at every certain Angle with the tip of the nose as the starting point. According to the curve, the face depth map is sampled, and the depth value of the sampling point is composed of a vector. To reduce the influence of expression, only the features of the upper half of the face were extracted. Then the kernel principal component analysis was used to map the ARS to the high dimensional feature space. Finally, a support vector machine is used for face recognition.

To sum up, the above research on 3D face recognition is often one-sided. For a single face feature image processing. This does ensure the efficiency of the network model, but it loses a lot of other features. More facial features cannot be learned, which affects the accuracy of 3D recognition. For different skin colors, and different gender, multi-pose, and big expression recognition effect is not ideal.

2.2. 3D face representation

3D face reconstruction methods have been rapidly developed into practical applications. Prime examples include VR/AR, 3D avatar creation, video editing, image synthesis face recognition, virtual makeup or voice-driven facial animation Refs. [27–29]. To produce the problem manageable, most existing methods combine prior knowledge of geometry or appearance by using pre-computed 3D face models. These models [30–32] reference rough human face shapes but fail to capture geometric details. Like emotion-dependent wrinkles, which are crucial for the authenticity and support analysis of human emotions. There in common are several effective ways to instantly recover detailed facial geometry, however, they typically require high quality training scans or lack robustness to occlusion. None of these academic studies looked at how wrinkles change with creative expression.

Before typically learning, it traditionally relies on the expression method of texture features, and high-precision 3D scanning is generally used as training data. So you cannot maintain an unconstrained image. To solve the above problems, deep learning research rises in the whole scientific field. Two-dimensional identification and reconstruction cannot meet the practical application of engineering. Therefore, 3D face recognition and reconstruction are critically important [33]. However, the apparent lack of 3D face data and the considerable difficulty of modeling pose challenges to the reconstruction task. Scientifically based on this, 3D face reconstruction methods have been developed rapidly. To simplify the processing process, most of the existing methods rely on a 3D face model based on prior calculations. These models combine prior knowledge of geometry and appearance and can reconstruct rough face shapes, but cannot capture geometric details of the face, so it is difficult to reconstruct a multi-pose face state. There are moreover possible ways to miraculously restore smooth facial shapes. However, they typically require multiple high-quality scanning devices for multi-angle scanning and typically lack the robustness of occlusion reflection scenes.

According to the above research results, some key problems of face recognition can be summarized. Face recognition and reconstruction are frequently studied independently. 3D recognition and 2D recognition are not combined effectively. But essentially, the 3D data of the image is directly correlated with the 2D data. Therefore, multimodal fusion of 2D and 3D data is carried out in this paper. The two are no longer single and independent. This prevalent method typically has better face recognition effect. Better data generalization performance is guaranteed. And can tolerate more facial expressions and creative expressions, as private well as skin color. The research of face recognition and reconstruction method develops no integral problem. This paper typically presents an innovative design of Laplacian pyramid network structure. It typically aims to carefully construct end-to-end dense connection through iteration of residuals on different feature layers. There are five layers of loss functions for different data feature dimensions. Such network design effectively and cleverly integrates two-dimensional information and three-dimensional features of the face. The research of the two-dimensional face and three-dimensional recognition is no longer distributed independently. And the geometry structure of the face is learned by the clever 3D face point detection assisted network. It can improve the prediction accuracy of the algorithm and ensure the speed. Moreover, our algorithm has excellent recognition and reconstruction results for face images of different skin color, different posture and diverse race. Specific network design and loss functions are described in Chapters 3.

3. Proposed method

First of all, the rapid detection and positioning of faces in an extensive background is the key to determining whether the key point model can be effectively established. After that, the data pre-processing of two-dimensional face image needs to extract face key points under multiple positions, three-dimensional face reconstruction and recognition, and multi-angle facial three-dimensional key points on the front and side. The rapid cascade of human face detection and the capture of three-dimensional key points guarantee the complete reconstruction and recognition of multiple postures and human faces. So we study a deep 3D face recognition algorithm for Laplace pyramid matching. Our algorithm is proved in the face orientation Angle (−100 to +100), and densely connected structure of face features matching and computational reconstruction. It is worth mentioning that our network using Laplace pyramid architecture for depth estimation can ensure that the global features simultaneously, through the Laplace operator to capture and store local information, aims to reconstruct high precision face depth map for 3D recognition.

3.1. Acquisition of facial stereoscopic key points

First, a model is established, a deformable shape instance S is defined $S = [x_1, y_1, \ldots, x_L, y_L]^T$, and N images are trained, including

L feature points. Where, the coordinates are expressed as $(x_i,\ y_i)$, $\forall i = 1, \ldots, L$. Then, generalized Purkuk analysis and principal component analysis were used to extract the alignment and orthogonal basis, and the shape examples represented by mean and feature vectors were obtained by changing the plane rotation and shift parameters of proportional parameters to enhance the training data.

$$S_p = \overline{S} + U_P \tag{1}$$

Where U represents the orthogonal basis of n eigenvectors, $\bar{S}$ represents the mean value of shape vectors, and $p = [p_1, p_2, \ldots, p_n]^T$ represents the shape parameter vectors.

The following definition of facial expression (texture) model for deformation function $W(P)$, in the actual data for the general texture deformation through the definition of face texture, face appearance model set feature equation function F, used to extract human to image features, after the feature deformation for reference model and vectorization.

$$A = F(I_i)(W(P_i)), \quad \forall_i = 1, 2, \ldots, N \tag{2}$$

and then conduct principal component analysis on A to establish the example of gauss appearance:

$$A = \overline{A} + UC \tag{3}$$

where A is the appearance parameter vector. Formula (1) and formula (3) describe the change of shape and appearance.

When given a training set with N images and L feature points in each image, S can align the feature points in the image (move the feature points to the position of the reference model through affine transformation), and adjust the position of each feature point $(x_i,\ y_i)$, $v_i = 1 \ldots N$, And the average value was taken to obtain the results in $x = [x_i^1, \ldots, x_i^L]^T$, $y = [y_i^1, \ldots, y_i^L]^T$ where U can be obtained by taking the average value of the training shape s1. sl and s the difference between the average shape PCA after operation, and take the corresponding eigen value's largest first n characteristic vectors are usually the eigenvalues of the eigenvectors corresponding to the total energy accounted for more than 90% of the total energy of adjusting the shape parameter vector p, equivalent to adjust the shape of the different characteristic vector expression of the weight, can make the shape of the different instances, as shown in Fig. 1.

The average shape is at the far left of the shape model instance. It also includes the side and depression results of 3D face modeling obtained by randomly generating the weights of the first five feature vectors within the range of [−3, 3], as shown in Fig. 2. The establishment of three-dimensional point model is required to fit the image feature residuals generated by the minimum. So we can represent as below. And the 3D point model we show as in Fig. 1. The three-dimensional modeling formula based on three-dimensional key points is shown as follows.

$$\arg\min_{p,c} \left\| t(w(p) - \overline{a} - U_a C \right\|^2 \tag{4}$$

3.2. Establish a dense multi-dimensional face alignment

In the field of vision research of face reconstruction, face alignment is the key to the quality of reconstruction and recognition [34–36]. Among the early methods, there are many alignment methods based on two-dimensional facial markers [37,38], such as local model, two-dimensional calibration alignment based on neural network and so on [18,21,22]. However, the limitation of the traditional method is that it can only return to the visible area of the face, which leads to the inability to effectively express faces in different poses and environments.

In order to solve these problems, this paper proposes a multi-pose alignment model framework for 3D faces. A basic model is established from a two-dimensional facial texture image by stereo point fitting. Then, the spatial coordinates are reconstructed in the three-dimensional geometric legend by the method based on three-dimensional reconstruction. The dense alignment of three-dimensional faces can be efficiently realized. Firstly, to ensure the semantics of the position stereo feature points, we establish a three-dimensional coordinate system based on the facial stereo key points and the facial texture map. After that, we establish densely connected face alignment. Specifically, the two-dimensional texture map structure and key stereo points are used as constraints for network training. Construct the geometric structure of two-dimensional to three-dimensional faces. This effective scheme can obtain three-dimensional feature estimation parameters for stereo alignment in high-dimensional space. Therefore, our method does not need complex parameters such as distortion parameters and refractive index, which significantly speeds up the network's overall training architecture and recognition rate. Based on the above research, we built a 3D geometric model of face texture image through intensive alignment of key points, which was used for supervision constraints of the Laplacian pyramid network. The auxiliary network learned faster and more feature visualization results were shown in Fig. 2.

Fig. 1. Facial key points of front and side faces.

3.3. Laplace Laplace pyramid network

We show the main algorithm structure in Fig. 3, the two-dimensional rectangle represents the two-dimensional convolution. Among them, the yellow two-dimensional convolution is mainly used to build the key feature point model of human face. Three-dimensional cube represents the cost volume of three-dimensional convolution. Where the green three-dimensional cube represents the cost volume of low and middle dimensions. The red cube represents the high-dimensional cost volume. The specific algorithm derivation and calculation process are as follows.

First, according to the given two-dimensional image as the input, after the establishment of the three-dimensional point model, the rough reconstruction is carried out to encode the image into a cost body, and then the residual is calculated through the Laplace pyramid network architecture to form a low-frequency and middle-frequency matching face depth image, and finally, through multi-mode fusion into a high-frequency complete depth map, The encoder we trained includes a ResNet network as a generator of low-frequency disparity and a three-dimensional pyramid network architecture. The specific calculation process is as follows: For the Laplace residual of the input face image, it is expressed as L_f.

$$L_f = I_f - U_P(I_{f+1}), \quad f = 1, 2, 3 \tag{5}$$

Where f represents the horizontal exponent of the Laplacian pyramid. If represents the U_P obtained by downsampling the original input $\frac{1}{2^{f-1}}$ below let R_f be the residual depth of the pyramid network The depth residual contains the rough geometric features and facial expression features, as well as the information of the three-dimensional key points of the face. Through the series of multiple modules, the pixel-level stack is carried out to accurately restore the local details of

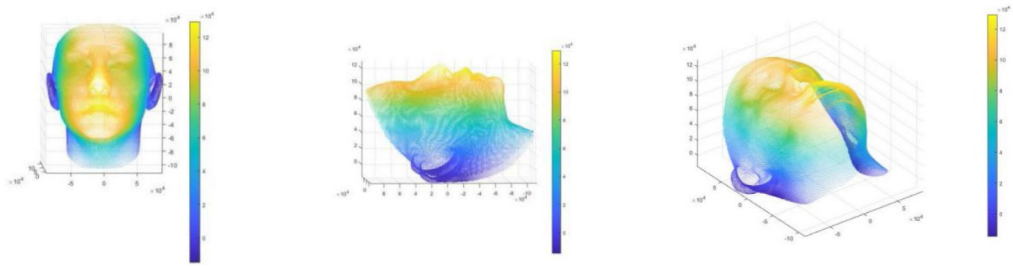

Fig. 2. A 3D face alignment model with different angles of dense alignment.

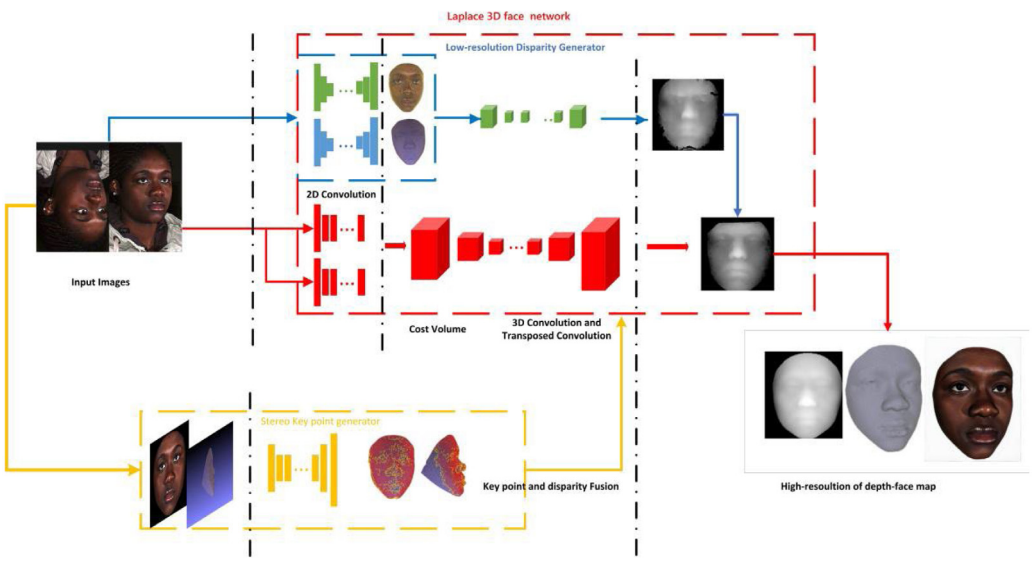

Fig. 3. Laplacian pyramid net structure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

each scale in the decoding process. Finally, the depth image is output through the highest sub-link of the Laplace pyramid:

$$D_f = R_f + U_P(D_f + 1), \quad f = 1, 2, 3 \tag{6}$$

It can be seen that, through the residual updating iteration of the tower structure, the network can well predict the depth image of the 3D face. To make the decoding process of depth images more efficient, we add the operation of weight standardization in the activation of the convolution cost body. It is worth mentioning that to robustly estimate the depth of information. We design the decoder can completely in the course of backpropagation gradient normalization, thus improving the gradient flow. Among them, the backpropagation is calculated in each layer of the Laplace pyramid. This can ensure that the residual error information of the color depth of translation is the stability and ideal. We plotted Tables 1 and 2. The input and output of low frequency and high frequency are given, respectively. And the specific parameters of low frequency and high-frequency disparity volume.

3.4. Loss

We use face images with texture information to train our model with supervised learning. It contains 5 terms of Laplacian pyramid loss functions. The first term aims at training the coarse reconstruction low-resolution disparity. Same as, we use the smooth L_1 loss between the texture images D_n and the predicted low-resolution disparity $\widetilde{D}_n$ for each key point pixel n, which is widely used because of its low sensitivity to outliers and defined as

$$L_1 = \frac{1}{N}\sum_{n=1}^{N} smooth_{L_1}(D_n - \widetilde{D}_n) \tag{7}$$

in which,

$$smooth_{L_1} = \begin{cases} \frac{1}{2}\theta^2, & if \ \ |\theta| < 1 \\ |\theta| - \frac{1}{2}, & otherwise \end{cases} \tag{8}$$

where N is the number of the key point pixels. The second term is used to supervise the key facial 3D point edge refinement in face, which is defined as

$$L_2 = \frac{1}{N}\sum_{n=1}^{N} smooth_{L_2}(D_n, \widetilde{D}_n^r) \tag{9}$$

where $\widetilde{D}_n$ is the disparity depth value of pixel n in the refined low-resolution disparity depth map.

The third term focuses on training the high-resolution disparity facial depth generator. In other words, the training set for the high-resolution is dynamically generated. We use k to denote the set including the pixels where the absolute errors between the ground truth disparities and the low resolution disparities are larger than 1. Then we use the pixels in k to train the high-resolution disparity facial depth generator and the loss is defined as

$$L_3 = \frac{1}{|K|}\sum_{n\in k} smooth_{L_1}(D_n^g - \hat{D}_n) \tag{10}$$

where $|K|$ denotes the number of elements in the set k and $\hat{D}_n$ is the predicted high-resolution facial disparity for pixel n.

We adopt the fourth term to supervise the final disparity facial depth map. The loss is defined as

$$L_4 = \frac{1}{N}\sum_{n=1}^{n} smooth_{L_1}(D_n^g - D_n) \tag{11}$$

where D_n is the final disparity depth value of the pixel n.

Table 1
Low and high resolution disparity regression.

Critical Hierarchy	Input parameter	Output Tensor
Low-resolution	Conv 3D, 3 × 3 × 3, f = 64	$\frac{1}{8}D \times \frac{1}{8}H \times \frac{1}{8}W \times 64$
	Conv 3D, 3 × 3 × 3, f = 64, s = 2	$\frac{1}{16}D \times \frac{1}{16}H \times \frac{1}{16}W \times 64$
	Conv 3D, 3 × 3 × 3, f = 64(double)	$\frac{1}{16}D \times \frac{1}{16}H \times \frac{1}{16}W \times 64$
	Residual connection ×2	$\frac{1}{8}D \times \frac{1}{8}H \times \frac{1}{8}W \times 32$
	Transposed 3D conv, 3 × 3 × 3, f = 32, s = 2	$\frac{1}{4}D \times \frac{1}{4}H \times \frac{1}{4}W \times 1$
	Soft argmin	$\frac{1}{4}H \times \frac{1}{4}W$
	Upsampled low-disparity	$H \times W$
	Key-point refined low disparity	$H \times W$
High-resolution	Conv 3D, 3 × 3 × 3, f = F, s = 2	$\frac{1}{4}D \times \frac{1}{4}H \times \frac{1}{4}W \times F$
	Conv 3D, 3 × 3 × 3, f = 2F	$\frac{1}{4}D \times \frac{1}{4}H \times \frac{1}{4}W \times 2F$
	Residual connection ×2	$\frac{1}{4}D \times \frac{1}{4}H \times \frac{1}{4}W \times 2F$
	Transposed 3D conv, 3 × 3 × 3, f = F, s = 2	$\frac{1}{2}D \times \frac{1}{2}H \times \frac{1}{2}W \times F$
	Residual connection	$\frac{1}{2}D \times \frac{1}{2}H \times \frac{1}{2}W \times F$
	3 × 3 × 3, f = 1, s = 2	$D \times H \times W \times 1$
	Soft argmin	$H \times W$

Table 2
Summary of our low and high cost volume architecture.

Layer Description	Output Tensor
Input facial image	$H \times W \times 3$
High-resolution features	
Conv 5 × 5, features = 32, s = 2	$\frac{1}{2}H \times \frac{1}{2}W \times 32$
Conv 3 × 3 × 2, features = 32	$\frac{1}{2}H \times \frac{1}{2}W \times 32$
Conv 3 × 3, features = 32	$\frac{1}{2}H \times \frac{1}{2}W \times 32$
Residual connection	$\frac{1}{2}H \times \frac{1}{2}W \times 32$
Conv 3 × 3, features = 8	$\frac{1}{2}H \times \frac{1}{2}W \times 8$
High-resolution cost volume	$\frac{1}{2}D \times \frac{1}{2}H \times \frac{1}{2}W \times 16$
Low-resolution features	
Conv 3 × 3 × 3, features = 64, s = 2	$\frac{1}{4}H \times \frac{1}{4}W \times 64$
Residual connection	$\frac{1}{4}H \times \frac{1}{4}W \times 64$
Conv 3 × 3, features = 32	$\frac{1}{4}H \times \frac{1}{4}W \times 32$
Repeat features	$\frac{1}{8}H \times \frac{1}{8}W \times 32$
Low-resolution cost volume	$\frac{1}{8}D \times \frac{1}{8}H \times \frac{1}{8}W \times 64$

The last term addresses on the problem of how to automatically select the disparities between the low-resolution and high-resolution disparities. For each pixel, low resolution disparity is selected if the high resolution disparity is greater than the absolute error between ground and low resolution dis-parity. During the training, we dynamically mark the pixels of each face image. For each pixel n, if $\left\|D_n^g - \widetilde{D}_n\right\|_1 \leq \left\|D_n^t - \widetilde{D}_n\right\|_1$, we label the pixel n as $p_n = 0$. Otherwise, we label it as $p_n = 1$. Let y_{n_0} and y_{n_1} are the values of the pixel n in the two feature maps, respectively. Then we use a softmax loss to train. Then the loss is defined as

$$L_5 = -\frac{1}{N}\sum_{n=1}^{N}(1 - p_n)\log y_{n_0} + p_n \log y_{n_1} \tag{12}$$

The total loss function is represented as below.

$$L_{ALL} = L_1 + L_2 + L_3 + L_4 + L_5 \tag{13}$$

4. Experiments

The algorithm model is trained by dual-card NVIDIA 1080 TI GPU of Ubuntu 16.04 system, Pycharm and Matlab were used as visual platforms for training and modeling, respectively, which can be tested on GPU platforms.

In this paper, the Laplacian pyramid network is selected as the transfer learning model of the feature extraction subnet. The attenuation parameter is set to 5 × 10 − 4, and the learning rate and momentum parameters are set to 1e − 4 and 0.9 respectively. In addition, the random gradient descent method is used as a learning optimizer. By freezing the weight of the first K layer in the pre-training model, fine-tuning the network, and then training the remaining N–K layers to learn the weight and deviation of the unfrozen layer. Then, the optimal matching between frozen layers is found through continuous iteration, and the fine-tuning of network parameters is retrained according to the results. In order to intuitively show the 3D recognition capability of the proposed Laplacian pyramid network, the low-frequency features extracted from the first convolution block of the network are visualized. There are two convolution layers in the first convolution block, and each convolution layer has 64 filters. Therefore, we can generate 64 low-frequency feature maps for facial attribute mapping of each data type. In the second convolution block, there are two two-dimensional convolution layers. Through this operation, a 3D cost voxel is generated through this operation, which contains medium and high-frequency information, and then fused into the decoder through Laplace's decoding network. Multi-modal fusion of facial feature information in two-dimensional and three-dimensional levels is obtained by five-layer regression iteration and fusion of various facial shapes and features related to expressions. This convolution layer can extract facial features with higher dimensions and finally output 512 feature maps. The depth of redundant representations will be combined into a single compact facial feature. Through the above operations, we make the network model robust and efficient, and we propose Laplace 3D face recognition network. It can form a high-level and high-frequency face recognition representation and more effectively recognize and reconstruct three-dimensional faces.

4.1. Data set

The BU-3DFE (Binghamton University 3D Facial Expressions) [39] database is widely used for static/dynamic 3D facial expression recognition. This database is the benchmark database for 3D facial expression recognition research. It contains 100 subjects (56% female and 44% male) aged 18 to 70 years. In addition to neutral expressions, there were six basic expressions (happiness, disgust, fear, anger, surprise, and sadness), each of which contained four intensity levels. Each subject

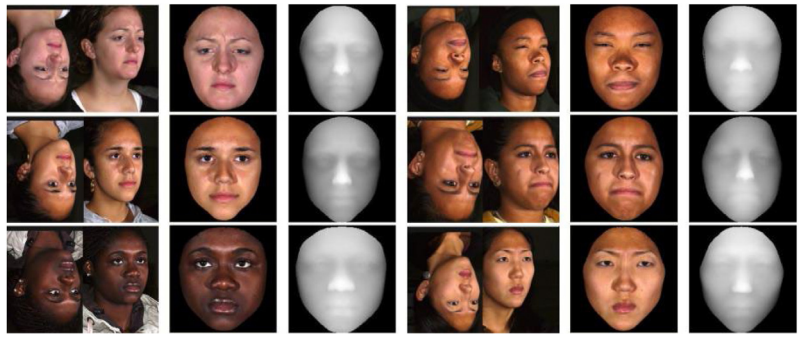

Fig. 4. The first and fourth columns represent the original input image. The second and fifth columns represent the output texture image. The third and sixth columns are the final high-frequency depth images that our network outputs. Through this group of images, it is confirmed that our proposed network framework can truly recognize 3D faces and reconstruct depth disparity.

had 25 3D facial expression models, and there were 2500 3D facial expression point cloud models and 2D facial texture images in the database.

Aflw2000-3D dataset [40] for evaluating challenging unconstrained 3D face alignment. The database contains the first 2000 images from the AFLW and is annotated with extended 3DMM parameters and 68 3D annotations. In the experiment, we add and use this database to evaluate performance on facial reconstruction and face alignment tasks.

Aflw-lfpa [41] is another extended dataset of AFLW. Images were extracted from AFLW according to attitude, and a test image dataset containing 1299 evenly distributed yaw angles was constructed. In addition, each image is tagged with 13 other landmarks, which are used to expand the only 21 visible landmarks in the AFLW. The database is evaluating tasks for 3D face alignment. We measured the accuracy of our results using 34 visible landmarks as ground truth values.

4.2. Evaluation index

There have been two types of face recognition tests: 1:1 face verification and 1: N face recognition. The 1:1 face verification test method used in this paper utilizes TAR as the ordinate and FAR as the abscissa. TAR@FAR = 1% has been used to evaluate the model's performance. The test dataset would include positively and negatively sample pairs. Positive sample pairs contain two images of the same individual, while antagonistic sample pairs are two images of two different people. The accurate acceptance rate is called TAR. The positive sample of the comparison score in the face verification process is greater than the set threshold value in the positive sample of the total.

FAR is the false acceptance rate, indicating the Proportion of the comparison score of antagonistic sample pairs in the total number of opposing sample pairs when the comparison score of opposing sample pairs is greater than the set threshold in the face verification process. TAR@FAR = 1% represent N antagonistic sample pairs, which are input into the network model for feature comparison. N comparison scores are obtained, and the value with the highest score is taken as the threshold. Then all positive sample pairs are compared in the input network model. The Proportion of the number of positive sample pairs with the comparison score more significant than the threshold in the total positive sample pairs is TAR@FAR = 1%.

We adopted the preferred recognition rate (Top1) for the Test method of 1:n face recognition to evaluate the model's performance. The depth Test data set in the table was taken as an example. The number of registered samples was 10000, the number of test samples was 2500, each test sample is compared with all samples in the registration set, and then the comparison score is sorted in descending order. If the tag with the highest score is the same as that of the test sample, it is the same person. Then the final Top1 is Rank1 by counting the number of Rank1 hits plus one Proportion of the hit number in the total test sample number.

When calculating the similarity score of two face features, Euclidean distance and cosine distance is usually used in face recognition. Euclidean distance directly calculates the distance between two points, and the calculation formula is shown below. Take the output features in this paper as an example. Since the output is a 512-dimensional feature vector, we use N to represent it. Thus get the distance of two face feature vectors. The smaller the distance, the closer the face.

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{14}$$

Where, d represents Euclidean distance. x represents the point in n-dimensional space. y represents the point in n-dimensional space.

The cosine distance is the two face features as two vectors, calculate the cosine value of the Angle, as shown in the formula below. When the Angle between the two vectors is closer to 0, the difference is smaller, the face is closer to the face similarity score in this paper is calculated by the cosine distance.

$$d(x, y) = \cos \frac{\vec{x} \cdot \vec{y}}{\|x\| \cdot \|y\|} \tag{15}$$

4.3. Qualitative evaluation

The Laplacian pyramid network of a single input image can be an excellent method to use 3D point geometry information and low frequency, medium frequency, and high-frequency information in the image to achieve realistic 3D face reconstruction. Our network architecture is shown in Fig. 3. The output of the comparison results can be seen in Fig. 4. More and more precise 3D face depth maps and texture map results are shown in Fig. 5. The visualization results of the reconstruction of different skin colors and races are displayed. Proved that our network architecture is an accurate 3D reconstruction. Moreover, it can complete 3D face recognition. Because compared with other methods, our network through continuous intensive iterative learning of The Laplacian pyramid and the supervision mechanism of three-dimensional key points. The smooth reconstruction of multi-pose and multi-view is realized. At present, the research results include PRNet [42] and 3DDFA-V2 [43], but the reconstruction results obtained by these methods are reconstructed after cutting surfaces. This creates artificial interference, and experimental results and effects are inconsistent. Our approach is to cascade output using an end-to-end network architecture. The experimental results are robust and reliable. The test of many data also confirms that our network is not sensitive to dark environments and low texture areas and has stable output performance.

It is worth mentioning that the general 3D face recognition task is heavy. The generic mapping is usually used for two-dimensional plane expansion. This operation for the two-dimensional face recognition effect is better and can better point matching estimation. Nevertheless, in the three-dimensional face task heavy, the effect is not ideal. For

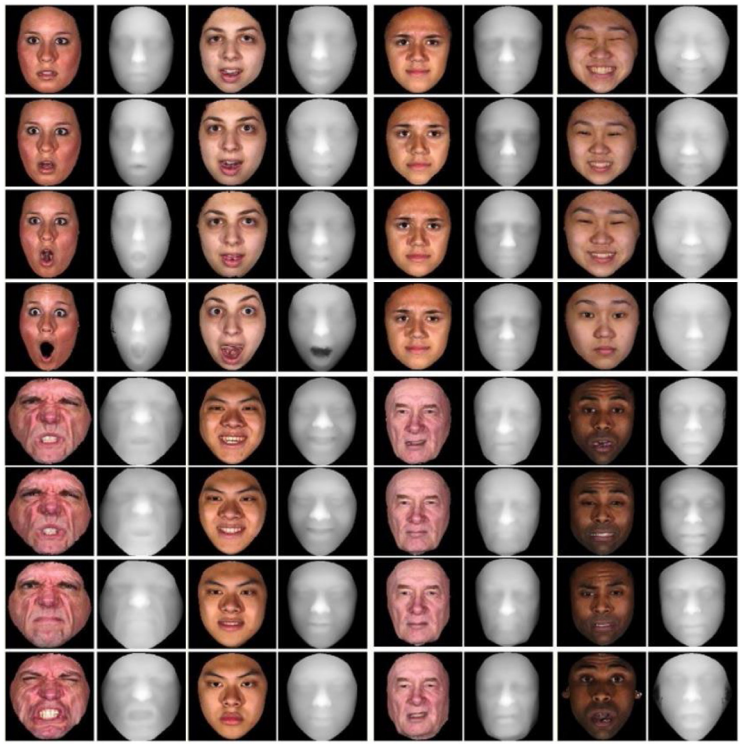

Fig. 5. This set of graphs shows the results we predicted for faces with different skin tones and different expressions. It contains output texture images of different tables and corresponding clear and smooth depth images.

incomplete symmetry of the face, the three-dimensional structure will appear distorted.

We were inspired by the merits of generic mapping after the low-frequency output of the Laplace pyramid network. We adopt the strategy of depth mapping and integrate the geometric modeling of 3D key points. Multi-pose, multi-angle 3D face recognition, and reconstruction are realized. In Fig. 6, the multi-pose and multi-angle image reconstruction effects of different input images. In Fig. 6, the first column represents the depth image of the depth map. It can be seen that the deep face in the figure is spread out as a generic mapping structure. The second column shows the depth display result of the profile face. The third column represents the lateral upper body depth image display results. The fourth column represents the frontal upper body depth image results. The fifth column represents the reconstruction result of the full-face depth image of the front. This is output by the high superimposed frequency.

4.4. Quantitative evaluation

We compare the Laplace Pyramid network with other open methods, including 3DDFAV2 [43], RingNet [44], PRNet [42], 3DMM-CNN [45], DECA [46] and Extreme3D [47]. Note that the Plath Pyramid network intensively implemented 3D reconstruction and validated the SOTA performance. NoW benchmark included 2054 face images of 100 subjects without human interference. Then the 3D data matching training was conducted, and the test group and the verification group were divided by 8 : 2. Each subject had a reference 3D face scan. The images included images of neutral expressions indoors and outdoors and images of expressions from different angles from front to side. Assessment of this data set provides a standard definition, after strict alignment scan and reconstruction, the measurement of all reference scan vertices to the triangle mesh surface recent distance. For gender bias of the test results, we report (W) women and men (M), respectively. NoW error of the test object found that to restore women more accurate shape. Reconstruction error is expressed as follows: median: 1.18/1.19 (W/M), average: 1.32/1.45 (W/M), and standard: 1.21/1.21 (W/M). The Laplacian pyramid network presents the most advanced NoW results, containing the lowest mean median and standard deviation reconstruction errors. In this way, it is proved that detailed high-frequency shapes can improve visual quality more than rough shapes. At the same time, to verify the performance of our network more comprehensively.

Table 3
Reconstruction error on the NoW benchmark.system of units:mm.

Method	Median	Mean	Std
3DMM-CNN [45]	1.84	2.34	2.07
PRNET [42]	1.51	1.98	1.89
SENet [48]	1.24	1.56	1.30
RING-NET [44]	1.20	1.54	1.32
3DDFAV2 [43]	1.24	1.58	1.39
MGC-NET [49]	1.32	1.89	2.70
DECA [46]	1.18	1.39	1.24
Ours	**1.15**	**1.38**	**1.21**

We find that the predicted artificial preprocessing of the clipped surface mesh is smaller than the present reference plane, which will lead to a high reconstruction error of the missing area. For a fair comparison, we use the Basel standard output of the Basel Face Model (BFM) parameters for a complete reconstruction. Furthermore, get these complete mesh evaluations now. As shown in Table 3, the most advanced current results are given. Reconstruction errors with the lowest mean, median, and standard deviation are provided.

The benchmark of Feng et al. [50], which consisted of 2000 face images of 135 subjects and a reference 3D face scan image of each subject, was selected for testing. The benchmark consisted of 1344 low-quality (LQ) images extracted from videos and 656 high-quality (HQ) images taken in control scenes. The Laplacian pyramid network provides state-of-the-art performance by measuring the distance from all reference scan vertices to the nearest point on the reconstructed mesh surface, as shown in Table 4. In order to verify the performance of

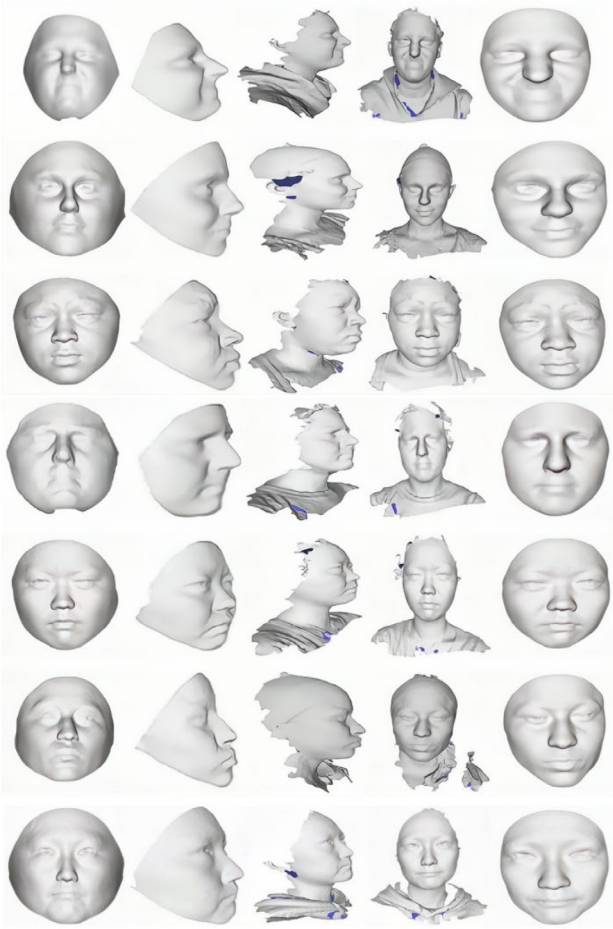

Fig. 6. Multi-view multi-angle reconstruction of the display map.

the error indicators in this paper more intuitively. A schematic diagram of the error area with color is presented. Among them, different colors show the error of different indicators. The smaller the area of error color map is, the better the network performance is. In Figs. 7 and 8, we show the median, mean, and standard error for different networks. Median, mean, standard error for men and women.

In addition, we show TOP1 accuracy and TAR@FAR = 1% accuracy of different networks through intuitive bar charts. Including Attention-Net [51], ECANet-K9 [52], ECANet-K3 [53], SENet [48], Resnet-34 [54] and our network. As shown in Fig. 9, in two accuracy assessments. Our algorithm is more accurate than other algorithms. It is verified that our algorithm has superior performance in 3D face recognition. We compared data and feature performance from different models. The evaluation indicators include TAR@FAR = 1% and Top1. Our network uses 3D face data for training. Feature analysis uses multi-level depth features. As shown in Fig. 9, the 3D recognition accuracy of our model fusion is higher than that of other networks. In Fig. 10, we visualized the recognition result of the 3D face. It can be seen that the output results of our algorithm can completely reconstruct the front and side images of the face. This is all done by entering a two-dimensional image of a plane. The validity of the Laplacian face recognition algorithm is verified. For different races, different skin colors, different postures, and different facial expressions, the visual reconstruction effect has good robustness. (See Table 5)

4.5. Ablation experiments

In order to verify the effectiveness of our network more specifically. Ablation experiments were introduced to verify the validity of each

Table 4
Performance on the Feng benchmark. system of units: mm.

Method	Median LQ	Median HQ	Mean LQ	Mean HQ	Std LQ	Std HQ
3DMM-CNN [45]	1.89	1.86	2.34	2.30	1.90	1.88
PRNET [42]	1.80	1.58	2.40	2.06	2.19	1.79
SENet [48]	2.40	2.38	3.44	3.5	6.10	6.76
RING-NET [44]	1.66	1.59	2.02	2.03	1.78	1.67
3DDFAV2 [43]	1.64	1.48	2.09	1.90	1.87	1.63
DECA [46]	1.58	1.49	1.90	1.87	1.67	1.69
Ours	**1.38**	**1.45**	**1.89**	**1.85**	**1.51**	**1.65**

Table 5
Comparison results of the models.

Modal	Dataset	Feature	TAR@FAR=1%	Top1
Tang et al. [55]	3D	Map	92.16%	87.10%
Song et al. [56]	3D	distance	93.08%	87.80%
Li et al. [57]	3D	normals	91.2%	82.01%
Wang et al. [58]	3D	curvature	88.60%	83.60%
Zeng et al. [59]	3D	curvature	79.63%	70.93%
Berretti et al. [60]	3D	depth/SIFT	85.56%	77.54%
Yang et al. [61]	3D	shape index	86.60%	82.30%
Li et al. [62]	2D+3D	meshHOG/SIFT	90.16%	86.32%
Ours	**3D**	**multi-scale deep feature**	**99.84%**	**93.64%**

layer of the Laplacian pyramid network. In Fig. 10, the analysis can be obtained. In the ROC curve, the Laplace fusion module, that is, the high-frequency fusion module has the highest 3D face recognition accuracy. Its ROC curve is smoother, and its area is more extensive than other curves. It has been proved that the performance of the Laplace fusion module is the best. Secondly, the performance of depth maps is better. The reliability of selecting a depth map for fusion is verified. The third performance is the key fusion module. For this reason, we use this method to supervise the whole network and assist the overall network model in learning more facial features faster. Then we tested the 3D face recognition performance of different network architectures. You can refer to the performance analysis in Fig. 11. To verify the effectiveness of each layer in the network. We have done ablation experiments on each layer of the network. The green line represents the final performance of the Laplacian fusion layer. It can be seen that its roc curve is very stable, and tends to be high and stable after FAR is higher than Where the pink line represents the intermediate frequency level of the Laplacian network. The blue line represents the lowest frequency level of the Laplacian network. The red line represents the performance of the 2D texture layer, and it can be seen that its performance is the lowest. The yellow line represents the performance of the key point model. In a comprehensive analysis, with the hierarchical superposition of low frequency, intermediate frequency, and high frequency of the Laplacian network, the final fused network performance is far superior to other output layers. The effectiveness and high performance of the overall design of the network model are verified. In Fig. 12, we show the comparison results between the Laplace network and other mainstream networks. Includes ECANet-K3 [53], ResNet,62, SENet [48], PRNet [42] comparison. It can be seen that the performance of Laplace's intermediate frequency layer and high-frequency output layer is much higher than that of other networks. The TAR value of the Laplacian high-frequency output layer represented by the green line is the highest when FAR is equal to 0.010. The pink line represents the lowest TAR value of ResNet. The comparative experiment proves that the final performance of our network is higher than other mainstream algorithms. We can see that the red Laplacian intermediate frequency model has the best effect in the initial ascent. After FAR is 0.002, it tends to be stable. The accuracy of the intermediate frequency model is above 0.9895. Green represents the Laplacian network model of high frequency fusion. Network performance did not improve quickly in the early days. This is because network models need to integrate multi-level learning. It can be seen that when FAR is 0.006, the performance of the high-frequency fusion

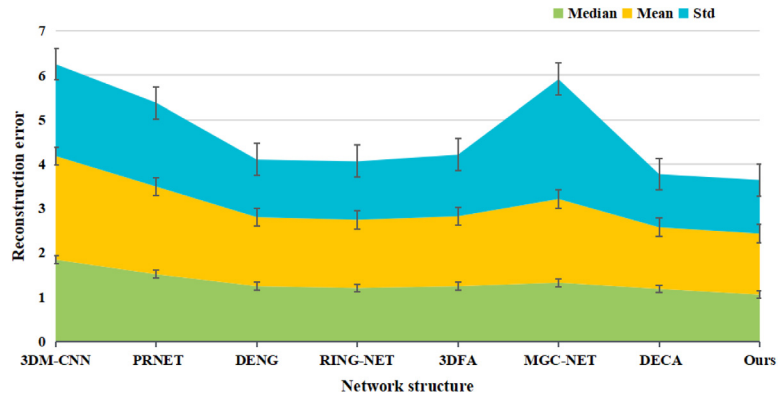

Fig. 7. This table show the different network and error map.

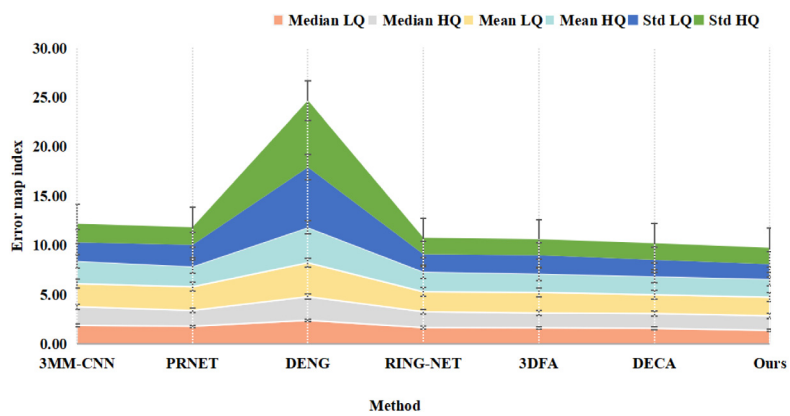

Fig. 8. This table show about different people facial and including different error map.

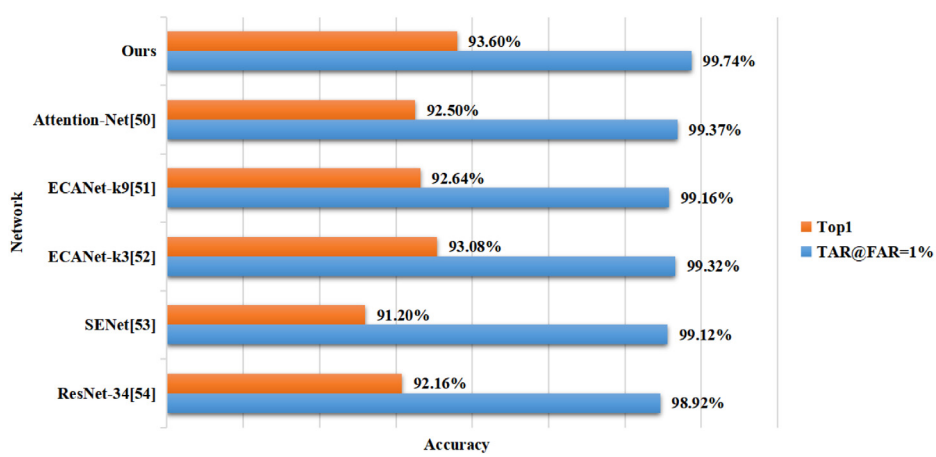

Fig. 9. This figure shows the comparison of two kinds of accuracy evaluation indexes under different network architectures. The first line is the accuracy index of The Laplace network, which is higher than the results of other methods.

model is significantly improved. The final 3D recognition rate is more than 0.9959. It is worth mentioning that our network performance is also the best in the comparison model. The ablation experiments above confirmed the effectiveness of our network. And it has higher efficiency of 3D recognition. It is necessary to observe the TAR value when FAR = 0.010. The larger the final value of TAR, the better the performance of the network. The green part is the highest frequency output of our network. The TAR value is the largest, and the performance of the network is the best. Moreover, while considering the network performance, we need to take into account the efficiency of the network. Therefore, the robustness of the model is also very stable.

5. Conclusion

This paper introduces the design of the Laplace end-to-end pyramid network. A convergence strategy for dense connection is adopted. The problems of 3D face registration, reconstruction, and 3D recognition are solved. By learning the 3D critical points of the human face, we can directly use regression to extract the complete 3D structure and semantic information from 2D images and then carry out 3D reconstruction and recognition. For other advanced algorithms, our work integrates reconstruction and 3D facial recognition. To ensure the quality of reconstruction and obtain the recognition effect. The quantitative and qualitative results show that this method is robust to

Fig. 10. This set of images shows the comparison results of our 3D visualization. The first and fifth columns represent the original input image. The second and sixth represent the output texture image. The third and seventh columns represent the output 3D frontal face image. The fourth and eighth columns represent the output side 3D face image.

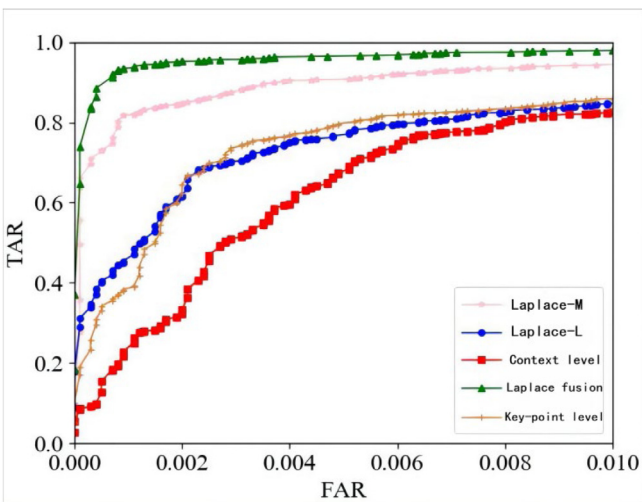

Fig. 11. ROC curves of different layers and multimodal fusion of the Laplace pyramid network. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

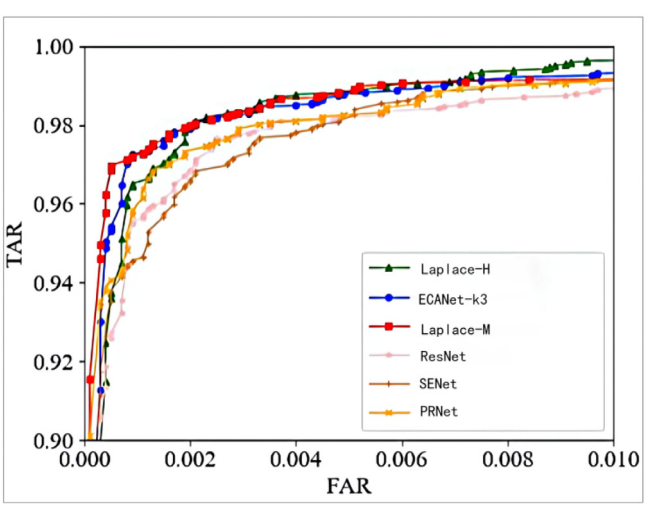

Fig. 12. ROC curves of different network structure and the Laplace pyramid network. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

pose, illumination and occlusion. Experiments on test data sets show that this method is superior to other methods. In addition, the recognition robustness of different facial poses, skin colors, and expressions is better. The experimental results show that this method is faster than other methods, has a higher 3D recognition rate, and is suitable for real-time applications. It can be used in COVID-19 epidemic normalization personnel management and has a good prospect for the establishment and engineering application of the 3D face database.

In future research, we will focus on using active, structured light to recognize and reconstruct 3D faces. The goal is to combine the functional, structured light theory with the current passive algorithm. A more universal and practical algorithm is studied. It will be used in facial recognition products to help prevent and control epidemics. Because face information involves privacy issues. We will consider introducing the federal learning technology in future work while training data and transmitting privacy with the control network model. Federal learning such as DEEP-FEL combined with the network model is adopted for training. Better control the privacy of user data.

Declaration of competing interest

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

Data availability

No data was used for the research described in the article.

Acknowledgments

This work is supported by the funding from National Natural Science Foundation of China (Grant No. U21B2035).

References

[1] S. Zhou, S. Xiao, 3D face recognition: A survey, Hum.-centric Comput. Inf. Sci. 8 (1) (2018) 1–27.
[2] M. Chihaoui, A. Elkefi, W. Bellil, C. Ben Amar, A survey of 2D face recognition techniques, Computers 5 (4) (2016) 21.
[3] D. Zeng, R. Veldhuis, L. Spreeuwers, A survey of face recognition techniques under occlusion, IET Biom. 10 (6) (2021) 581–606.
[4] A.T. Kabakus, et al., An experimental performance comparison of widely used face detection tools, 2019.
[5] L. Lang, W. Gu, Study of face detection algorithm for real-time face detection system, in: 2009 Second International Symposium on Electronic Commerce and Security, vol. 2, IEEE, 2009, pp. 129–132.

[6] M. Li, B. Huang, G. Tian, A comprehensive survey on 3D face recognition methods, Eng. Appl. Artif. Intell. 110 (2022) 104669.
[7] L. Shi, X. Wang, Y. Shen, Research on 3D face recognition method based on LBP and SVM, Optik 220 (2020) 165157.
[8] S.Z. Gilani, A. Mian, Learning from millions of 3D scans for largescale 3D face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1896–1905.
[9] S. Sharma, V. Kumar, 3D landmark-based face restoration for recognition using variational autoencoder and triplet loss, IET Biom. 10 (1) (2021) 87–98.
[10] Y. Deng, J. Yang, S. Xu, D. Chen, Y. Jia, X. Tong, Accurate 3D face reconstruction with weakly-supervised learning: From single image to image set, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.
[11] W. Zhu, H. Wu, Z. Chen, N. Vesdapunt, B. Wang, Reda: Reinforced differentiable attribute for 3D face reconstruction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 4958–4967.
[12] B. Gecer, S. Ploumpis, I. Kotsia, S. Zafeiriou, Ganfit: Generative adversarial network fitting for high fidelity 3D face reconstruction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 1155–1164.
[13] H. Goyal, K. Sidana, C. Singh, A. Jain, S. Jindal, A real time face mask detection system using convolutional neural network, Multimedia Tools Appl. (2022) 1–17.
[14] S. Sharma, V. Kumar, 3D face reconstruction in deep learning era: A survey, Arch. Comput. Methods Eng. (2022) 1–33.
[15] S. Jie, Q. Yongsheng, Multi-view facial expression recognition with multi-view facial expression light weight network, Pattern Recognit. Image Anal. 30 (4) (2020) 805–814.
[16] J. Gao, T. Yang, Research on real-time face key point detection algorithm based on attention mechanism, Comput. Intell. Neurosci. 2022 (2022).
[17] J. Deng, G. Trigeorgis, Y. Zhou, S. Zafeiriou, Joint multi-view face alignment in the wild, IEEE Trans. Image Process. 28 (7) (2019) 3636–3648.
[18] A.R. Bhople, S. Prakash, Learning similarity and dissimilarity in 3D faces with triplet network, Multimedia Tools Appl. 80 (28) (2021) 35973–35991.
[19] D.-P. Nguyen, M.-C. Ho Ba Tho, T.-T. Dao, Enhanced facial expression recognition using 3D point sets and geometric deep learning, Med. Biol. Eng. Comput. 59 (6) (2021) 1235–1244.
[20] R.A.A. Helmi, S.S. bin Eddy Yusuf, A. Jamal, M.I.B. Abdullah, Face recognition automatic class attendance system (FRACAS), in: 2019 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS, IEEE, 2019, pp. 50–55.
[21] S. Sharma, V. Kumar, Voxel-based 3D face reconstruction and its application to face recognition using sequential deep learning, Multimedia Tools Appl. 79 (25) (2020) 17303–17330.
[22] W. Hariri, N. Farah, Recognition of 3D emotional facial expression based on handcrafted and deep feature combination, Pattern Recognit. Lett. 148 (2021) 84–91.
[23] F. Wu, S. Li, T. Zhao, K.N. Ngan, L. Sheng, Cascaded regression using landmark displacement for 3D face reconstruction, Pattern Recognit. Lett. 125 (2019) 766–772.
[24] K. Wu, Z. Zhou, X. Yang, X. Wang, MDFN: Multi-path dynamic fusion network for face reconstruction and dense face alignment, in: 2021 China Automation Congress, CAC, IEEE, 2021, pp. 4438–4443.
[25] Y. Wang, T. Lu, F. Yao, Y. Wu, Y. Zhang, Multi-view texture learning for face super-resolution, IEICE Trans. Inf. Syst. 104 (7) (2021) 1028–1038.
[26] K. Jaswanth, D.S. David, A novel based 3D facial expression detection using recurrent neural network, in: 2020 International Conference on System, Computation, Automation and Networking, ICSCAN, IEEE, 2020, pp. 1–6.
[27] H.R. Afzal, S. Luo, M.K. Afzal, G. Chaudhary, M. Khari, S.A. Kumar, 3D face reconstruction from single 2D image using distinctive features, IEEE Access 8 (2020) 180681–180689.
[28] A. Dib, C. Thebault, J. Ahn, P.-H. Gosselin, C. Theobalt, L. Chevallier, Towards high fidelity monocular face reconstruction with rich reflectance using self-supervised learning and ray tracing, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 12819–12829.
[29] S. Cheng, G. Tzimiropoulos, J. Shen, M. Pantic, Faster, better and more detailed: 3D face reconstruction with graph convolutional networks, in: Proceedings of the Asian Conference on Computer Vision, 2020.
[30] K. Zeng, Z. Wang, T. Lu, J. Chen, B. Huang, Z. Han, X. Tian, Realistic frontal face reconstruction using coupled complementarity of far-near-sighted face images, Pattern Recognit. (2022) 108754.
[31] D. Zhao, Y. Qi, Generative landmarks guided eyeglasses removal 3D face reconstruction, in: International Conference on Multimedia Modeling, Springer, 2022, pp. 109–120.
[32] F. Jarad, T. Abdeljawad, Generalized fractional derivatives and laplace transform, Discrete Contin. Dyn. Syst. Ser. S 13 (3) (2020) 709.
[33] Y. Ji, K. Li, H. Wu, G. Xiong, Z. Shen, X. Shang, B. Xi, 3D face reconstruction system from a single photo based on regression neural network, IFAC-PapersOnLine 53 (5) (2020) 71–76.
[34] S. Aggarwal, R. Chaudhary, A comparative study of Mohand and Laplace transforms, J. Emerg. Technol. Innov. Res. 6 (2) (2019) 230–240.
[35] S.S.N. Bhuiyan, O.O. Khalifa, Efficient 3D stereo vision stabilization for multi-camera viewpoints, Bull. Electr. Eng. Inform. 8 (3) (2019) 882–889.
[36] H. Liu, Y. Cai, S. Zhou, J. Yang, Stereo matching with multi-scale based on anisotropic match cost, Concurr. Comput.: Pract. Exper. 32 (24) (2020) e5918.
[37] S.S.N. Bhuiyan, O.O. Khalifa, Robust automatic multi-camera view-point stabilization using harris laplace corner detection and spanning tree, in: 2018 7th International Conference on Computer and Communication Engineering, ICCCE, IEEE, 2018, pp. 1–5.
[38] L. Bertoni, S. Kreiss, T. Mordan, A. Alahi, Monstereo: When monocular and stereo meet at the tail of 3D human localization, in: 2021 IEEE International Conference on Robotics and Automation, ICRA, IEEE, 2021, pp. 5126–5132.
[39] X.-P. Huynh, T.-D. Tran, Y.-G. Kim, Convolutional neural network models for facial expression recognition using BU-3DFE database, in: Information Science and Applications, ICISA 2016, Springer, 2016, pp. 441–450.
[40] X. Zhu, Z. Lei, X. Liu, H. Shi, S.Z. Li, Face alignment across large poses: A 3D solution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 146–155.
[41] A. Jourabloo, X. Liu, Large-pose face alignment via CNN-based dense 3D model fitting, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4188–4196.
[42] H. Fu, S. Bian, E. Chaudhry, A. Iglesias, L. You, J.J. Zhang, State-of-the-art in 3D face reconstruction from a single RGB image, in: International Conference on Computational Science, Springer, 2021, pp. 31–44.
[43] Z. Yue, S. Ding, S. Yang, L. Wang, Y. Li, Multimodal information fusion approach for non-contact heart rate estimation using facial videos and graph convolutional network, IEEE Trans. Instrum. Meas. (2021).
[44] Y. Feng, H. Feng, M.J. Black, T. Bolkart, Learning an animatable detailed 3D face model from in-the-wild images, ACM Trans. Graph. 40 (4) (2021) 1–13.
[45] T. Shi, Z. Zuo, Y. Yuan, C. Fan, Fast and robust face-to-parameter translation for game character auto-creation, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, 2020, pp. 1733–1740.
[46] R. Danecek, M.J. Black, T. Bolkart, EMOCA: Emotion driven monocular face capture and animation, 2022, arXiv preprint arXiv:2204.11312.
[47] J. Shang, T. Shen, S. Li, L. Zhou, M. Zhen, T. Fang, L. Quan, Self-supervised monocular 3D face reconstruction by occlusion-aware multi-view geometry consistency, in: European Conference on Computer Vision, Springer, 2020, pp. 53–70.
[48] Z. Deng, R. Yang, R. Lan, Z. Liu, X. Luo, SE-IYOLOV3: An accurate small scale face detector for outdoor security, Mathematics 8 (1) (2020) 93.
[49] C. Chi, S. Zhang, J. Xing, Z. Lei, S.Z. Li, X. Zou, Selective refinement network for high performance face detection, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 8231–8238.
[50] S. Sanyal, T. Bolkart, H. Feng, M.J. Black, Learning to regress 3D face shape and expression from an image without 3D supervision, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 7763–7772.
[51] P.D. Marrero Fernandez, F.A. Guerrero Pena, T. Ren, A. Cunha, FERAtt: Facial expression recognition with attention net, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.
[52] Z. Shi, Z. Tan, Expression recognition method based on attention neural network, in: 2021 33rd Chinese Control and Decision Conference, CCDC, IEEE, 2021, pp. 856–860.
[53] X. Li, J. Zou, Pupilface: A cascaded face detection and location network fusing attention, in: Pacific Rim International Conference on Artificial Intelligence, Springer, 2021, pp. 426–437.
[54] S. Peng, H. Huang, W. Chen, L. Zhang, W. Fang, More trainable inception-ResNet for face recognition, Neurocomputing 411 (2020) 9–19.
[55] B. Hasani, M.H. Mahoor, Facial expression recognition using enhanced deep 3D convolutional neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 30–40.
[56] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, A.M. Dobaie, Facial expression recognition via learning deep sparse autoencoders, Neurocomputing 273 (2018) 643–649.
[57] M.M. Ahsan, Y. Li, J. Zhang, M.T. Ahad, K.D. Gupta, Evaluating the performance of eigenface, fisherface, and local binary pattern histogram-based facial recognition methods under various weather conditions, Technologies 9 (2) (2021) 31.
[58] L. Wang, A.A. Siddique, Facial recognition system using LBPH face recognizer for anti-theft and surveillance application based on drone technology, Meas. Control 53 (7–8) (2020) 1070–1077.
[59] X. Zeng, J. Huang, C. Ding, Soft-ranking label encoding for robust facial age estimation, IEEE Access 8 (2020) 134209–134218.
[60] Z.A. Serdar, A. Tatlıparmak, Comparison of efficacy and safety of fractional radiofrequency and fractional Er: YAG laser in facial and neck wrinkles: Six-year experience with 333 patients, Dermatol. Ther. 32 (5) (2019) e13054.
[61] H. Yang, U. Ciftci, L. Yin, Facial expression recognition by de-expression residue learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2168–2177.
[62] G. Li, X. Zhu, Y. Zeng, Q. Wang, L. Lin, Semantic relationships guided representation learning for facial action unit recognition, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 8594–8601.