RESEARCH ARTICLE

RP-Net: A PointNet++ 3D Face Recognition Algorithm Integrating RoPS Local Descriptor

YIBO CAO, SHUN LIU, PENGFEI ZHAO, AND HAIWEN ZHU

School of Software, South China Normal University, Guangdong 510631, China

Corresponding author: Yibo Cao (422705879@qq.com)

ABSTRACT As a biometric identification method in the post-epidemic era, face recognition owing more and more attention in practical applications to its non-contact and interaction-friendly advantages. Researchers more favor 3D faces because they have richer spatial information than 2D faces and are not easily affected by the environment. However 3D faces are not all collected in normal environments. To enhance the facial features of 3D faces and improve the recognition degree of 3D faces in weak-light or dark environments, a 3D face recognition algorithm based on point cloud depth learning is proposed. First, 3D faces are automatically detected from 3D raw data and preprocessed, including nose-tip detection and face cropping, spike removal and hole filling, and surface normals. Then, rotated projection statistical local feature descriptors (RoPS) are integrated into the PointNet++ network to describe and classify local features. Finally, feature matching is performed using the nearest neighbor distance ratio. The algorithm was tested on the Bosphorus and CASIA-3D datasets, and good results were obtained in a simulated weak-light environment.

INDEX TERMS PointNet++, RoPS, 3D face recognition, deep learning.

I. INTRODUCTION

With the increasing degree of information in current society, the related information security issues have received increasing attention, and all information security is ultimately inseparable from the authentication of personal identity. Whether it is personal privacy information, property security, or government confidential documents and management authority, it is necessary to authenticate the identity of the relevant people to ensure security. Traditional identity authentication methods such as certificates, passwords, seals, and cards have their disadvantages and hidden dangers, such as certificates, cards, and other authentication tools are easy to be damaged or lost, passwords are easy to be confused and forgotten, etc. Due to the advantages of reliability and convenience, emerging biometric recognition technology has incomparable advantages over traditional identity recognition and authentication technology, It has been widely concerned and used by the society [1], [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Zahid Akhtar.

Among many biometric recognition technologies, fingerprint and palmprint recognition have low costs and simple operations, but there are serious security risks due to easy forgery. Iris recognition has high accuracy and is not easy to be forged, but its high cost of acquisition and recognition equipment makes it unable to be widely used. Face recognition technology combines the advantages of other recognition technologies and has the advantages of simple acquisition, the advantages of safety and reliability, have wider application and practicability. Although the two-dimensional face recognition method has the advantages of friendly interaction, convenient acquisition, and low cost, it still has shortcomings in dealing with illumination changes, posture changes, anti-counterfeiting attacks, etc. [3], [4]. Compared with the above methods, 3D face recognition technology not only takes into account the advantages of 2D face recognition but also has its unique advantages. These advantages are reflected in the following points:

1) The collected three-dimensional shape data of the face can be regarded as not changing with the change of light and view, and accessories such as makeup significantly

impact on the image but have no obvious impact on the three-dimensional data. Therefore, 3D face recognition is considered to have the characteristics of constant illumination and pose [5].

2) 3D data has explicit spatial shape representation, so it is richer in information than 2D images [6].

In the existing traditional 3D face recognition methods and 3D face recognition methods combined with deep learning, the traditional 3D face recognition methods generally recognize global face features, such as principal component analysis, or local face features, such as using local feature descriptors to describe the local features of the human face, to achieve the purpose of face recognition. However, compared with the 3D face recognition method combined with deep learning, the traditional 3D face recognition method relies too much on the face alignment algorithm and feature descriptor, which limits the scalability.

Due to point cloud data's excellent performance in representing the position information and depth information of 3D images, the point cloud depth learning framework has shown remarkable performance in 3D object classification, semantic segmentation, and target recognition. PointNet [7] structure is the first structure that uses the symmetric function to aggregate point-to-point and order invariant features on the original point cloud. However, PointNet only aggregates global features and ignores local features, which are very important for the representation of 3D shapes. PointNet++ [8] applies PointNet hierarchically to capture local and global features for object recognition and segmentation on point clouds. We pay more attention to facial features, use the method of the surface normal in the preprocessing process, and propose to integrate RoPS local feature descriptor into PointNet++ to describe face features in more detail, and can be easily integrated into PointNet++. The improved PointNet++ extracts face features directly from the original face point cloud.

II. RELATED WORK
A. 3D FACE RECOGNITION
The existing 3D face recognition methods can be roughly divided into two categories: methods based on deep learning or traditional methods. The method based on deep learning is to train the three-dimensional face model through the neural network model to detect and extract the three-dimensional face features and then match the extracted three-dimensional face features with the target face to determine whether the two faces belong to the same person, to achieve the purpose of face recognition. Luo *et al.* [9] fused the two-dimensional face abstract features output by a convolutional neural network with the depth maps representing three-dimensional face information, and output the fused data through the full connection layer as the input of the classifier. Kangming *et al.* [10] proposed a method of 3D face recognition using depth dual neural network to fuse 3D depth and 2D texture. Unlike Luo *et al.* [9] and others who directly fused two-dimensional

face features with depth map, Kangming *et al.* [10] fused depth features extracted based on a convolutional neural network. Dutta *et al.* [11] proposed a lightweight deep learning network sppcanet for feature extraction and used a linear support vector machine (SVM) to classify the extracted features.

Based on the traditional method, face recognition is carried out by describing the feature information of three-dimensional face. According to the different types of features, face recognition can be divided into three methods: global-based, local-based, and hybrid.

Global-based methods usually recognize 3D faces as a whole feature vector. Russ *et al.* [12] proposed a general method of 3D face recognition based on principal component analysis. It avoids eliminating size information by scaling 3D reference to realize the alignment of principal component analysis training, synthesis, and recognition of critical facial points. Liu *et al.* [13] proposed a competitive method for 3D face recognition using the spherical harmonic feature (SHF). This method is based on the spherical harmonic feature (SHF), namely spherical depth map (SDM), reliably calculated using a standardized three-dimensional face representation. mohammadzade *et al.* [14] used the iterative nearest normal point method (ICNP) to find the nearest normal point between the ordinary reference plane and each input plane. These points are effectively aligned on all faces so the discriminant analysis method can effectively apply to 3D face recognition.

The local-based method first detects landmarks or representative facial regions and then uses these landmarks / regions to calculate the similarity measure between faces. Lei *et al.* [15] proposed an automatic 3D particle filter method based on single sample input, which can effectively represent some 3D faces by counting the angle and distance information of multiple spatial triangular regions around key points. Yu *et al.* [16] proposed a new 3D direction vertex (3d2v) method, which effectively represents and matches 3D surfaces through fewer sparse distributed structured vertices. Soltanpour *et al.* [17] proposed a new 3D face recognition descriptor based on the local derivative pattern. They can extract more detailed unique information from 3D facial images by calculating high-order LNDP.

The hybrid-based method combines global and local features. Taghizadegan *et al.* [18] proposed a method of automatic face recognition using three-dimensional images, which uses the nose point for positioning, two-dimensional principal component analysis for feature acquisition, and Euclidean distance method for classification. Huang *et al.* [19] proposed a new geometric face representation and local feature hybrid matching scheme, which uses multi-scale extended local binary pattern to describe facial depth information and sift matching strategy to realize local and overall analysis.

B. POINT CLOUD DEEP LEARNING
Point cloud data refers to a set of vectors in a three-dimensional coordinate system. Point cloud data

generally scans the object to be measured through a 3D scanner and stores its data in the form of points. Each point contains three-dimensional coordinates, and some may contain color information or reflection intensity information. Due to the remarkable performance of point cloud data in 3D object classification, semantic segmentation, and target recognition, it is favored by researchers, and combined with the neural network, and it is widely used in feature extraction and registration. PointNet [7] began to use the deep learning network for point cloud processing. PointNet uses symmetric functions to aggregate point state features, which solves the problem of dealing with disordered point sets. However, PointNet cannot capture the local structure generated by metric space points, which limits its generalization ability to complex scenes. PointNet++ [8] introduces a multi-level feature extraction structure based on PointNet to extract local and global features on the point cloud in an iterative way. In addition, KD-Networks [20] uses kd-trees instead of a unified grid to build calculation diagrams and share learnable parameters. Compared with uniform voxel grid, kd-trees have improved the ability to index and construct 3D data. KD-Networks occupies less memory and has higher computational efficiency in training and testing. Unsupervised point fractal network pf-net estimates the missing point cloud hierarchically by using the multi-scale network based on feature points and proposes a new multi-resolution encoder. A new feature extractor combined with multi-layer perception (CMLP) is used to extract multi-layer features from local point clouds and their low-resolution feature points. Finally, the multi-stage completion loss and antagonistic loss are added to generate a more real missing area. Patchmatch-Net [21] introduces the Patchmatch idea into the end-to-end trainable deep learning based on MVS framework and embeds the model into the coarse to the fine framework to speed up the calculation speed. In addition, Patchmatch-Net uses a learnable adaptive module to enhance the traditional propagation and cost evaluation steps of Patchmatch, improving of image processing accuracy. Point-GNN [22] introduced graph convolution neural networks into point cloud processing. The GNN merge operation is designed to automatically reduce the variance of multiple translation points according to the GNN merge operation. Point cloud deep learning shows its robust in 3D object classification and segmentation performance. However, we know that a 3D point cloud neural network is relatively less used in face recognition.

III. METHOD

A. FACE PREPROCESSING

In order to enhance the local features of the face and reduce the impact of acquiring the original 3D scanning surface, we preprocess the acquired face before further operation [23], [24]. The operation consists of three parts: nose tip detection and face cutting, removing spikes and filling holes, and surface normals (as shown in Figure 1).
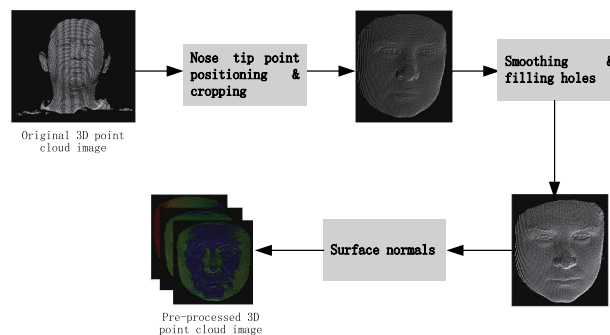

FIGURE 1. An illustration of 3D facial data preprocessing.

1) NOSETIP DETECTION AND FACE CROPPING

Given the original facial scan obtained from above the shoulder, we first detect the tip of the nose and remove unwanted points outside the three-dimensional area of the face. Firstly, a set of horizontal face scans are used to slice the three-dimensional face, and a set of horizontal contours of the three-dimensional face are obtained. For each horizontal contour, the points on the contour are evenly interpolated to fill the hole. Then, a set of probe points are located on each section and a circle is placed at each point to get two intersections with the horizontal section. The probe point forms a triangle with two intersections. The probe point with the maximum height h of its related triangle along the section is regarded as the candidate point of the nose tip. Repeat this process for all horizontal planes to obtain a set of nose tip candidates. Then, a random sample consensus (RANSAC) algorithm is used to screen these candidate points. The remaining candidates can be considered as a group of points on the bridge of the nose, and the one with the highest section height is considered the nose's tip. Once the nose tip is detected, the 3D face is cropped from the face scan by eliminating points more than 90 mm from the nose tip.

2) SMOOTH FACE AND FILL HOLES

Spikes mainly appear in three areas: eyes, the nose tip, and teeth. To remove these spikes, we apply a median filter to the vertices of 3D faces. The filter first sorts the coordinates in a neighborhood, finds the median, and finally replaces the original coordinates with the median. However, the process of removing spikes will produce unwanted holes in the three-dimensional surface. In addition, these holes may also be caused by other factors, including light absorption in dark areas, specular reflection of lower surfaces such as sclera, pupils and eyelashes, open mouth and eyes, and occlusion. For 3D faces, these holes can be filled with cubic interpolation.

3) SURFACE NORMALS

Compared with the simple point cloud image, the normal image of the face can display more local features, achieve the effect of enhancing the local features of the face, and facilitate better extraction and recognition of facial features. The facial surface normal components are estimated by adopting
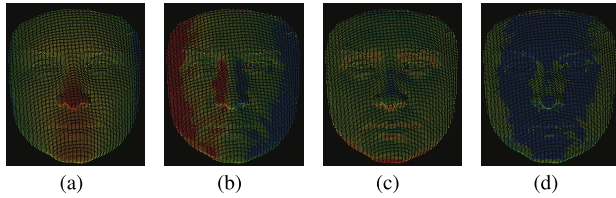
the optimization-based method [25]. The plane fits the local neighbors of a given point by minimizing the cost function (for example, the average distance to the neighbors). For a set of nth points, the data matrix is:

$$P = [p_1, p_2, \ldots, p_n]^T, \quad p_i \in R^3 \quad (1)$$

The 3D coordinates are defined as $p_i = [p_{ix}, p_{iy}, p_{iz}]^\top$. The normal vector $n_i = [n_{ix}, n_{iy}, n_{iz}]^T$ is estimated by the adjacent point Q_i around q_i.

It is calculated by solving the optimization problem as:

$$\min A(p_i, Q_i, n_i) \quad (2)$$

Here A is the cost function. We use a neighborhood of 5 × 5 matrix. Therefore, each 3D point has its normal component in the X, Y, and Z channels.

$$N(P) = Nx(j, k), N\, y(j, k), N\, z(j, k)$$
$$1 \le j \le m, \quad 1 \le k \le n \quad (3)$$

Figure 2 shows the normalized point cloud images from the Bosphorus database and their estimated three normal component images. Not surprisingly, the normal image contains more feature information than its corresponding point cloud image, and the point cloud image looks very subdued. In the normal images, the shape details around the eyes and mouth are nicely highlighted with other colors.

B. RoPS LOCAL FEATURE DESCRIPTOR

The local image feature is the local expression of image features, which can reflect the local characteristics of the image. After obtaining the key points detected from the 3D face, the local feature descriptor describes the local information around each key point in the form of a matrix, histogram, etc. In this paper, the rotation projection statistical descriptor [26] is used to encode the geometric information of the corresponding local surface. Then, we introduce a rollover protection structure into the rotating projection statistic for 3D face recognition.

Given a key point Q and its supporting radius r, the adjacent points around the key point Q whose distance is less than r are clipped from the three-dimensional surface to produce a point set Q = {q1, q2, . . . , qM}. Then, the RoPS descriptor is generated by following the steps below.

First, rotate the cloud of points to be measured Q by a set of angles around the x-axis $\{\theta k\}, k = 1, 2, 3, \ldots, K$. Generate rotation point cloud $Q'(\theta k)$, the point cloud after

rotation is $Q'(\theta k)$ It will be projected on XY, XZ and YZ planes respectively to obtain three projection point clouds $Q'_i(\theta k), i = 1, 2, 3$. 2D projection can describe 3D local surface in a concise and effective way, retain the geometric information under specific viewpoints, and significantly reduce the dimension, to achieve the purpose of recording the geometric information of different point sets Q to be measured.

Secondly, to extract the corresponding geometric information, we divide each projected 2D point set into N_{RoPS}^{div} * N_{RoPS}^{div} meshes on average. By calculating the number of point clouds in each grid, and N_{RoPS}^{div} * N_{RoPS}^{div} distribution matrix D is obtained. The distribution matrix D is further normalized to achieve invariance to the change of grid resolution. We use the central moment μ_{mn}, such as formula, and Shannon entropy e, such as formula, to further compress the information in distribution matrix D. The compressed distribution matrix D can improve the efficiency of calculation and storage. According to the literature [26], we use four central moments $\{\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}\}$.

Then, the central moment generated by rotation and projection is connected with Shannon entropy to form a sub-feature descriptor $f_x(\theta k)$ rotating around the x-axis. In order to encode more information about the local surface, the point set Q is rotated and projected around the y-axis and z-axis respectively in the same way to generate two sub feature descriptors $f_y(\theta k), f_z(\theta k)$ rotating around the y-axis and z-axis. All these sub-feature descriptors will be connected into a vector to form an overall RoPS feature descriptor, That is:

$$f = \{f_x(\theta_k), f_y(\theta_k), f_z(\theta_k)\}, \quad k = 1, 2, \ldots, T. \quad (4)$$

Finally, based on principal component analysis (PCA), the RoPS feature descriptor is further compressed. Select a set of training RoPS features and calculate its covariance matrix C. Then C is decomposed into eigenvalues to obtain its eigenvector. These eigenvectors are arranged according to the descending order of eigenvalues. The first N_{sf} eigenvector is used to form a matrix V_{sf}. N_{sf} is determined so that the fidelity of the training RoPS feature is maintained in the compressed feature θ Ratio of. Usually, θ Is a positive number close to 1. For a RoPS feature f_i, its compressed RoPS feature $\hat{f}_i$ is calculated as:

$$\hat{f}_i = V_{sf}^T f_i \quad (5)$$

C. NETWORK ARCHITECTURE

Inspired by PointNet++'s [8] introduction of a multi-level feature extraction structure to extract local features-global features on point clouds by iteration, we made improvements in the sampling layer of PointNet++ to make it more effective in capturing local surface information of 3D faces. The proposed network structure is shown in Figure 3. We use three ensemble abstraction (SA) modules containing sampling, grouping, and MLP layers to extract local-global features. The first two layers focus on local features of different fields of view by defining the center of mass of the local area and
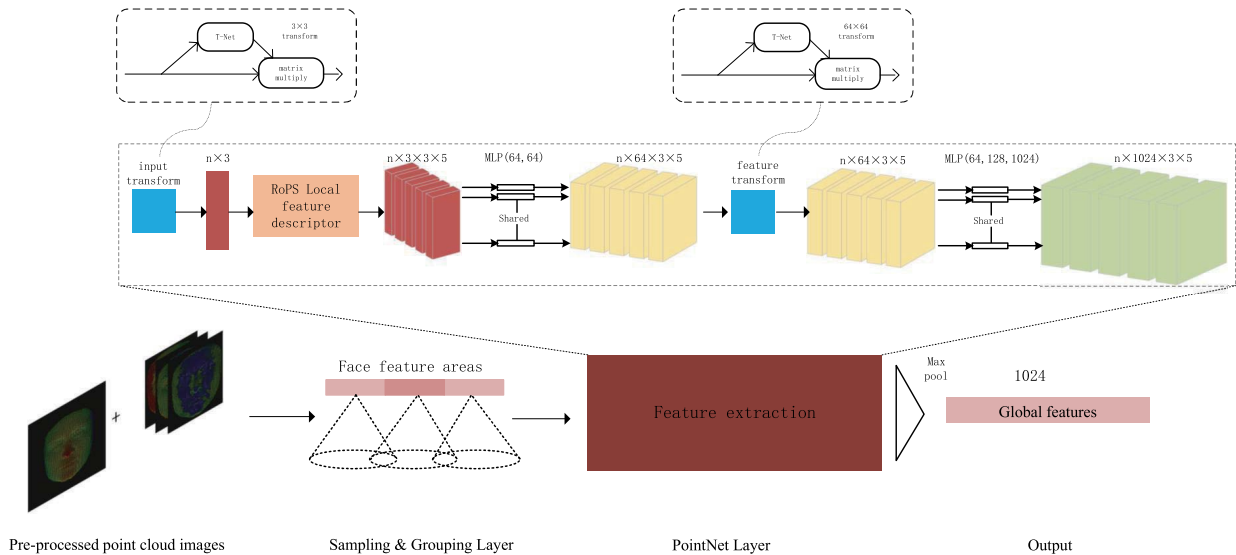
FIGURE 3. Improved PointNet++ modeling framework.The pre-processed point cloud image determines the face feature region by sampling layer and grouping layer, the face features are extracted by the PointNet layer, the local features are described by adding RoPS local feature descriptors in the PointNet layer to extract more accurate local features. Finally, the global features of the face are output by maximum pooling.

finding the "neighboring" points around the center of mass to build the local area set, while the MLP layer focuses on global features. Key point data are collected in the sampling layer by RoPS and directly grouped accordingly to better extract local features of the face.

In PointNet++ local feature selection, we use Rotational Projection Statistic (RoPS) for 3D keypoint detection and extract feature descriptors with sufficient recognition power from the local surface around these key points since RoPS crops out a circular region such that the local region has information around a single point and has a more geometric shape compared to the global information. To extract local features, we can directly use PointNet, and after sampling the key points several times and covering most of the global region, we can finish the process. After processing these feature key points by PointNet, we can get the final global features.

We use the global features proposed in FaceNet [27] as face embedding. This algorithm can be used to calculate the cosine similarity between faces. If the distance between two scans is greater than a given threshold, the two scans are considered to belong to the same constant equation and vice versa [28].

D. FEATURE MATCHING

Suppose $\mathcal{F}^i = \{f_n^i\}$ and $\mathcal{F}^j = \{f_m^j\}$ are RoPS feature sets extracted from 3D face $\boldsymbol{P}^i$ and $\boldsymbol{P}^j$, respectively. The nearest neighbor distance ratio (NNDR) method is used for feature matching. Specifically, each feature f_n^i in $\mathcal{F}^i$ matches all features in $\mathcal{F}^i$ to obtain its closest feature $f_{m'}^j$ the second closest feature $f_{m''}^j$, that is:

$$f_{m'}^j = \arg \min_{f_m^j \in \mathcal{F}^j} \left\| f_n^i - f_m^j \right\|_2,$$

$$f_{m''}^j = \arg \min_{f_m^j \in \mathcal{F}^j \setminus f_{m'}^j} \left\| f_n^i - f_m^j \right\|_2, \tag{6}$$

where $\mathcal{F}^j \setminus f_{m'}^j$ is the feature set F^j excluding feature $f_{m'}^j$. NNDR r_{dis} is calculated as:

$$r_{dis} = \frac{\left\| f_n^i - f_{m'}^j \right\|_2}{\left\| f_n^i - f_{m''}^j \right\|_2} \tag{7}$$

If the ratio r_{dis} is less than the threshold τ_f, $\left(f_n^i, f_{m'}^j \right)$ is considered as a potential feature matching. In order to achieve robust feature matching, $f_{m'}^j$ also matches all features in $\mathcal{F}^i$. If f_n^i is the closest feature in $\mathcal{F}^i$ to $f_{m'}^j$ and meets the NNDR criterion, then $\left(f_n^i, f_{m'}^j \right)$ is finally considered as feature matching. Threshold τ_f determines the number and accuracy of feature matching. The number of feature matching generated by a small threshold is limited, which is not enough to achieve accurate transform estimation. In contrast, a large threshold will lead to a large number of false positive matches, which reduces the performance of transform estimation. The face recognition performance under different thresholds is further analyzed in Section IV-B2. Figure 4 shows the same face matching at the same angle (Figure 4(a)), the same face matching at different angles (Figure 4(b)), different faces matching at the same angle (Figure 4(c)), and different faces matching at different angles (Figure 4(d)). For 3D point cloud faces from the same body at different angles, most features can be matched correctly. While most of the 3D point cloud faces from different individuals cannot match the corresponding features. We match all of the features in $\mathcal{F}^i$ against the features in $\mathcal{F}^j$, resulting in a set of matched keypoints $\mathcal{C}^{ij} = \left\{ c_1^{ij}, c_2^{ij}, \ldots, c_n^{ij} \right\}$, where $c_n^{ij} = \left\{ q_n^i, q_n^j \right\}$ is a pair of matched keypoints.
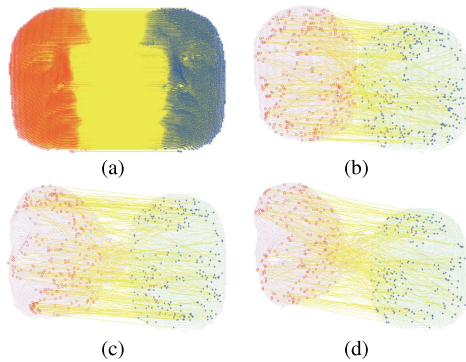
FIGURE 4. An illustration of the feature matching results: (a) indicates the matching result of the same facial feature from the same angle; (b) indicates the matching result of the same facial feature from different angles; (c) indicates the matching result of different facial features from the same angle; (d) indicates the matching result of different facial features from different angles.
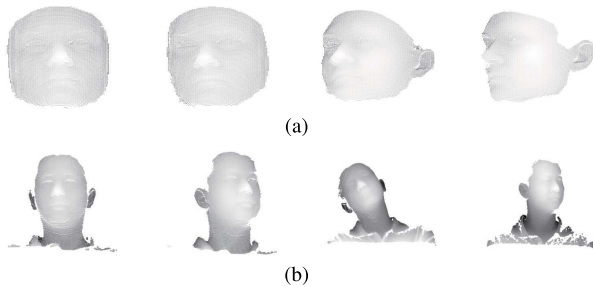

FIGURE 5. Selected face point clouds from the Bosphorus dataset and the CASIA-3D dataset: (a) Derived from the Bosphorus dataset. (b) Derived from the CASIA-3D dataset.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL SETUP

1) EXPERIMENTAL SIMULATION ENVIRONMENT SETTING

In this paper, we simulate 3D face recognition in normal and weak light environments.

In the classification training and fine-tuning under a normal environment, the input of our network is face scanning of 8192 point cloud data; In the classification and fine-tuning in a weak light environment, the number of point cloud data input for face scanning is reduced to 4096. Each point cloud data has six-dimensional features, including three European coordinates x, y, z, and their corresponding normal vector coordinates n_x, n_y, n_z.

We adopt the frequently used Rank-1 Identification Rate (R1IR) to measure the performance [29]. The R1IR is the percentage of the probe faces that are correctly recognized using the first rank.

2) DATASET DESCRIPTION

In this paper, we use the two publically available datasets (i.e., the Bosphorus dataset [30] and the CASIA-3D dataset [31]) (as shown in Figure 5) to test our proposed PR-Net in both normal and weak light environments.

The Bosphorus dataset includes 4666 3D facial scans from 105 individuals (61 men and 44 women) aged 25 to 35 [30]. There are more than 31 scans for each individual, and these

TABLE 1. The performance was represented by using different feature description radii r on the Bosphorus dataset.

r(mm)	5	10	15	30
Accuracy (%)	80.0	96.0	97.4	96.6

scans were acquired under different expressions, poses, and occlusions.

The CASIA-3D database was created by the pattern recognition and security technology research center of the Institute of automation, Chinese Academy of Sciences, using Minolta vivid910 3D digital scanner. The database was built from August to September 2004. There are 123 people in the database. Each person has 37 or 38 different three-dimensional data. Each data contains different expressions, gestures, lighting, and different combinations of the above conditions. There are 4626 face models [32].

B. ABLATION STUDY

In this section, we test the radius r and the threshold τ of feature matching used by our algorithm in different feature descriptions Performance in. We conduct comparative experiments on each parameter.

1) RADIUS USED IN THE FEATURE DESCRIPTION

The support radius determines the discriminative power and robustness of the expression. We tested our face verification algorithm and the support radius was set to 5mm, 10mm, 15mm, 20mm. The threshold τ was set to 0.8, and no feature compression was performed in the experiments. The experimental results are shown in Table 1. It can be seen that the facial verification performance improves significantly as the support radius increases from 5mm to 10mm. This is because the recognition capability of feature descriptors is insufficient when the support radius is small. The method achieves the best performance when the support radius is further increased from 10 mm to 15 mm. The face verification performance decreases when the support radius is further increased. This is due to the trade-off between discriminative power and robustness when the support radius is set to 15 mm. A larger support radius makes the extracted feature descriptors sensitive to expressions, which reduces the overall verification performance. In this paper, the feature descriptors are set to 15 mm using radius.

2) FEATURE MATCHING THRESHOLD

The threshold τ determines the number and accuracy of matched features. A smaller τ can improve feature matching accuracy, but the number of matched features is smaller. In this section, we will set the threshold τ to 0.6, 0.7, 0.8, and 0.9, set the radius r to 15 mm, and test the performance of face verification. The results are shown in Table 2. It is obvious that the best performance is obtained when τ is set to 0.8. When the threshold value increases, the recognition performance slightly decreases. This is because when the threshold is larger, many incorrect feature matches are encountered, and therefore these incorrect matches reduce the recognition

TABLE 2. The performance was expressed using different feature matching thresholds τ on the Bosphorus dataset.

τ	0.6	0.7	0.8	0.9
Accuracy (%)	97.60	98.74	98.22	97.20

TABLE 3. Comparison of the rank-1 identification rates achieved on the Bosphorus dataset set in a normal and weak light environment.

Method	Normal (%)	Weak light(%)
Ours(RP-Net)	98.00	**97.20**
Koushik Dutta et al. [33], 2020	**98.54**	96.50
Dutta et al. [34],2019	95.21	95.14
Soltanpour et al. [17], 2017	97.30	95.23
PointNet++. [8], 2017	80.20	82.36
Gilani et al. [35],2016	96.30	95.78

TABLE 4. Comparison of the rank-1 identification rates achieved on the CASIA-3D dataset set in a normal and weak light environment.

Method	Normal (%)	Weak light (%)
Ours(RP-Net)	97.90	**97.34**
Koushik Dutta et al. [33], 2020	97.97	95.43
Chandrakala et al. [36], 2018	**98.40**	96.04
Torkhani et al. [37], 2017	98.37	95.97
PointNet++. [8], 2017	82.40	85.36
Chouchane et al. [38], 2015	96.75	93.20

Chandrakala *et al.* [36] achieved the best recognition results with a rank-1 recognition rate of 98.4%. However, in a weak light environment, our algorithm has higher robustness.

V. CONCLUSION

In this paper, we propose a PointNet++ 3D face recognition algorithm integrating RoPS local feature descriptor (RP-Net). Firstly, the original 3D face is preprocessed by nose tip detection, face clipping, removing spikes, filling holes, and surface normals. The RoPS local feature descriptor is used to describe the facial features extracted from PointNet++ to enhance the extraction of face features in weak light or dark environment. The local face features extracted in each level will be spliced into global face features for output. So as to achieve the purpose of using local global facial features to classify. Finally, the nearest neighbor distance ratio algorithm is used for recognition. The experimental results on Bosphorus and CASIA-3D database show that our proposed algorithm achieves not only high face recognition rate in a normal environment but also has high face robustness in a weak light environment. This research, however, is subject to several limitations. the RP-Net extracts global-local features of 3D faces, which brings a large number of parameters and relatively slow recognition speed. In future work, we will try to improve the feature extraction structure of RP-Net in the direction of light weight to achieve a lightweight network model that balances speed and accuracy.

performance. In this paper, the threshold value τ is set to 0.8 for the subsequent experiments.

C. COMPARISON RESULTS OF NORMAL AND WEAK LIGHT ENVIRONMENT ON BOSPHORUS DATASET

The face recognition performance of the algorithm was tested using the Bosphorus dataset. First, key points are identified using the PointNet++ framework described in Section III-C, and key points are detected using the sampling layer iterative farthest-away technique (FPS). For each key point, its neighboring points (with a radius of 15 mm) are transformed into a local reference frame generated by RoPS features. These RoPS descriptors are then compressed and matched using the technique described in Section III-B. Finally, face recognition is performed using the similarity metric NNDR (Section III-D).

We have conducted classification training for 3D face recognition on the PointNet++ network without RoPS local descriptor. It can be clearly seen from table 3 that the recognition rate of RP-Net is much higher than that of PointNet++, which proves the effectiveness of our method with RoPS local descriptor. In order to compare our recognition results with the latest performance implemented on the Bosphorus dataset, we give the recognition rate of ranking 1 of the existing algorithms in Table 3. It can be seen that Dutta *et al.* [33] achieved the best recognition results in the normal environment, and the rank-1 recognition rate was 98.54%. The rank-1 recognition rate of RP-Net algorithm is 98.0%; In the weak light environment, RP-Net is 0.7% higher than the algorithm proposed by Koushik Dutta *et al.*

D. COMPARATIVE RESULTS ON THE CASIA-3D DATASET

The face recognition performance of the algorithm was further tested using the CASIA-3D dataset. The 3D face recognition results on the CASIA-3D dataset are shown in Table 4. It can be seen that under normal environment, our algorithm achieves 97.9% face recognition rate in the CASIA-3D dataset. In a weak light environment, our algorithm achieves 97.34% face recognition rate in the CASIA-3D dataset.

Again we compare our recognition results with the latest performance implemented on the CASIA-3D dataset, and we give the rank-1 recognition rate of the existing algorithms in Table 4. It can be seen that under a normal environment,

REFERENCES

[1] A. E. R. Shabayek and D. Aouada, "Dense and sparse 3D deformation signatures for 3D dynamic face recognition," *IEEE Access*, vol. 9, pp. 38687–38705, 2021.

[2] M. Abdelmaksoud, E. Nabil, I. Farag, and H. A. Hameed, "A novel neural network method for face recognition with a single sample per person," *IEEE Access*, vol. 8, pp. 102212–102221, 2020.

[3] Y. Jiang and Q. Ruan, "Multi-feature tensor neighborhood preserving embedding for 3D facial expression recognition," *IEEE Access*, vol. 9, pp. 106303–106316, 2021.

[4] B.-S. Kim and S. Seo, "Intelligent digital human agent service with deep learning based-face recognition," *IEEE Access*, vol. 10, pp. 72794–72805, 2022.

[5] M. Rouhsedaghat, Y. Wang, S. Hu, S. You, and C.-C.-J. Kuo, "Low-resolution face recognition in resource-constrained environments," *Pattern Recognit. Lett.*, vol. 149, pp. 193–199, Jun. 2021.

[6] X. Tu, J. Zhao, Q. Liu, W. Ai, G. Guo, Z. Li, W. Liu, and J. Feng, "Joint face image restoration and frontalization for recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1285–1298, Mar. 2022.

[7] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 652–660.

[8] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–10.

[9] J. Luo, F. Hu, and R. Wang, "3D face recognition based on deep learning," in *Proc. IEEE Int. Conf. Mechatronics Automat. (ICMA)*, Aug. 2019, pp. 1576–1581.

[10] K. Xu, X. Wang, Z. Hu, and Z. Zhang, "3D face recognition based on twin neural network combining deep map and texture," in *Proc. IEEE 19th Int. Conf. Commun. Technol. (ICCT)*, Oct. 2019, pp. 1665–1668.

[11] K. Dutta, D. Bhattacharjee, and M. Nasipuri, "SpPCANet: A simple deep learning-based feature extraction approach for 3D face recognition," *Multimedia Tools Appl.*, vol. 79, nos. 41–42, pp. 31329–31352, Nov. 2020.

[12] T. Russ, C. Boehnen, and T. Peters, "3D face recognition using 3D alignment for PCA," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2006, pp. 1391–1398.

[13] P. Liu, Y. Wang, D. Huang, Z. Zhang, and L. Chen, "Learning the spherical harmonic features for 3-D face recognition," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 914–925, Mar. 2013.

[14] H. Mohammadzade and D. Hatzinakos, "Iterative closest normal point for 3D face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 381–397, Feb. 2012.

[15] Y. Lei, Y. Guo, M. Hayat, M. Bennamoun, and X. Zhou, "A two-phase weighted collaborative representation for 3D partial face recognition with single sample," *Pattern Recognit.*, vol. 52, pp. 218–237, Apr. 2016.

[16] X. Yu, Y. Gao, and J. Zhou, "3D face recognition under partial occlusions using radial strings," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3016–3020.

[17] S. Soltanpour and Q. J. Wu, "High-order local normal derivative pattern (LNDP) for 3D face recognition," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2811–2815.

[18] Y. Taghizadegan, H. Ghassemian, and M. Naser-Moghaddasi, "3D face recognition method using 2DPCA-Euclidean distance classification," *ACEEE Int. J. Control Syst. Instrum.*, vol. 3, no. 1, pp. 1–5, 2012.

[19] D. Huang, M. Ardabilian, Y. Wang, and L. Chen, "3-D face recognition using eLBP-based facial description and local feature hybrid matching," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 5, pp. 1551–1565, Oct. 2012.

[20] R. Klokov and V. Lempitsky, "Escape from cells: Deep kd-networks for the recognition of 3D point cloud models," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 863–872.

[21] F. Wang, S. Galliani, C. Vogel, P. Speciale, and M. Pollefeys, "Patchmatch-Net: Learned multi-view patchmatch stereo," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14194–14203.

[22] W. Shi and R. Rajkumar, "Point-GNN: Graph neural network for 3D object detection in a point cloud," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1711–1719.

[23] D. Gera and S. Balasubramanian, "Landmark guidance independent spatio-channel attention and complementary context information based facial expression recognition," *Pattern Recognit. Lett.*, vol. 145, pp. 58–66, May 2021.

[24] J. Rivero-Hernández, A. Morales-González, L. G. Denis, and H. Méndez-Vázquez, "Ordered weighted aggregation networks for video face recognition," *Pattern Recognit. Lett.*, vol. 146, pp. 237–243, Jun. 2021.

[25] T. Song, F. Xi, S. Guo, Z. Ming, and Y. Lin, "A comparison study of algorithms for surface normal determination based on point cloud data," *Precis. Eng.*, vol. 39, pp. 47–55, Jan. 2015.

[26] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, and J. Wan, "Rotational projection statistics for 3D local surface description and object recognition," *Int. J. Comput. Vis.*, vol. 105, no. 1, pp. 63–86, 2013.

[27] A. Tunç, M. Yildirim, C. S. Taşdemir, and A. A. Altun, "Development of face recognition system by using deep learning and face-net algorithm in the operations processes," in *Proc. Int. Conf. Artif. Intell. Appl. Math. Eng.* Springer, 2020, pp. 93–105.

[28] J. Chen, Z. Guo, and J. Hu, "Ring-regularized cosine similarity learning for fine-grained face verification," *Pattern Recognit. Lett.*, vol. 148, pp. 68–74, Aug. 2021.

[29] Y. Guo, Y. Lei, L. Liu, Y. Wang, M. Bennamoun, and F. Sohel, "EI3D: Expression-invariant 3D face recognition based on feature and shape matching," *Pattern Recognit. Lett.*, vol. 83, pp. 403–412, Nov. 2016.

[30] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis," in *Proc. Eur. Workshop Biometrics Identity Manage.* Berlin, Germany: Springer-Verlag, 2008, pp. 47–56.

[31] (2004). C. for Biometrics and S. Research. *Note on CASIA-3D FaceV1*. Accessed: Nov. 10, 2021. [Online]. Available: http://biometrics.idealtest.org

[32] C. Xu, T. Tan, S. Li, Y. Wang, and C. Zhong, "Learning effective intrinsic features to boost 3D-based face recognition," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer-Verlag, 2006, pp. 416–427.

[33] S. Ram, W. Tang, A. J. Bell, C. Spencer, A. Buschhaus, C. R. Hatt, M. P. di Magliano, J. J. Rodriguez, S. Galban, and C. J. Galban, "Lung cancer lesion detection in histopathology images using graph-based sparse PCA network," 2021, *arXiv:2110.14728*.

[34] K. Dutta, D. Bhattacharjee, M. Nasipuri, and A. Poddar, "3D face recognition based on volumetric representation of range image," in *Advanced Computing and Systems for Security*. Kolkata, India: Springer-Verlag, 2019, pp. 175–189.

[35] S. Z. Gilani and A. Mian, "Towards large-scale 3D face recognition," in *Proc. Int. Conf. Digit. Image Computing: Techn. Appl. (DICTA)*, Nov. 2016, pp. 1–8.

[36] M. Chandrakala and S. Ravi, "Effective 3D face recognition technique based on Gabor and LTP features," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 2S, pp. 284–290, 2018.

[37] C. Li, Y. Tan, D. Wang, and P. Ma, "Research on 3D face recognition method in cloud environment based on semi supervised clustering algorithm," *Multimedia Tools Appl.*, vol. 76, no. 16, pp. 17055–17073, 2017.

[38] A. Chouchane, M. Belahcene, and S. Bourennane, "3D and 2D face recognition using integral projection curves based depth and intensity images," *Int. J. Intell. Syst. Technol. Appl.*, vol. 14, no. 1, pp. 50–69, 2015.

YIBO CAO was born in Hengyang, Hunan, China, in 1971. He received the B.S. and M.S. degrees from the Department of Electrical and Mechanical Engineering, Xi'an University of Electronic Science and Technology, in 1994 and 1997, respectively, the Ph.D. degree from the School of Mechanical Engineering, South China University of Technology, Guangzhou, China, in 2007, and the Ph.D. degree from the Department of Precision Instruments, Tsinghua University, Beijing, China, in 2009.

He joined the Software College, South China Normal University, Guangzhou, in 2009, and is currently an Associate Professor. His research interests include AI artificial intelligence technology research, optical electromechanical integration technology, the Internet of Things technology, pattern recognition, and simultaneous localization and mapping (SLAM).

SHUN LIU was born in Yiyang, Hunan, China, in 1999. He is currently pursuing the Graduate degree with the School of Software, South China Normal University. His current research interests include computer vision, pattern recognition, and 3D face recognition.

PENGFEI ZHAO was born in Zhoukou, Henan, China, in 1997. He is currently pursuing the Graduate degree with the School of Software, South China Normal University. His current research interests include computer vision, vision simultaneous localization, mapping, and indoor positioning.

HAIWEN ZHU received the B.E. degree from the College of Earth Science and Technology, Southwest Petroleum University, Chengdu, China, in 2020. He is currently pursuing the master's degree with the School of Software, South China Normal University. His research interests include autonomous robot navigation and environmental mapping.

• • •