

Local Feature Tensor Based Deep Learning for 3D Face Recognition

Shisong Lin^{1,2}, Feng Liu^{1,2}, Yahui Liu^{1,2} and Linlin Shen^{1,2}

¹ Computer Vision Institute, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, P.R. China

²Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University, Shenzhen, P.R. China

Abstract—A local feature tensor similarity based deep learning approach is proposed in this paper for 3D face recognition. Once a set of salient points on the 3D mesh are detected, three scale and rotation invariant features are extracted to represent local surface around each salient point. The local features of all the salient points are concatenated to produce a 3rd order feature tensor to represent a 3D face. Similarity of two 3D faces can thus be measured by a similarity tensor calculated using the two feature tensors. To address the unavailability of large 3D face samples, a feature tensor based data augmentation approach is proposed to augment the number of feature tensors. Experimental results show that the ResNet model trained using the augmented feature tensors achieves the best performance among state of the art competitors, i.e. 99.71% and 96.2% accuracy are achieved for Bosphorus and BU3DFE database, respectively.

I. INTRODUCTION

Face recognition is one of the most researched topics in Biometrics. With the huge success of deep learning, especially Convolutional Neural Network (CNN), 2D face recognition techniques make repaid progresses in various domains, such as security, surveillance, and entertainment. In the past decade, 3D face recognition receives widely attention due to its advantages over 2D face recognition in terms of illumination, pose and scale robustness.

However, 3D face recognition still faces some critical challenges. The first challenge is the shape deformations caused by changes in facial expressions. Since the geometric shape of 3D face is the primary information to distinguish different subjects, how to extract expression-invariant features becomes critical to 3D face recognition. The second challenge is the unavailability of large 3D face database, which makes the 3D face recognition unable to benefit from the recent progress in deep learning. To address these two challenges, various approaches have been proposed for 3D face recognition. These approaches can be divided into two broad categories: the conventional methods, such as holistic [1], [2], feature-based [3], [4], and hybrid matching methods, and the deep learning based methods [5]. For instance, Kim *et al.* [6] proposed a 3D face recognition algorithm based on a deep convolutional neural network with a 3D augmentation technique. To cope with the limited amount of available 3D data, they fine-tuned the existing network trained for 2D face recognition to apply in 3D face recognition. In order to improve the performance of 3D face recognition under different expressions, they generated a lot

of synthesized 3D faces from a single raw 3D scan by changing expression to augment their 3D face database. In additon, a number of surveys have discussed conventional 3D face recognition approaches by summarizing the advantages and disadvantages, limitations and solutions, characteristics and challenges, as well as the 3D face databases accordingly [7], [8]. From the discussion of these surveys, we can see that there is still room to improve the performance of the existing methods.

In order to fully integrate geometric features in 3D face into the powerful framework of CNN, we propose a local feature tensor based deep learning method to improve the performance of 3D face recognition.

The novelties of this paper are listed as follows:

- 1) A 3rd order feature tensor based on meshSIFT feature is designed to represent the local features around salient points on the 3D face;
- 2) A similarity tensor is proposed to measure the similarity of two feature tensors, i.e. two 3D faces;
- 3) A feature tensor based data augmentation approach is used to generate a large number of intra-personal and extra-personal similarity tensors;
- 4) A ResNet based deep network is successfully trained to predict whether two 3D faces are from the same person, or not. The network achieves very competitive performance on both Bosphrous and BU3DFE databases.

The rest of the paper is organized as follows. Section II introduces the construction and matching of the local feature tensor of 3D mesh. Section III describes the deep learning details for similarity tensor classification. Section IV presents the experimental settings and discusses the experimental results. Finally, section V concludes the paper.

II. LOCAL FEATURE TENSOR EXTRACTION AND MATCHING

In this section, the salient points are firstly detected based on the meshSIFT algorithm. Subsequently, three local features are extracted from each salient point to describe the shape index, slant angle, and the relative position of the local patch around the salient point in the 3D mesh. Afterwards, the feature tensor, which integrates features of all the salient points, is defined to represent the surface mesh of the 3D face. Finally, the tensor matching is implemented and a corresponding RGB image is generated to represent the

similarity tensor for further classification. Fig. 1 illustrates the framework of the system.

A. Salient Point Detection

Inspired by SIFT [9], meshSIFT algorithm [10] has been proved to be a powerful technique to extract invariant and distinctive features from 3D images. It can be considered as an extension of SIFT on 3D meshes. Similarly, the salient point detection is implemented on scale space.

To detect the salient points on the mesh, a series of approximating Gaussian filters \widehat{G}_{σ_s} are firstly implemented on an input surface mesh $M = (\gamma, \varepsilon)$, where γ are vertices connected by edges ε , for scale space construction. The scale space containing smoothed versions of the input mesh is expressed as:

$$M_s = \begin{cases} M & \text{if } s = 0 \\ \widehat{G}_{\sigma_s} \otimes M & \text{otherwise} \end{cases} \quad (1)$$

where M is the original input mesh, \widehat{G}_{σ_s} is the approximated Gaussian filter with standard deviation σ_s , and the symbol \otimes represents the convolution operation. The standard deviation $\sigma_s = 2^{\frac{s}{k}} \cdot \sigma_0$ varies exponentially to describe different scales, and can be discretely approximated as:

$$\tilde{\sigma}_s = \bar{e} \cdot \sqrt{\frac{2 \cdot \sigma_0}{3}} \cdot 2^{\frac{s}{k}} \quad (2)$$

where \bar{e} is the average length of edges ε , k is the number of total scales, and $s = 0, \dots, n_{scales} + 2$. Subsequently, the mean curvature at each vertex v_i with each scale $\tilde{\sigma}_s$ is computed as:

$$H_i^s = \frac{k_{i,max}^s + k_{i,min}^s}{2} \quad (3)$$

where $k_{i,max}^s$ and $k_{i,min}^s$ stand for the maximal and minimal curvatures at vertex v_i with scale $\tilde{\sigma}_s$. The difference between consecutive scales is computed as:

$$dH_i^s = H_i^{s+1} - H_i^s \quad (4)$$

To detect the salient points, each dH_i^s is compared with its 20 neighbors on the same scale as well as on the upper and lower scales. Finally, the vertexes with the largest or the smallest dH_i^s among its neighbors are selected as salient points. Meanwhile, the scale $\tilde{\sigma}_s$ at which the extremum is obtained is associated with the salient point.

B. Feature Tensor Extraction

Using the salient points, the local surface around the detected vertex can be expressed by an orientation invariant descriptor to extract the shape index, the slant angle, and the relative position of the surface patch.

For each salient point, the normal to the local surface, as well as the canonical orientation, are calculated to construct a local reference frame in which the vertices of the neighborhood can be described. Specifically, a spherical region with radius $9\tilde{\sigma}_s$ around the salient point are located. All the vertexes in this spherical region are selected as neighborhood points, at which the normal vectors and the geodesic

distances to the salient point are calculated. After that, all the normal vectors are projected onto the tangent plane to the local surface at the salient point. Then a histogram with 360 bins (stand for 360 degrees) are constructed with entries weighted by their corresponding geodesic distances. The highest peak in the histogram and every peak above 80% of this highest peak value are selected as canonical orientations. If there are more than one peaks selected, the salient point is regarded as multiple salient points, each assigned one of the canonical orientations. Therefore, the scale and the orientation of each salient point is extracted. The local feature descriptors are then developed to summarize the neighborhood around each salient point. As shown in Fig. 2, 9 sub-regions, with radius of $3.75\tilde{\sigma}_s$, are selected around the salient point. And the geodesic distances between the region centers to the salient point are $4.5\tilde{\sigma}_s$ (regions 2, 4, 6, 8) and $4.5\sqrt{2}\tilde{\sigma}_s$ (regions 3, 5, 7, 9) respectively. For each of the 9 regions, two histograms with eight bins each are calculated. The first histogram p_s contains the shape index expressed as:

$$S_i = \frac{2}{\pi} \cdot \arctan\left(\frac{k_{i,max} + k_{i,min}}{k_{i,max} - k_{i,min}}\right) \quad (5)$$

where $k_{i,max}$ and $k_{i,min}$ stand for the maximal and minimal curvatures at vertex v_i . The second histogram p_θ contains the slant angles, which are defined as the angles between the projected normal and the canonical orientation. Both histograms p_s and p_θ are then Gaussian weighted with the geodesic distances to the salient point. The two histograms of the 9 regions are concatenated respectively into two feature vectors i.e. $f_s = [p_{s,1}, p_{s,2}, \dots, p_{s,9}]^T$ and $f_\theta = [p_{\theta,1}, p_{\theta,2}, \dots, p_{\theta,9}]^T$ to describe the shape index and the slant angle for each salient point, respectively.

For the relative position description, the 68 face landmarks [11] on the 2D face image are utilized as reference points. Given a salient point V , its 2D coordinates in corresponding RGB image are $[x_0, y_0]^T$, for each landmark $L_j = [x_j, y_j]^T$, where $j = 1, \dots, 68$, the L2-norm $d_j = \|V - L_j\|$ is calculated. All the L2-norm values are then concatenated and normalized between 0 to 1 to form a feature vector $f_d = [d_1, d_2, \dots, d_{68}]^T$. The maximum, minimum, mean, and standard deviation of $[d_1, d_2, \dots, d_{68}]^T$ are also calculated and connected to the end of f_d . Finally, the $f_d = [d_1, d_2, \dots, d_{68}, max, min, mean, std]^T$ is extracted to describe the relative position of the i_{th} salient point on the 3D face mesh.

Consequently, three feature channels of each salient point are extracted as $f_i = (f_s, f_\theta, f_d)$, where $i = 1, \dots, N$. Finally, the feature tensor is formed as $F = [f_1, \dots, f_N]$ to represent the whole 3D surface.

C. Feature Tensor Matching

Using the feature tensor extracted from 3D mesh, two 3D faces G and T can be represented by their feature tensors $F_G = [F_{G,1}, \dots, F_{G,M}]$ and $F_T = [F_{T,1}, \dots, F_{T,N}]$, where M and N are the numbers of salient points in G and T respectively. In order to measure the similarity between

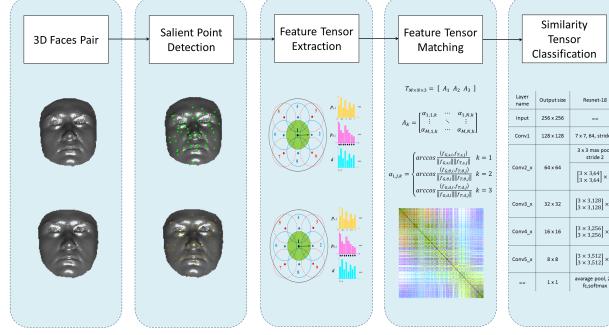


Fig. 1. System framework

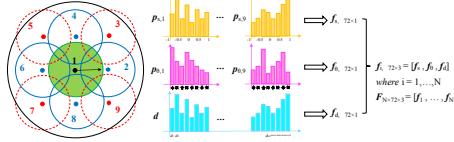


Fig. 2. Feature tensor extraction illustration

two feature tensors, a 3^{rd} order tensor Γ , namely similarity tensor, is defined to measure the similarity between the two faces:

$$\Gamma_{M \times N \times 3} = [A_1 \quad A_2 \quad A_3] \quad (6)$$

where

$$A_k = \begin{bmatrix} \alpha_{1,1,k} & \dots & \alpha_{1,N,k} \\ \vdots & \ddots & \vdots \\ \alpha_{M,1,k} & \dots & \alpha_{M,N,k} \end{bmatrix} \quad (7)$$

where $k = 1, 2, 3$, and

$$\alpha_{i,j,k} = \begin{cases} \arccos \frac{\langle f_{G,s,i}, f_{T,s,j} \rangle}{\|f_{G,s,i}\| \|f_{T,s,j}\|} & k = 1 \\ \arccos \frac{\langle f_{G,\theta,i}, f_{T,\theta,j} \rangle}{\|f_{G,\theta,i}\| \|f_{T,\theta,j}\|} & k = 2 \\ \arccos \frac{\langle f_{G,d,i}, f_{T,d,j} \rangle}{\|f_{G,d,i}\| \|f_{T,d,j}\|} & k = 3 \end{cases} \quad (8)$$

where $i = 1, \dots, M$, and $j = 1, \dots, N$.

By mapping the $\alpha_{i,j,k}$ values from range $[0, \pi]$ to $[0, 255]$, the similarity tensor can be visualized as a RGB image. Fig. 3 illustrates the genuine similarity tensor calculated from two faces of the same person, as well as the impostor similarity tensor of two faces from different persons, respectively. The problem of 3D face verification can now be converted to a binary classification problem, i.e. the similarity tensor image is from the same person, or different persons.

III. DEEP LEARNING BASED SIMILARITY TENSOR CLASSIFICATION

Inspired by the great success of deep learning, especially Convolutional Neural Network (CNN), for image classification, a deep CNN method is utilized for the similarity tensor classification. We first generate a large number of similarity tensor images using 3D faces of the same and different persons, and then train a deep CNN to predict whether a similarity tensor is intra-person or extra-person.

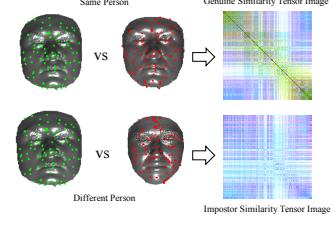


Fig. 3. Feature tensor matching visualization

A. Network

We adopt the *ResNet - 18* [12] for classification. As shown in Fig. 4, the network contains seventeen convolutional layers with eight residual shortcuts and one fully connected layer (FC). After each convolutional layer, the BN and ReLU activation layers are followed. The last FC layer is adopted for classification with the binary cross-entropy loss.

B. Feature Tensor Based Data Augmentation

Typically, the deep CNN needs millions of samples for training. However, building such a large 3D face dataset is beyond the capabilities of most research groups. A possible solution is generating 3D face samples from 2D face images based on 3D face reconstruction or 3D face modeling algorithms. Instead of generating 3D face samples, we generate the feature tensors directly based on the Voronoi diagram subdivision.

Given a set of points in a plane, a Voronoi diagram partitions the space such that the boundary lines are equidistant from neighboring points. Every boundary line passes through the center of two points. Fig. 5 shows an example of a Voronoi diagram generated from the 68 landmark points on a human face. The landmarks points are detected using the pre-trained facial landmarks detector proposed by Kazemi [11]. Since the 3D face dataset used in this paper consists of 3D points-clouds with their co-registered 2D RGB images, all 3D points can be mapped into 2D coordinates in the corresponding RGB images. Therefore, all the salient points can be divided into different subdivisions based on the Voronoi diagram. Specifically, given a 3D face points cloud with its co-registered 2D face image, we detect the salient points and facial landmarks simultaneously. Afterwards, a Voronoi

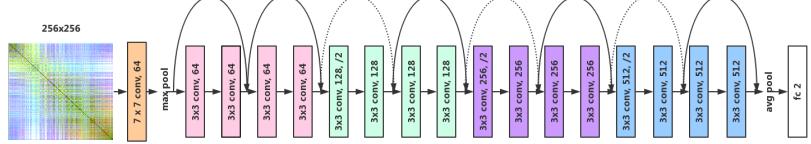


Fig. 4. Architecture of the CNN for similarity tensor classification

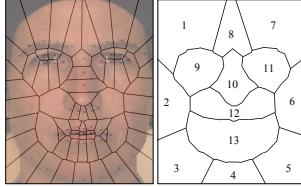


Fig. 5. Voronoi segmentation (13 sub-areas) based on face landmarks

diagram is generated according to the 68 landmark points. To avoid over-division, the 68 subdivisions are clustered into 13 areas, as shown in Fig. 5. Consequently, each feature set $F = [f_1, \dots, f_N]$ is divided into 13 sub feature sets defined as $SubF_i, i = 1, \dots, 13, F = \bigcup_{i=1}^{13} SubF_i$.

For a group of samples from the same person, it is safe to assume that the same sub feature sets of different samples describe the same area of the person's face. Given a set of K 3D face samples, a feature tensor $F^k = \bigcup_{i=1}^{13} SubF_i^k$ can be generated for the k^{th} sample using the procedure presented in section II.B. A new feature set $F^n = SubF_1^{n_1} \cup \dots \cup SubF_i^{n_i} \dots \cup SubF_{13}^{n_{13}}$ ($1 \leq n_i \leq K$) could be synthesized by randomly choosing the i^{th} sub feature set from the K samples. A million similarity tensors, including 0.5 million genuine similarity tensors and 0.5 million impostor similarity tensors were generated using the synthesized tensors. In the training stage, 90% of the tensor images are used for training, and 10% of the tensor images are used for validation. An accuracy rate beyond 99% is achieved after 20 epochs.

C. Implementation Details

We resize the tensor images to $256 \times 256 \times 3$ as the input of the proposed network. We use Adam with mini-batch size of 200. The learning rate starts from 0.001 and is updated with cosine annealing schedule [13]. The network was trained for 20 epochs. We use a weight decay of 0.0001 and do not use dropout.

IV. EXPERIMENTAL RESULTS

A. Bosphorus Database

The Bosphorus 3D face database [14] contains 4666 faces captured from 105 subjects, with 7 types of facial expressions and 28 types of action units. Meanwhile, the 2D and 3D facial landmarks are recorded as coordinates in 2D RGB image and the co-registered 3D points cloud. The typical landmarks have 22 labels and some special cases have 15 to 24 labels.

Since most of the existing papers focus on expression-robust face recognition on the Bosphorus database, we use similar strategy to evaluate the performance of our method. The first neutral sample of each 105 subjects is used as gallery, and the remaining samples with different expressions and different action units are used as probes (2797 samples).

B. BU3DFE Database

The BU3DFE 3D face database [15] consists of 2500 scans of 100 individuals displaying expression of neutral, happiness, disgust, fear, angry, surprise and sadness. All non-neutral expressions were acquired at four levels of intensity. In addition, the 3D facial landmarks with 83 labels were recorded as coordinates in the 3D points cloud. The first neutral sample of each 100 subjects is used as gallery, and the remaining samples with different expressions are used as probes (2400 samples).

C. Results

1) *Bosphorus database*: For each of the 2797 probe samples, a feature tensor was extracted and matched with the 105 gallery samples to generate a similarity tensor, which is input to the trained ResNet to decide whether the two samples are from the same person, or not with a threshold 0.5. However, if there are multiple tensors probabilities higher than 0.5, the final prediction is the one has the highest probability. Table I lists the rank-one accuracies of different approaches for Bosphorus database, where *w.* and *w.o.* denote whether the feature tensor based augmentation is applied, or not. As shown in the Table I, augmentation improves the accuracy of our approach from 98.68% to 99.71%, which beats all of the competing approaches.

2) *BU3DFE database*: Table II lists the rank-one accuracies of different approaches for BU3DFE database. Again, the proposed augmentation approach significantly increases the accuracy of the network from 77.32% to 96.2%, which is the same with the best approach in literature [24].

V. CONCLUSIONS AND FUTURE WORKS

In this paper, a novel local feature tensor based deep learning method is proposed for expression-robust 3D face recognition. A data augmentation method is proposed based on the facial landmarks and Voronoi diagram subdivision. By integrating the geometric features of 3D face into the powerful frame of CNN, a 99.71% rank-one recognition rate is achieved on Bosphorus database, and a 96.2% rank-one recognition rates is achieved on BU3DFE database.

TABLE I

RANK-ONE RECOGNITION RATES (%) FOR THE BOSPHROUS DATABASE.

Approaches	Rank-one accuracies(%)
Alyuz et al. [17].	98.20
Ocegueda et al. [18]	98.20
Smeets et al. [10]	97.70
Zhang et al. [19]	98.65
Li et al. [20]	95.40
Li et al. [16]	98.82
Emambakhsh et al. [21]	95.35
Li et al. [5]	97.89
Ours (w.o.)	98.68
Ours (w.)	99.71

TABLE II

RANK-ONE RECOGNITION RATES (%) FOR THE BU3DFE DATABASE.

Approaches	Rank-one accuracies(%)
Kim et al. [6]	95.0
Li et al. [16], [20]	92.2
Lei et al. [22]	94.0
Mian et al. [23]	95.9
Gilani et al. [24]	96.2
Ours (w.o.)	77.32
Ours (w.)	96.2

There will be several improvements in the future. The automatic detection of 3D facial landmarks will be studied. More experiments will be implemented on the other 3D face databases, such as FRGC v 2.0 [25] and several new datasets [26], [27]. Moreover, the features described by other 3D SIFT-like algorithms [28], as well as the features from specific regions of the face instead of the hand-crafted features, will be extracted for additional experiments and comparison.

VI. ACKNOWLEDGMENTS

The work is supported by National Natural Science Foundation of China (Grant No. 61672357 and U1713214), and the Science and Technology Project of Guangdong Province (Grant No. 2018A050501014).

REFERENCES

- [1] T. Russ, C. Boehnen and T. Peters, "3D Face Recognition Using 3D Alignment for PCA", in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, New York, NY, USA, 2006, pp. 1391-1398.
- [2] X. Lu and A. K. Jain, "Deformation Modeling for Robust 3D Face Matching", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 8, pp. 1346-1357, Aug. 2008.
- [3] A. S. Mian, M. Bennamoun and R. Owens, "Keypoint Detection and Local Feature Matching for Textured 3D Face Recognition", *International Journal of Computer Vision*, vol. 79, no. 1, pp. 1-12, 2008.
- [4] D. Zeng, S. Long, J. Li and Q. Zhao, "A Novel Approach to Mugshot Based Arbitrary View Face Recognition", *Journal of the Optical Society of Korea*, vol. 20, no.2, pp. 239-244, 2016.
- [5] H. Li, J. Sun and L. Chen, "Location-Sensitive Sparse Representation of Deep Normal Patterns for Expression-Robust 3D Face Recognition", in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, Denver, CO, 2017, pp. 234-242.
- [6] D. Kim, M. Hernandez, J. Choi and G. Medioni, "Deep 3D Face Identification", *arXiv preprint arXiv:1703.10714*, 2017.
- [7] Y. Lei, M. Bennamoun, M. Hayat and Y. Guo, "An Efficient 3D Face Recognition Approach Using Local Geometrical Signatures", *Pattern Recognition*, vol. 47, no. 2, pp. 509-524, 2014.
- [8] S. Soltanpour, B. Boufama and Q. M. J. Wu, "A Survey of Local Feature Methods for 3D Face Recognition", *Pattern Recognition*, vol. 72, pp. 391-406, 2017.
- [9] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2014.
- [10] D. Smeets, J. Keustermans, D. Vandermeulen and P. Suetens, "MeshSIFT: Local Surface Features for 3D Face Recognition Under Expression Variations and Partial Data", *Computer Vision and Image Understanding*, vol. 117, no. 2, pp. 158-169, 2013.
- [11] V. Kazemi and J. Sullivan, "One Millisecond Face Alignment with An Ensemble of Regression Trees", in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, 2014, pp. 1867-1874.
- [12] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition", in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016, pp. 770-778.
- [13] I. Loshchilov and F. Hutter, "SGDR: Stochastic Gradient Descent with Warm Restarts", *arXiv preprint arXiv:1608.03983*, 2016.
- [14] A. Savran and L. Akarun, "Bosphorus Database for 3D Face Analysis", in *Proc. First Cost 2101 Workshop on Biometrics and Identity Management*, Springer, Berlin, Heidelberg, 2008, pp. 47-56.
- [15] L. Yin, X. Wei, Y. Sun and J. Wang, "A 3D Facial Expression Database for Facial Behavior Research", in *International Conference on Automatic Face and Gesture Recognition*, Southampton, 2006, pp. 211-216.
- [16] H. Li, D. Huang, J. M. Morvan and L. Chen, "Towards 3D Face Recognition in the Real: A Registration-Free Approach Using Fine-Grained Matching of 3D Keypoint Descriptors", *International Journal of Computer Vision*, vol. 113, no. 2, pp. 128-142, 2015.
- [17] N. Alyuz, B. Gokberk and L. Akarun, "Regional Registration for Expression Resistant 3-D Face Recognition", *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 425-440, 2010.
- [18] O. Ocegueda, G. Passalis, T. Theoharis, S. K. Shah and I. A. Kakadiaris, "UR3D-C: Linear Dimensionality Reduction for Efficient 3D Face Recognition", *International Joint Conference on Biometrics*, 2011, pp. 1-6.
- [19] L. Zhang, Z. Ding, H. Li, Y. Shen and J. Lu, "3D Face Recognition Based on Multiple Keypoint Descriptors and Sparse Representation", *Plos One*, vol. 9, no. 6, pp. e100120, 2014.
- [20] H. Li, D. Huang, J. M. Morvan, L. Chen and Y. Wang, "Expression-Robust 3D Face Recognition Via Weighted Sparse Representation of Multi-Scale and Multi-Component Local Normal Patterns", *Neurocomputing*, vol. 133, no. 14, pp. 179-193, 2014.
- [21] M. Emambakhsh and A. Evans, "Nasal Patches and Curves for An Expression-Robust 3D Face Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 5, pp. 995-1007, 2016.
- [22] 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 Recognition*, vol. 52, pp. 218-237, 2016.
- [23] A. Mian, M. Bennamoun and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition", *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 11, pp. 1927-1943, 2007.
- [24] S. Z. Gilani, A. Mian, F. Shafait and I. Reid, "Dense 3D Face Correspondence", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 7, pp. 1584-1598, 1 July 2018.
- [25] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek, "Overview of the Face Recognition Grand Challenge", in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, 2005, pp. 947-954, vol. 1.
- [26] V. Vijayan, K. W. Bowyer, P. J. Flynn and D. Huang, "Twins 3D Face Recognition Challenge", in *International Joint Conference on Biometrics*, Washington, DC, 2011, pp. 1-7.
- [27] J. Zhang, D. Huang, Y. Wang and J. Sun, "Lock3DFace: A Large-Scale Database of Low-Cost Kinect 3D Faces", in *International Conference on Biometrics*, Halmstad, 2016, pp. 1-8.
- [28] P. Scovanner, S. Ali and M. Shah, "A 3-Dimensional Sift Descriptor and Its Application to Action Recognition", in *Proceedings of the 15th ACM international conference on Multimedia*. ACM, 2007, pp. 357-360.