HIGH-ORDER LOCAL NORMAL DERIVATIVE PATTERN (LNDP) FOR 3D FACE RECOGNITION

Sima Soltanpour, Q.M. Jonathan Wu

University of Windsor
Department of Electrical and Computer Engineering
Windsor, Canada

ABSTRACT

This paper proposes a novel descriptor based on the local derivative pattern (LDP) for 3D face recognition. Compared to the local binary pattern (LBP), LDP can capture more detailed information by encoding directional pattern features. It is based on the local derivative variations that extract high-order local information. We propose a novel discriminative facial shape descriptor, local normal derivative pattern (LNDP) that extracts LDP from the surface normal. Using surface normal, the orientation of a surface at each point is determined as a first-order surface differential. Three normal component images are extracted by estimating three components of normal vectors in x, y, and z channels. Each normal component is divided into several patches and encoded using LDP. The final descriptor is created by concatenating histograms of the LNDP on each patch. Experimental results on two famous 3D face databases, FRGC v2.0 and Bosphorus illustrate the effectiveness of the proposed descriptor.

Index Terms— 3D face, surface normal, local derivative pattern, high-order local pattern

1. INTRODUCTION

The face has proved to be the most popular biometrics to identify a person compared to other modalities like iris, fingerprint, and etc. It is easily captured with non-contact sensors and publicly acceptable. During past decades, the 2D face has been applied in many applications. However, there are limitations and disadvantages including pose and illumination variations. Recently, to overcome such these degradation conditions by 2D face, 3D face has widely attracted researchers' attention. Generally, face recognition approaches based on the type of the features are divided into global and local feature-based methods [1]. Unlike global representation, local descriptors extract prominent points, patches or regions of the face to handle facial expression, occlusion, and missing data effectively. There are some regions on the face that are invariant under facial expression, such as regions around the nose that are employed in the recognition task. One group of the local feature-based methods is based on the keypoints detection and descriptor extraction around each interest point such as SIFT-based keypoints [2], and keypoints detection using minimum and maximum curvatures [3]. Shape curvature-based SIFT-like method also has been applied to the 3D face as local features in combination with the 2D facial descriptor in [4]. In some works, keypoints are detected directly from the facial mesh to handle large pose variations and occlusions such as meshSIFT [5], and meshDOG [6]. Another type of local features that are less sparse compared to keypoints are computed as facial curves that contain rich geometrical information [7]. In addition, local features can be extracted from several patches of the facial surface or from several regions of the face such as applying local patterns [8]. One of the most efficient and popular local surface patterns is the local binary pattern (LBP) that is first proposed in [9] for texture description. LBP on the 3D face is presented by Huang et al. [10] as the 3DLBP local descriptor on the depth map in combination with global matching. Later, a generalized version of 3DLBP, multiscale extended local binary patterns (eLBP) [11], is proposed for histogram-based 3D face recognition using SIFT-based matching strategy. An LBP feature-based 3D face division scheme is proposed in [12] that extracts and encodes depth and normal information from each region of the face using LBP descriptor. In [13], a multi-scale and multi-component local normal pattern (MSMC-LNP) is presented as a discriminative facial surface descriptor where each normal component image is encoded using LBP scheme. LBP has been applied directly on 3D facial mesh by Werghi et al. [8] to construct a grid of facial regions. The idea of LBP is based on the micropatterns composition of the image that presents the first order circular derivative pattern.

In this paper, we propose a novel local descriptor, LNDP that works based on the high-order local derivative pattern [14] for local feature-based 3D face recognition. The high-order LDP provides the general scheme to compute directional features for 2D images. The change of derivative directions among local neighbors in LBP-based pattern is captured using second-order LDP. Unlike LBP that is a non-directional first-order local pattern, n^{th}-order LDP calculates directional relationship that has achieved superior performance compared to LBP. We have extended LDP to surface normal components as LNDP to propose the robust descriptor that is computationally efficient and training free for 3D face recognition. The surface normal that computes the facial surface orientation extracts discriminative and detailed information compared to range data. In this work, we assume that always there are local regions on semi-rigid and rigid areas that are invariant under facial expression.

This paper is organized as follows. The proposed descriptor for 3D face recognition is described in section 2 in detail. Experimental results and comparison with other researchers' work are illustrated in section 3. Section 4 concludes the paper and gives the future directions.

2. PROPOSED METHOD

2.1. Surface normal

The surface normals contain more detailed and robust information compared to depth image for 3D data [13]. In this paper, encoded facial surface components in three x, y, and z channels are proposed to extract discriminative information. The facial surface normal components are estimated by adopting the optimization based method

Fig. 1. From left to right: range image, normal component x, normal component y, and normal component z

presented in [15]. The normal vector of a given point is computed using the plane normal vector that the point belongs to it. The plane is fitted to the local neighbors of the given point by minimizing a cost function, for instance, the average distance from the neighbors. For a set of n points, the data matrix is considered as follows:

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

where the 3D coordinate is defined as $p_i = [p_{ix}, p_{iy}, p_{iz}]^T$. A normal vector $n_i = [n_{ix}, n_{iy}, n_{iz}]^T$ is estimated using l neighbor points Q_i around p_i.

$$Q_i = [q_{i1}, q_{i2}, ..., q_{il}]^T \quad (2)$$

It is calculated by solving the optimization problem as:

$$minA(p_i, Q_i, n_i) \quad (3)$$

where A is the cost function and we consider a neighborhood of 5×5 window. Consequently, each 3D point has its normal component in x, y, and z channel that results in three images as follows:

$$N(P) = Nx(j,k), Ny(j,k), Nz(j,k), \\ 1 \leq j \leq m, 1 \leq k \leq n \quad (4)$$

Fig. 1 illustrates a sample range image and its corresponding three normal component images. As the figure shows, the surface normal component images contain more information compared to the range image.

2.2. Local Normal derivative pattern (LNDP)

After calculating normal components from 3D face data, the LBP-based facial encoding is applied to compute facial information. Inspired by LBP which is introduced as a powerful tool to describe texture images and applied in other applications such as face recognition, the local derivative pattern [14] is applied to encode facial normal components. Unlike LBP which is a non-directional first-derivative binary operator, LDP is defined as directional high-order derivative pattern that contains more detailed features. For a normal component N(P), the first-order derivatives for α equal to 0°, 45°, 90°, and $135^\circ$are computed as follow:

$$N'_{0^\circ}(P_0) = N(P_0) - N(P_4) \quad (5)$$

$$N'_{45^\circ}(P_0) = N(P_0) - N(P_3) \quad (6)$$

$$N'_{90^\circ}(P_0) = N(P_0) - N(P_2) \quad (7)$$

$$N'_{135^\circ}(P_0) = N(P_0) - N(P_1) \quad (8)$$

where P_0 is a point in N(P) and P_i are the neighboring points around P_0 as it is shown in Fig. 2a. The second-order normal local derivative pattern is calculated using the below equation:

$$LNDP^2_\alpha(P_0) = (f(N'_\alpha(P_0), N'_\alpha(P_1)), f(N'_\alpha(P_0), N'_\alpha(P_2))..., \\ f(N'_\alpha(P_0), N'_\alpha(P_8))) \quad (9)$$

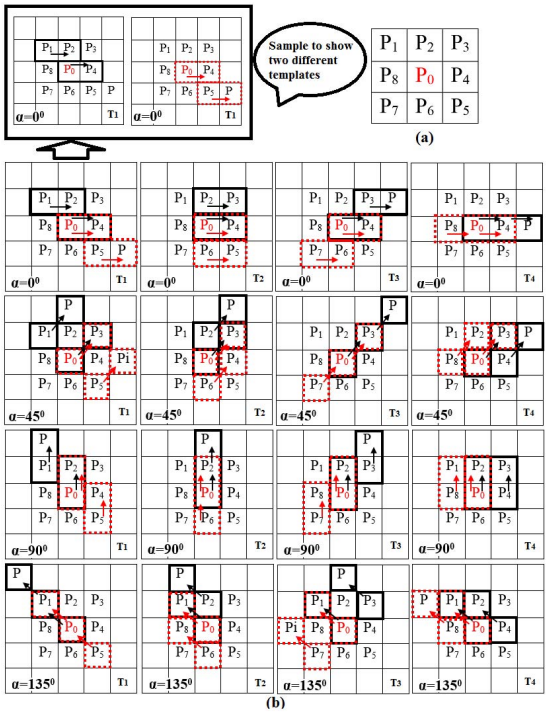

Fig. 2. (a) 8-neighborhood around P_0, (b) 32 templates for $\alpha = 0^\circ, 45^\circ, 90^\circ, 135^\circ$. The black and dashed red lines represent two different templates as the top left sample shows, i for four templates in each row from left to right is: Templates T1 ($i = 1, 5$), templates T2 ($i = 2, 6$), templates T3 ($i = 3, 7$), and templates T4 ($i = 4, 8$).

where binary coding function $f(N'_\alpha(P_0), N'_\alpha(P_i))$ is equal to 0 if $N'_\alpha(P_0).N'_\alpha(P_i) > 0$ and equal to 1 if $N'_\alpha(P_0).N'_\alpha(P_i) <= 0$. To calculate f in different directions we apply 32 templates shown in Fig. 2b. The second-order LNDP is used to encode the change of the neighborhood derivative directions in the local region. Different templates are applied to calculate $f(N^{n-1}_\alpha(P_0), N^{n-1}_\alpha(P_i))$ for different α based on the Fig. 2b.

Using concatenation of 8-bits LNDP for four directions the second-order descriptor is constructed:

$$LNDP^2(P) = \{LNDP^2_\alpha(P) | \alpha = 0^\circ, 45^\circ, 90^\circ, 135^\circ\} \quad (10)$$

Since there are three components for surface normal based on the equation 4, the multi-components second-order LNDP is considered as:

$$LNDP^2(P) = \{LNDPx^2(j,k), LNDPy^2(j,k), \\ LNDPz^2(j,k)\}, 1 \leq j \leq m, 1 \leq k \leq n \quad (11)$$

High-order LNDP can extract more discriminative information by encoding derivative direction variations. In LBP, the relationship between the central point and its 8 neighbors is calculated. However, LNDP contains more spatial information in a local area by encoding different distinct spatial relations. The third-order local normal derivative pattern is computed using the following equation:

$$LNDP^3_\alpha(P_0) = (f(N''_\alpha(P_0), N''_\alpha(P_1)), f(N''_\alpha(P_0), N''_\alpha(P_2))..., \\ f(N''_\alpha(P_0), N''_\alpha(P_8))) \quad (12)$$

As it is found from the above equation, the first step is the calculation of the second-order derivatives for different directions. The high-order descriptor extracts more local details from the image. Increasing the order of the operator extracts more and more detailed patterns that tend to be a noisy image instead of the image with detailed discriminative features. Therefore, the performance of the system is evaluated under different orders and the best order is selected for our proposed descriptor. The general scheme to compute n^{th}-order of LNDP is calculated as follows:

$$LNDP_{\alpha}^{n}(P_0) = (f(N_{\alpha}^{n-1}(P_0), N_{\alpha}^{n-1}(P_1)), f(N_{\alpha}^{n-1}(P_0), N_{\alpha}^{n-1}(P_2))..., f(N_{\alpha}^{n-1}(P_0), N_{\alpha}^{n-1}(P_8))) \quad (13)$$

where f is calculated based on the explanation of the equation 9. The decimal value is calculated to generate the integer value of the descriptor as follows:

$$LNDP_{\alpha}^{n}(P_0) = \sum_{l=1}^{L} LNDP_{\alpha}^{n}(P_0) \times 2^{l-1} \quad (14)$$

where L is the number of the neighborhood points around the central point P_0. To describe the micropatterns of the shape variations for a direction α in the 3D image and extend LNDP to a local region pattern, it is modeled to the histogram-based descriptor as follows:

$$HLNDP_{\alpha}^{n} = \sum_{j,k} LNDP_{\alpha}^{n}(A(N_{x,y,z}(j,k)) = v), \quad j = 1, ..., m, k = 1, ..., n, v = 0, ..., V-1 \quad (15)$$

where N is the normal component in three different channels, $V-1$ is the maximum value of $HLNDP_{\alpha}^{n}$, j and k are the numbers of rows and columns in the normal component images respectively, $HLNDP_{\alpha}^{n}(A) = 1$ if A is true, else it is equal to 0. To create the discriminative descriptor with spatial information of the facial surface, we divide normal components of x, y, and z channels into several local patches. We re-size each range image into 120×96 to compute facial normal components. Each normal image is divided into local patches with 12×12 size and 10×8 windows. The histogram of the local normal derivative pattern ($HLNDP$) is extracted from each local patch and the final proposed descriptor is constructed using the concatenation of the histograms from all local patches. The final descriptor consists of three histograms including $HLNDP_x, HLNDP_y$, and $HLNDP_z$ for four different directions (See Fig. 3). The summary of the n^th-order LNDP is presented in Algorithm 1. The proposed descriptor is training free and histogram intersection is used for the recognition step. The similarity between the gallery sample with histogram H_G and a probe face with histogram H_Q is measured as follows:

$$S(H_G, H_Q) = \sum_{i=1}^{C} min(H_G(i), H_Q(i)) \quad (16)$$

where S is the similarity between two histograms and C is the number of histograms bins. This measurement computes common areas of two histograms. This simple operation makes our proposed method efficient with low computational complexity.

3. EXPERIMENTAL RESULTS

In this section, the proposed descriptor is evaluated on different 3D face databases including FRGC v2.0 [16] and Bosphorus [17].

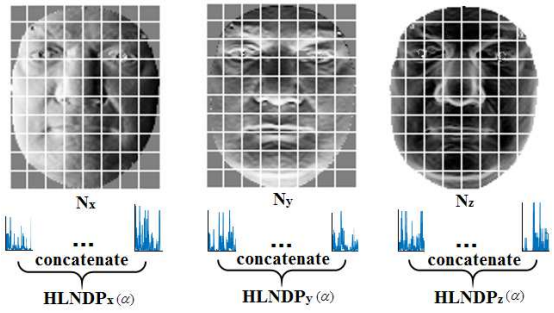

Fig. 3. $HLNDP(\alpha)$ descriptor construction for x, y, and z facial normal components

Algorithm 1 n^{th}-order LNDP

Input: 3D face data P
for each point in P **do**
  Calculate normal components (N_x, N_y, and N_z)
end for
for each N **do**
  Divide into 10×8 patches
end for
for $alpha = 0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}$ **do**
  for each patch **do**
    for each pixel in patch of N **do**
      Apply Equation (13)
    end for
    Encode $LNDP$ using Equation (14)
    Histogram construction
  end for
  Concatenate the histogram for different patches
end for
Concatenate the histogram for different α
return $HLNDP_x^n, HLNDP_y^n, HLNDP_z^n$

FRGC v2.0 database consists of 4007 scans of 466 persons with neutral and non-neutral expressions, and minimum pose changes. The Bosphorus database comprises of 4666 scans of 105 subjects including 34 facial expressions (action units and six emotions), 13 pose variations (yaw, pitch, and cross-rotations), 4 occlusions (eye, and mouth with hand, hair, and eyeglasses). We apply 3D face models pre-processing tool [18] as follows: to remove spike and noises using median filter, hole filling by fitting square surface, nose detection by curvature-based method, and region of interest (ROI) cropping. To correct pose, five frontal scans with neutral expression from each database are used as models to align faces by applying the iterative closest point (ICP) algorithm [19]. The comparison of different orders of the proposed descriptor has been presented in Fig. 4 using first (by applying LBP on normal components), second, third, and fourth-order local normal patterns on FRGC v2.0 database. As results demonstrate the performance increases from first to second and to the third-order descriptor. It means high-order local pattern extracts more detailed and discriminative information. However, the result of forth-order LNDP shows extracting more and more detailed information by increasing the order converts the image to the noisy data that decreases the performance. Based on the results in Fig. 4, the recognition rate of the score level fusion of third-order descriptor outperforms other orders. In the next experiments, the third-order

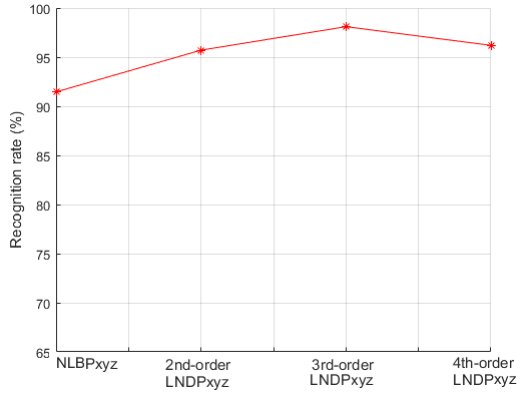

Fig. 4. Comparison of different orders of proposed descriptor on FRGC v2.0

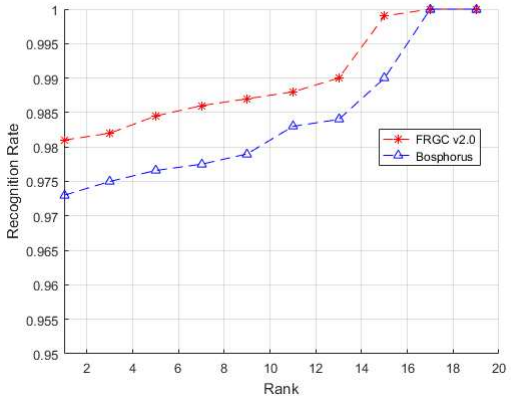

Fig. 5. CMC curves of score-level fusion third-order LNDP on FRGC v2.0 and Bosphorus

Table 1. Comparison of RR1 on FRGC v2.0 for LBP, LDP, and LNDP descriptors

Descriptor	RR1
$DepthLBP$	86.2%
$DLDP^3$	89.08%
$LNDP^3_x$	92.53%
$LNDP^3_y$	91.18%
$LNDP^3_z$	96.04%
$LNDP^3_{xyz}$	98.1%

Table 2. Comparison of results from LBP-based methods

Methods	RR1 (FRGC v2.0)	RR1 (Bosphorus)
MS-eLBP-DFs[11]	97.6%	97%
V-LBP[12]	94.89%(900/150)	-
MSMC-LNP[13]	96.3%	95.4%(2797/105)
DLBP[21]	90%	90%
region-based-eLBP[20]	97.8%	-
$LNDP^3_{xyz}$	**98.1**%	**97.3**%

descriptor is applied to show the proposed descriptor effectiveness. The third-order local normal derivative pattern $LNDP^3_x$, $LNDP^3_y$, $LNDP^3_z$, and their score level fusion $LNDP^3_{xyz}$ are compared with the descriptors including first-order local pattern extracted on depth map (DepthLBP), and the proposed local derivative pattern on depth map (DLDP) with the same matching approach. Rank one recognition rate, $RR1$, for neutral versus all experiment [16] on FRGC v2.0 database is presented in Table 1. The comparison between results of LBP and LDP on the depth image shows DLDP performs 2.88% better than DepthLBP. Applying third-order LDP on the normal components and their score-level fusion $LNDP^3_{xyz}$ outperforms other descriptors and improves the recognition rate effectively. In our experiments, the first sample of each person with a neutral expression is considered as the gallery and other scans as the probe. Based on the results in the Table 1, score-level fusion third-order LNDP descriptor performs much better than other descriptors. In this way, we continue our experiments with this descriptor. We have also evaluated our proposed descriptor on Bosphorus database to further validate the proposed descriptor effectiveness. In the experiment, we consider nearly frontal images with a facial expression or partial occlusion which consist of 3301 samples. Like the test on FRGC v2.0, we consider the first neutral sample of each subject as the gallery and the remain scans as the probe. In this way, we have 105 scans for the gallery and 3196 scans for the probe. To illustrate recognition efficiency, the Cumulative Match Characteristics (CMC) curves on two databases are presented in Fig. 5. The descriptor has been compared with the LBP-based methods in Table 2 to demonstrate its effectiveness compared to state-of-the-art. From the table, it is obvious that our new descriptor outperforms other methods with high recognition rate on both databases under the same experimental conditions. The result of the method in [11] is close to this paper. However, the extended LBP in combination with SIFT-like matching strategy in a hybrid way increases computational complexity. Moreover, the works [13, 20] with high recognition rates apply sparse representation classifier that needs training step and increases the system complexity.

4. CONCLUSION

A novel descriptor for 3D face recognition using high-order local pattern is proposed based on the conventional LBP framework. After pre-processing to remove noise and spikes, make pose correction, and extract ROI, facial surface normal components from 3D data, in x, y, and z channels are computed. Then, high-order LNDP is calculated to extract more detailed distinct information from the 3D facial image. The score-level fusion of three high-order $LNDP_x$, $LNDP_y$, and $LNDP_z$ is used to improve the performance. The proposed method is training free and computationally efficient with high recognition rate. Experimental results show the score-level fusion of histogram-based LNDP descriptors is much more discriminative than the LDP on normal components, LBP and LDP on depth images. The proposed descriptor can be applied in other 3D object recognition as well. In this work, we concentrated on the faces under expression. In the future, we plan to apply more powerful classifier that can enhance the recognition results under extreme pose variations and missing data.

5. ACKNOWLEDGMENTS

This research was supported in part by the Canada Research Chair Program and the NSERC grant.

6. REFERENCES

[1] M. Bennamoun, Y. Guo, and F. Sohel, "Feature selection for 2D and 3D face recognition," *Wiley Encyclopedia of Electrical and Electronics Engineering*, 2015.

[2] 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.

[3] H. Li, D. Huang, J.-M. Morvan, Y. Wang, 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.

[4] S. Soltanpour and Q. J. Wu, "Multimodal 2D-3D face recognition using local descriptors: pyramidal shape map and structural context," *IET Biometrics*, vol. 6, no. 1, pp. 27–35, 2016.

[5] 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.

[6] S. Berretti, N. Werghi, A. Del Bimbo, and P. Pala, "Selecting stable keypoints and local descriptors for person identification using 3D face scans," *The Visual Computer*, vol. 30, no. 11, pp. 1275–1292, 2014.

[7] H. Drira, B. B. Amor, A. Srivastava, M. Daoudi, and R. Slama, "3D face recognition under expressions, occlusions, and pose variations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2270–2283, 2013.

[8] N. Werghi, C. Tortorici, S. Berretti, and A. Del Bimbo, "Boosting 3D LBP-based face recognition by fusing shape and texture descriptors on the mesh," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 5, pp. 964–979, 2016.

[9] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[10] Y. Huang, Y. Wang, and T. Tan, "Combining statistics of geometrical and correlative features for 3D face recognition," in *BMVC*, 2006, pp. 879–888.

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

[12] H. Tang, B. Yin, Y. Sun, and Y. Hu, "3D face recognition using local binary patterns," *Signal Processing*, vol. 93, no. 8, pp. 2190–2198, 2013.

[13] 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, pp. 179–193, 2014.

[14] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor," *IEEE transactions on image processing*, vol. 19, no. 2, pp. 533–544, 2010.

[15] K. Klasing, D. Althoff, D. Wollherr, and M. Buss, "Comparison of surface normal estimation methods for range sensing applications," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2009, pp. 3206–3211.

[16] 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 (CVPR)*, 2005, pp. 947–954.

[17] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis," in *European Workshop on Biometrics and Identity Management*. Springer, 2008, pp. 47–56.

[18] P. Szeptycki, M. Ardabilian, and L. Chen, "A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking," in *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, 2009, pp. 1–6.

[19] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *International journal of computer vision*, vol. 13, no. 2, pp. 119–152, 1994.

[20] S. Lv, F. Da, and X. Deng, "A 3D face recognition method using region-based extended local binary pattern," in *IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 3635–3639.

[21] A. Aissaoui, J. Martinet, and C. Djeraba, "DLBP: A novel descriptor for depth image based face recognition," in *IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 298–302.