Face Recognition with Multi-channel Local Mesh High-order Pattern Descriptor and Convolutional Neural Network

M. Daud Abdullah Asif, Yongsheng Gao
School of Engineering,
Griffith Univeristy,
Queensland, Australia
Email: daud.asif@griffithuni.edu.au,
yongsheng.gao@griffith.edu.au

Jun Zhou
School of Information Communication Technology,
Griffith University,
Queensland, Australia
Email: jun.zhou@griffith.edu.au

***Abstract*— In this paper, we propose a novel Local Mesh High-order Pattern Descriptor (LMHPD) for face recognition. This description is constructed in a high-order derivative space and is integrated with a Convolutional Neural Network (CNN) architecture. Based on the information collected at a local neighborhood of reference pixel with diverse radiuses and mesh angles, a vectorized feature representation of the reference pixel is generated to provide micro-patterns. They are then converted to multi-channels to use in conjunction with the CNN. The CNN adopted in the proposed architecture is generic and very compact with a small number of convolutional layers. However, LMHPD is derived in such a way that it can work with most of the available CNN architectures. For keeping the computational cost and time complexity at the minimum, we propose a lighter approach of high-order texture descriptor with CNN architecture that can effectively extract discriminative face features. Extensive experiments on Extended Yale B and CMU-PIE datasets show that our method consistently outperforms several alternative descriptors for face recognition under various circumstances.**

Keywords—Face Recognition, High-Order Local Pattern Descriptor, Multi-channel Micro-patterns, Convolutional Neural Network.

I. INTRODUCTION

Face recognition has been intensively studied during the past decades [1] – [4]. Like in many recognition tasks, feature descriptor is a key step that affects the recognition performance. Face descriptors shall be able to maximize the inter-person margin, minimize the intra-person overlapping, and be computed with low computational cost with large amount of data. Generating an effective and efficient face descriptor not only requires an optimized classifier but also requires optimal extraction of face features. The proposed descriptor presented in this paper balances all these criteria in conjunction with CNN to produce highly accurate recognition results.

Widely used face descriptors include local pattern descriptors [5]-[8], Eigenface [9], Fisherface [10] and manifold based learning methods [11] – [12]. The importance of texture description has been well recognized in face recognition to represent the spatial structure information from input data. For example, Local Binary Pattern (LBP) [13] – [15] is a descriptor with low computational cost. It has achieved better performance in the past as compared to Eigenface [9] and Fisherface [10].

LBP divides a face image into sub-regions to extract the micro-patterns. To generate this micro-pattern, each reference pixel is compared with its neighbours, which can be demonstrated by spatial histograms to represent the texture information. The extension of this model to higher order derivative space was proposed by Zhang et al. in local derivative pattern (LDP) [16]. This descriptor was proved to be more effective than the LBP as it can more successfully capture the discriminative information by extracting face features in a high-order derivative space by computing the derivatives of face input image in four directions and encoding the relationship of the reference pixel and its neighbours in different directions. Inspired by the idea of LDP, Murala et al. presented local tetra patterns (LTrP) which maps two distinct values into four distinct values by using two high-order derivative direction patterns for generating more robust feature descriptors [17]. In comparison with the previous approaches, local vector pattern (LVP) computes the vector in four different directions. Micro-patterns are computed based on the encoding scheme of comparative space transform (CST) [18]. The reason high-order pattern descriptors are more effective is that the first-order derivative pattern descriptors fail to capture the discriminative information at distance neighbourhood.

Another factor to consider in face recognition is the relationship between the traditional hand-crafted features and how they can be used with deep neural networks, e.g. convolutional neural networks (CNNs). Over the last several years, CNNs have shown outstanding performance in computer vision and image analysis domains. Inspired by human natural visual perception mechanism, CNNs produce discriminative feature representation from input samples without any prior manipulation [19] – [20].

Deep learning approaches have been applied effectively for face recognition. The LBCNN architecture [25] was based on the local binary patterns with the reformulation of all of its texture analysis algorithms. LBCNN uses up to 150 convolutional layers, which makes it memory as well as time wise, a complex and heavy architecture. On the other hand another analysis of deep learning representation for face recognition was presented recently [26], which used two types CNN architectures. One architecture is VGG-Face network [27], which is pre-trained network on more than 2.6 million images. The second architecture is lightened CNN [28] which uses extra max-feature-map with two networks of four and

978-1-5386-6602-9/18/$31.00 ©2018 IEEE

five convolutional layers, both have two fully connected layers. Although lightened CNN architecture is computationally efficient but the accuracies reported on this architecture are very low [26]. Similar approaches with multiple CNN architectures have also been used for face recognition [29] – [32].

In this paper, we propose a novel high-order local pattern descriptor, Local Mesh High-order Pattern Descriptor (LMHPD), for face recognition. We aim to capture information at different radiuses and angles at reference pixels and their local neighbourhood, and then bridge the gap between high order texture descriptors and CNNs. We have developed a multi-channel framework-encoding scheme within LMHPD, which can work in conjunction with the basic CNN architecture with only eight convolutional layers connected with three average pooling layers to enhance and maximize the robustness of face recognition. This framework is lighter than state-of-the-arts mentioned previously. In our work, the LMHPD can work as a dynamic descriptor in varying directions without obsoleting the information at smaller neighbourhood pixels at all mesh angles.

The rest of the paper is organized as follows. The proposed LMHPD is described in section II. The multi-channel framework for CNN is described in section III. Experimental results on Extended Yale B and CMU-PIE datasets are described in Section IV. Finally, conclusions are drawn in section V.

II. THE PROPOSED LOCAL PATTERN DESCRIPTOR

The proposed Local Mesh High-order Pattern Descriptor (LMHPD) generates multi-channel micro-patterns based on the discriminative information it captures from the dynamic radiuses and mesh angles at 0°, 45°, 90° and 135°. Moreover, it captures all the information from greater mesh angles at the referenced pixel location including (0°, 45°), (0°, 90°), (45°, 90°), (0°, 135°), (45°, 135°) and (90°, 135°). This descriptor captures all the information from the local neighbourhood across dynamic radiuses, angular radiuses and high-order inter-derivative spaces. In addition to the proposed LMHPD, micro-patterns have optimized the multi-channel convergence to the CNN architecture to produce discriminative feature representation of the input datasets.

A. Local Mesh High-order Pattern Descriptor

The proposed Local Mesh High-order Pattern Descriptor (LMHPD) is designed to represent the micro-patterns as a Multi-channel local pattern descriptor, which can be optimized for the CNN architecture. The micro-patterns are obtained by calculating the values between the reference pixel and the neighbourhood pixels at diverse radiuses and mesh angular directions.

Given an image I with local sub-region $I(P_c)$ centered at P_c with various radiuses and angles are shown in Fig. 1., the four first order derivatives along the direction of 0°, 45°, 90° and 135° at the reference pixel P_c are denoted by:

$$I_{\partial,R}(P_c) = I(P_c) - I\left(P_{\left(\frac{\partial}{45}+1\right),R}\right) \tag{1}$$

where R represent the radius of the descriptor, and ∂ is the angle.

We can optimize the descriptor to any radial width by considering the values for R as the number for the width of the radius. For extracting the second-order $LMHPD^2$ at radius R, the pair-wise angular derivatives are derived using the following function:

$$LMHPD^2(P_c) = \{F(I_{\partial,R}(P_0), I_{\beta,R}(P_0)), F(I_{\partial,R}(P_1), I_{\beta,R}(P_1)), \ldots, F(I_{\partial,R}(P_n), I_{\beta,R}(P_n))\} \tag{2}$$

where $F(.,.)$ is a function which encodes the pair-wise binary values between two different mesh angles ∂ and β. These two angles range among 0°, 45°, 90° and 135°. The indices of neighbouring pixels range between $n = 1, 2, \ldots, 8$. This encoding function at any reference pixel P is defined as:

$$F\left(I_{\partial,R}(P_c), I_{\beta,R}(P_c)\right) = \begin{cases} 1, & if\ I_{\partial,R}(P_c) > I_{\beta,R}(P_c) \\ 0, & else \end{cases} \tag{3}$$

If the value of angle ∂ with any radius R at a reference pixel P_c ($I_{\partial,R}(P_c)$) is greater than the value of other pixel P_c at angle β with another radius, then the function assigns a value of one to the pixel under consideration, otherwise the value of zero will be assigned to the reference pixel.

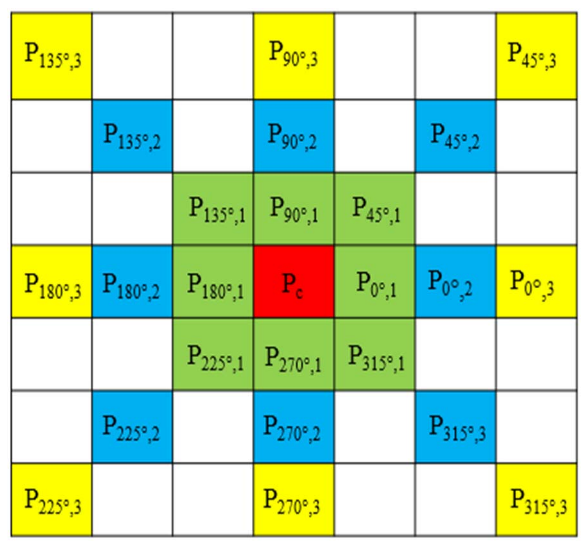

Fig. 1: Adjacent pixels of the center pixel (P_c) with dynamic radiuses and different angles.

With these derivatives at higher order using this encoding function, the $LMHPD^2$ can be calculated by concatenating different ∂, β angle pairs as:

$$LMHPD^2_{\partial,\beta,R}(P_C) = \{LMHPD^2_{\partial,\beta,R}(P_C) | (\partial,\beta)\} \tag{4}$$

where $(\partial,\beta) \epsilon \{(0°,45°), (0°,90°), (45°,90°), (0°,135°), (45°,135°), (90°,135°)\}$.

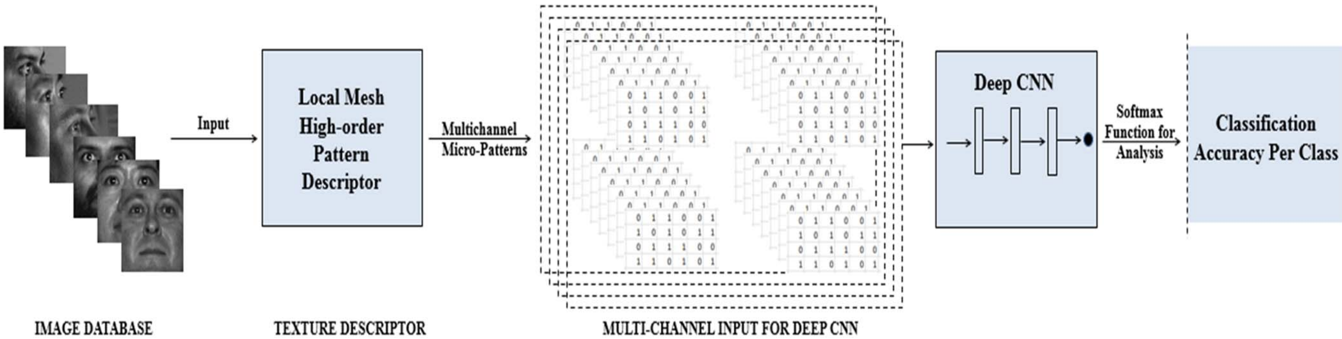

Figure 3. Architecture of the Multi-channel LMHPD configuration with Deep CNN framework.

In this way, for the second-order LMHPD^2, all the mesh angle-pairs (0°, 45°), (0°, 90°), (45°, 90°), (0°, 135°), (45°, 135°) and (90°, 135°) at variable radiuses can be calculated to form micro-patterns. These micro-patterns from all mesh pairs are concatenated to get the final pattern description of all the pixels by considering any pixel as the reference pixel. An example on the calculation of second order LMHPD^2 is given in Fig. 2.

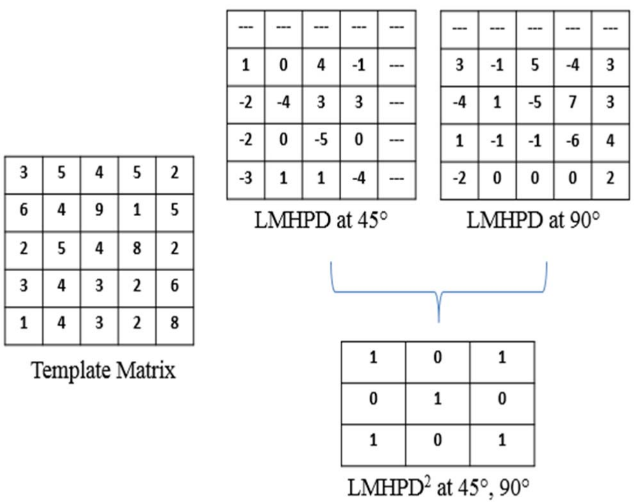

Fig. 2. Example to obtain second-order LMHPD^2 with radius 1 at angle pair of (45°, 90°).

The final feature vector is then considered for the encoding scheme by our proposed model. This encoding scheme converts the final pattern into multi-channels, which become the input to the deep CNN architecture. In the multi-channel framework, we used one CNN architecture with very less number of convolutional layers.

III. CONVOLUTIONAL NETWORK FOR MULTI-CHANNEL LMHPD

For the purpose of optimization of LMHPD with the deep convolutional neural network architecture, the micro-patterns obtained by our proposed second-order descriptor are considered by integrating a multi-channel output layer to CNN. The encoding scheme considers the binary micro-patterns of every pixel in the input image and utilizes them as the input channel. Since the radius can be dynamic in our descriptor, we use the radius of three to get 25 multi-channels for CNN architecture. In case the radius of the descriptor is decreased, the multi-channels of the input for CNN are also decreased because number of bits from the LMHPD will be less than the radius 3.

Table I. Architecture to maximize the genericity of CNN with convolution layers followed by nonlinear ReLU

Layer type	#Filters	Window	Stride
Convolution	96	7×7	
Convolution	96	7×7	
Average pooling		2×2	2
Convolution	96	5×5	
Convolution	96	5×5	
Average pooling		2×2	2
Convolution	96	3×3	
Convolution	96	3×3	
Average pooling		2×2	2
Convolution	192	3×3	
Convolution	192	1×1	
Convolution	# class	1×1	
Global averaging			
Softmax			

We then optimize the deep CNN with all the parameter settings. As in our proposed model, the mesh framework leads to dynamic multi-channel layers for CNN layers. The multi-channel architecture can be varied by mesh capability of the feature descriptor in which all angle pairs are considered for every pixel inside image without losing any information at smaller radiuses. Overall architecture of our proposed approach is shown in Fig. 3.

For the input to the deep CNN architecture, as explained earlier, the radius for our LMHPD was considered three to get 25 multi-channels, as this strategy gets highly efficient optimization with different convolution layers in the deep network with the layers of average pooling and a softmax function to get the class accuracies. Average pooling is not utilized in previous architecture [20]. The CNN architecture used in the proposed architecture is basic and generic with each convolution layer followed by a ReLU nonlinearity.

Table I gives the complete layer information of the CNN architecture used in proposed method. These layers in the architecture are much lighter and less time consuming, computationally as compared to the complex deep CNN architectures. Meanwhile, our framework demonstrates that even with using one CNN and having less number of convolutional layers, the texture descriptor LMHPD proved more efficient than using multiple CNN's with complex architectures, which will increase the complexity of the model as well as time complexity.All the parameter settings throughout the experimental stage were taken constant so that the performance of our texture descriptor can be cross-validated.

In order to utilize the CNN architecture in our method efficiently, we used twelve convolutional layers of 96 and 192 channels. At the initial stage, we integrate the multi-channel micro-pattern input from our feature descriptor LMHPD^2 to the CNN architecture and the last layer softmax function was used for discriminative analysis to compute the average classification accuracy for each class. The configuration settings of the proposed Deep CNN architecture is shown in Table II.

Table II. Configuration parameters of the Deep CNN used in proposed method

Parameter/ Setting	**Configuration/Value**
Optimization method	Adam et al. [21]
Learning rate	0.001
Weight decay	1e-4
Dropout rate	0.3
Batch size	128
Nonlinearity	ReLU

In the proposed approach, 25 Multi-channels input to the Deep CNN architecture was used by the encoding scheme with the LMHPD^2 descriptor. This encoding scheme utilized dynamic radius of the descriptor as multi-channel input integration to the Deep CNN model. Since the optimization method for the whole architecture was extensively tested, we used micro-patterns of six bits to represent 150 bits of each pixel and input all these micro-patterns at three radius of LMHPD^2 to CNN. The original image pixel values were used as the top layer for multi-channels so that they represented the original input image with its micro-patterns from the LMHPD^2 descriptor.

Comparing with other local pattern descriptors, the merits of our proposed LMHPD method can be summarized as follows:

1. The LMHPD further reduces the feature length than LTrP by using the pair-wise mesh angular description of the neighbourhood pixels at varying radiuses.
2. LMHPD descriptor computes the relationship in directional derivatives with mesh angle pairs of (0°, 45°), (0°, 90°), (45°, 90°), (0°, 135°), (45°, 135°) and (90°, 135°) at any radius without losing any information at smaller radiuses. This contributes to the discriminative capability that LDP and LVP descriptors cannot achieve.
3. The proposed architecture is the first approach for face recognition that has bridged the gap between conventional approaches and deep CNN architectures by proposing the Multi-channel encoding scheme for the micro-patterns derived from our LMHPD with mesh ability of the varying the micro-patterns at any referenced pixel location. By combining the deep CNN architecture with our handcrafted high-order pattern descriptor, we enhance the classification accuracy as compared with state-or-the-art techniques. Another thing to be considered here is that instead of using multiple pre-trained CNN architectures, we only use single CNN model in our proposed framework.
4. The deep CNN framework used in this approach contains only one CNN with multi-channel input from LMHPD^2, which makes it less complex and computationally efficient compared to other CNN frameworks. This method does not require any complex re-formulation of traditional texture pattern descriptors and arrangement for computationally complex convolutional layers with high-end deep architectures. This approach obsoletes all the possibilities of complex architectures and frameworks to be made like LBPCNN, LDPCNN and LVPCNN. One of these complex approaches recently developed was LBCNN [25].

IV. Experimental results

We have conducted extensive experiments to demonstrate the performance of our proposed and comparative methods. We used two publicly available face datasets, Extended Yale B [22], [23] and CMU-PIE [24]. All the original face images were normalized based on the location of two eyes. Moreover, the multi-channels of micro-patterns were obtained through the encoding scheme of the proposed LMHPD.

To be fair with our extensive experimental analysis for the proposed texture descriptor, we referenced to the deep CNN with only raw data values from the input images as the input to CNN architecture so that our descriptor can be cross-validated [26]. With only raw data input as multi-channel data, the CNN was unable to achieve the highest accuracies even with much larger number training samples. We performed five folds cross-validation and reported the average class accuracies on the datasets. Results shown in the coming section are evident that our feature descriptor outperforms several benchmark feature extraction classification techniques.

We compared our method with several alternative approaches, including LBP [13], LDP2 [16], LDP3[16], LTrP [17], LVP2 [18], and LVP3[18]. For fair comparisons, equal number of testing and training samples were used in all the experiments. The LBP method was combined with CNN as proposed in this method. This is to give a comparison on how effective our LMHPD method was in the proposed architecture. The other approached were implemented by their original features followed by classification of SVM or KNN.

A. Results on Extended Yale B Database

The Extended Yale B face database was used to conduct experiments with severe illumination variations. It contains 38 total subjects including ten subjects from the original Yale B database. Each subject has 9 poses and 64 different illumination conditions. The cropped images have dimension of 192 x 168 pixels for public downloading. We normalized the images to 96 x 84 pixels for fair comparison with other comparative techniques. Fig. 4. displays samples of these frontal face images from one subject in the datasets. Every subject in the dataset has equal number of pose and illumination variations.

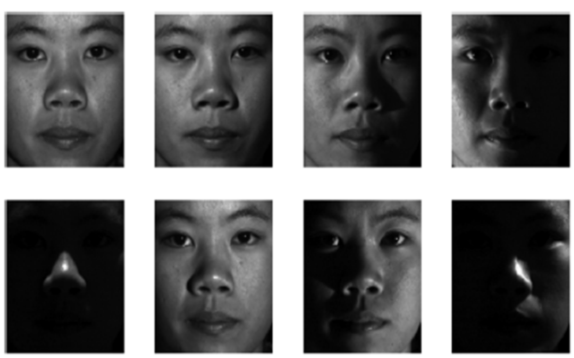

Fig. 4. Samples of the frontal face images from Extended Yale B dataset.

The average classification accuracies are shown in the Table III. Apparently, the LMHPD significantly improves the classification accuracies over other methods even under severe illumination variations.

Most of the results were taken directly from published papers. The latest work [26] provided the comparisons for raw image data with two different CNN architectures. With VGG-Face network [27] and Lightened CNN [28], the accuracies were reported on different subsets of the main dataset, ranging from 100% to 18.28%. On average, the recognition rate is much lower than our proposed architecture, although VGG-Face architecture is computationally complex and contains more than twice convolutional layers as compared to the CNN architecture we used in proposed approach.

Table III. The average classification accuracies (%) on Extended Yale B dataset.

Classification method	Accuracies (%)
LBP [13]* + CNN	85.5 ± 0.22
LDP^2 [16]*	76.4 ± 0.25
LDP^3[16]*	87.9 ± 0.18
LTrP [17]*	91.2 ± 0.24
LVP^2 [18]*	91.6 ± 0.26
LVP^3[18]*	93.3 ± 0.16
$LMHPD^2$ (R=1)	90.2 ± 0.22
$LMHPD^2$ (R=2)	94.8 ± 0.24
$LMHPD^2$ (R=3)	**97.6 ± 0.15**

*Results obtained from published papers.

We cross validated all the results. In every cross validation step, the dataset the training and testing samples were randomly selected. This technique claims the robustness of our texture descriptor that it performed better in every step than other compared classification techniques.

B. Results on CMU-PIE database

We further compared the results with publicly available face datasets CMU-PIE. This dataset contains face images in varying pose, illumination and expression from 68 subjects. The dimension of each image in the dataset was cropped and normalized to 80 x 80 pixels based on the location of two eyes.

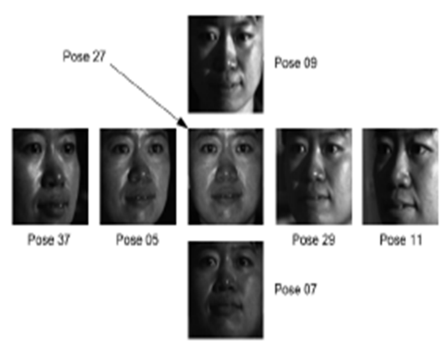

Fig. 5. Samples of the normalized face images of one subject from the CMU-PIE dataset.

Fig. 5. shows examples of the normalized images from one subject in the CMU-PIE dataset. Every subject was exposed to the same varying pose, illumination and expressions variations. The experiments on all these subjects were performed with five folds cross-validation and we calculated average class accuracies.

Our descriptor performed quite well as compared with the previous available classification techniques. The experimental results based on the proposed architecture are shown in Table IV. The LMHPD exhibits better performance than the other methods under same parameters. Although the number of images in this dataset were small, our proposed method still outperforms all the techniques under comparison. Usually, this is not the case in complex deep CNN architectures as they require high number of images to get high accuracies. In some cases, the number of images used for deep architectures are in millions. Therefore, it will be effective to utilize the proposed strategy by adopting the multi-channel framework. This experiment is a typical example that shows the advantage of combining handcrafted features with deep learning framework when the number of training samples is in moderate level. Some other researchers utilized very deep CNN frameworks which were computationally complex to achieve high accuracies and used more than one CNN's. Though the accuracies can be improved, they are computationally expensive than our proposed method.

Table IV. The average classification accuracies (%) on CMU-PIE dataset.

Classification method	**Accuracies (%)**
VGG-Face [26][27]*	93.16 (Average)
Lightened CNN [26][28]*	20.51 (Average)
LBP [13]* + CNN	89.5 ± 0.13
LDP^2 [16]*	88.6 ± 0.22
LDP^3 [16]*	91.8 ± 0.25
LTrP [17]*	89.8 ± 0.18
LVP^2 [18]*	92.5 ± 0.28
LVP^3 [18]*	94.2 ± 0.23
$LMHPD^2$ (R=1)	91.6 ± 0.21
$LMHPD^2$(R=2)	95.2 ± 0.16
$LMHPD^2$(R=3)	**96.8 ± 0.16**

*Results obtained from published papers.

Similarly, VGG-face [27] and Lightened CNN [28] have reported average accuracies on this dataset as well. As shown in Table IV, the accuracies achieved by [26] are much lower than our proposed framework. Due to the dynamic mesh capability of the descriptor, if we increase or decrease the radius of the descriptor, we can vary the multi-layer output from the LMHPD and it can increase the multi-layers either if R is increased or decrease the layers if R is decreased as explained earlier in Equation (2). In both cases, the mesh architecture of angle pairs is preserved. We get the optimized results when R is chosen to be three with the descriptor and these form 25 layers for input to CNN architecture. The deep convolutional neural network DCNN chosen in our experiments has the generic layers and very basic and simple architecture. With complex architectures, we can achieve higher accuracies as well but that will increases the time complexity of our framework.

Another thing to be considered here is that our model is at the second order derivative and some other descriptors under comparison are at 3^{rd} order derivative. From this approach, it is evident from experiments that our descriptor has inherent characteristics of robustness in extraction of discriminative features.

V. Conclusion

In this paper, a novel descriptor is devised and investigated for generating effective representation for texture analysis in face recognition. First, we developed a novel descriptor LMHPD and then, we made it capable of bridging the gap between conventional and deep neural network architectures by making the encoding scheme with the descriptor. The descriptor is a mesh with varying radiuses and multiple angle pairs to get the powerful discriminative capabilities over several other classification techniques.

The advantage of using such an approach is that, it has low computational complexity and the deep CNN architecture is generic and much lighter which is evident from less number of convolutional layers. Even with less number of images than required by the traditional deep neural network architectures, our texture descriptor is capable of handling fewer number of public database images with less complex CNN framework which is evident by experiments on public datasets with less number of images. The experimental analysis on face databases with different lighting conditions, expression variations and different poses demonstrates that the proposed approach has strong discriminative capabilities.

The proposed model has potential of being applied to other pattern recognition applications. These applications include object recognition, gender recognition, natural scenarios classification and many more.

References

[1] X. Xie, and K.-M. Lam, "Gabor-based kernel PCA with double nonlinear mapping for face recognition with a single face image," IEEE Trans. Image Process., vol. 15, no. 9, pp. 2481–2492, Sep. 2006.

[2] X. Chai, S. Shan, X. Chen, and W. Gao, "Locally linear regression for pose-invariant face recognition," IEEE Trans. Image Process., vol. 16, no. 7, pp. 1716–1725, Jul. 2007.

[3] W.-P. Choi, S.-H. Tse, K.-W. Wong, and K.-M. Lam, "Simplified Gabor wavelets for human face recognition," Pattern Recognit., vol. 41, no. 3, pp. 1186–1199, Mar. 2008.

[4] J. Y. Choi, K. N. Plataniotis, and Y. M. Ro, "Face feature weighted fusion based on fuzzy membership degree for video face recognition," IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 42, no. 4, pp. 1270–1282, Aug. 2012.

[5] T. Jabid, M. H. Kabir, and O. Chae, "Local directional pattern (LDP) for face recognition," in Proc. IEEE Int. Conf. Consum. Electron.,Mar. 2010, pp. 329– 30.

[6] J. Y. Choi, Y. M. Ro, and K. N. Plataniotis, "A comparative study of preprocessing mismatch effects in color image based face recognition," Pattern Recognit., vol. 44, no. 2, pp. 412–430, 2011.

[7] J. Y. Choi, Y. M. Ro, and K. N. Plataniotis, "Color local texture features for color face recognition," IEEE Trans. Image Process., vol. 21, no. 3, pp. 1366– 380, Mar. 2012.

[8] R. Rivera, J. R. Castillo, and O. Chae, "Local directional number pattern for face analysis: Face and expression recognition," IEEE Trans. Image Process., vol. 22, no. 5, pp. 1740–1752, May 2013.

[9] M. Turk, and A. Pentland, "Eigenfaces for recognition," J. Cognit. Neurosci., vol. 3, no. 1, pp. 71–86, 1991.

[10] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 711–720, Jul. 1997.

[11] D. Cai, X. He, J. Han, and H. Zhang, "Orthogonal Laplacianfaces for face recognition," IEEE Trans. Image Process., vol. 15, no. 11, pp. 3608–3614, Nov. 2006.

[12] H. F. Hu, "Orthogonal neighborhood preserving discriminant analysis for face recognition," Pattern Recognit., vol. 41, no. 6, pp. 2045–2054, 2008.

[13] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971–987, Jul. 2002.

[14] M. Pietikäinen, T. Ojala, and Z. Xu, "Rotation-invariant texture classification using feature distributions," Pattern Recognit., vol. 33, no. 1, pp. 43–52, 2000.

[15] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 12, pp. 2037–2041, Dec. 2006

[16] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with higher-order local pattern descriptor," IEEE Trans. Image Process., vol. 19, no. 2, pp. 533–544, Feb. 2010.

[17] S. Murala, R. P. Maheshwari, and R. Balasubramanian, "Local tetra patterns: A new feature descriptor for content-based image retrieval," IEEE Trans. Image Process., vol. 21, no. 5, pp. 2874–2886, May 2012.

[18] K.-C. Fan, and T.-Y. Hung, "A novel local pattern descriptor—Local vector pattern in high-order derivative space for face recognition," IEEE Trans. Image Process., vol. 23, no. 7, pp. 2877–2891, May 2014.

[19] Zeiler, M.D., and Fergus, R., "Visualizing and understanding convolutional networks," European Conference on Computer Vision, Springer, 818–833, 2014.

[20] Krizhevsky, A., Sutskever, and I., Hinton, G.E., "Imagenet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, 1097–1105, 2012.

[21] Kingma, D.P., Ba, and J. Adam, "A method for stochastic optimization," International Conf. on Learning Representation, 2015.

[22] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 6, pp. 643-660, Jun.2001.

[23] K. C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 5, pp. 684-698, May 2005.

[24] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, pp. 1615–1618, Dec. 2003.

[25] F. J. Xu, V. N. Boddeti, and M. Savvides, "Local binary convolutional neural networks," IEEE Conference on Computer Vision and Pattern Recognition, July 2017.

[26] M. M. Ghazi, and H. K. Ekenel, "A comprehensive analysis of deep learning based representation for face recognition," IEEE computer society workshop on biometrics, 2016.

[27] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," Birtish Machine Vision Conference, 2015.

[28] X. Wu, R. He, and Z. Sun, "A lightened cnn for deep face representation," Computing Research Repository (CoRR), 2015.

[29] L. He, H. Li, Q. Zhang, and Z. Sun, "Dynamic feature learning for partial face recognition," IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[30] J. Zhao, Y. Cheng, Y. Xu, L. Xiaong, J. Li, and F. Zhao, "Towards pose invariant face recognition in the wild," IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[31] K. Cao, Y. Rong, C. Li, X. Tang, and C. C. Loy, "Pose-robust face recognition via deep residual equivariant mapping," IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[32] X. Sun, P. Wu, S. C. H. Hoi, "Face detection using deep learning: an improved faster RCNN approach," Neurocomputing, vol. 299, pp. 42-50, July 2018.