International Journal of Research (IJR) Vol-1, Issue-7, August 2014 **ISSN 2348-6848**

Face and Expression Analysis Using Local Directional Number Pattern

P.Manju Bala[1], S.Vijayalakshmi[2] & M.O.Ramkumar[3]

[1]Assistant professor, Department of Computer science and Engineering
IFET College of Engineering, Villupuram, Tamilnadu, India
e-mail: pkmanju26@gmail.com

[2]Associate professor, Department of Computer science and Engineering,
IFET College of Engineering, Villupuram, Tamilnadu, India
e-mail: svlakshmi_81@yahoo.co.in

[3]Assistant professor, Department of Computer science and Engineering
IFET College of Engineering, Villupuram, Tamilnadu, India
e-mail: ramkumar.mo86@gmail.com

ABSTRACT

LDN encodes the directional information of the face's textures in a compacted way, generating a more discriminative code than current methods. LDN encodes the structure of a local neighborhood by analyzing its directional information. The edge response in the neighborhood is computed in eight different directions with a compass mask. From all the directions the top positive and negative directions are chosen to produce a meaningful descriptor for different textures with similar structural patterns. The extracted information is encoded using the prominent direction indices such as directional numbers and sign which allows us to distinguish among similar structural patterns that have different intensity transitions. The face is divided into several regions and extract the distribution of the LDN features from them. These features are concatenated into a feature vector and it is used as a face descriptor. Several experiments are performed in which this descriptor performs consistently under noise, expression, illumination, and time lapse variations.

KEYWORDS: Directional number pattern, expression recognition, face descriptor, face recognition, image descriptor

I. INTRODUCTION

In face analysis, a key issue is the descriptor of the face appearance. The efficiency of the descriptor depends on its representation and the ease of extracting it from the face. Ideally, a good descriptor should have a high variance among classes (between different persons or expressions), but little or no variation within classes (same person or expression in different conditions).

There are two common approaches to extract facial features: geometric-feature-based and appearance-based methods. The former encodes the shape and locations of different facial components, which are combined into a feature vector that represents the face. An instance of these methods are the graph-based methods, which use several facial components to create a representation of the face and process it. Facial features are widely used in expression recognition[4], as the pioneer work of Ekman and Friesen identifying six basic emotions produced a system to categorize the expressions, known as Facial Action Coding System, and later it was simplified to the Emotional Facial Action Coding System[1]. However, the geometric-feature-based methods usually require accurate and reliable facial feature detection and tracking, which is difficult to accommodate in many situations. The appearance-based methods use image filters, either on the whole-face, to create holistic features, or some specific face-region, to create local features, to extract the appearance changes in the face image. The performance of the appearance-based methods is excellent in constrained environment but their performance degrade in environmental variation.

The local-feature methods compute the descriptor from parts of the face, and then gather the information into one descriptor. Among these methods are Local Features Analysis, Gabor features, Elastic Bunch Graph Matching, and Local Binary Pattern (LBP). LBP achieved better performance than previous methods, thus it gained popularity, and was studied extensively. Newer methods tried to overcome the shortcomings of LBP, like Local Ternary Pattern (LTP), and Local Directional Pattern (LDiP). Both methods use other information, instead of intensity, to overcome noise and illumination variation problems. In this paper, a face descriptor Local Directional Number Pattern (LDN) proposed for robust face recognition that encodes the structural information and the intensity variations of the face's texture. LDN encodes the structure of a local neighborhood by analyzing its directional information. This method computes the edge responses in the neighborhood, in eight different directions with a compass mask. Then, from all the directions the top positive and negative directions are chosen to produce a meaningful descriptor for different textures with similar structural patterns. This approach allows us to distinguish intensity changes (*e.g.*, from bright to dark and vice versa) in the texture, that otherwise will be missed. Furthermore, our descriptor uses the information of the entire neighborhood, instead of using sparse points for its computation like LBP. Hence this approach conveys more information into the code, yet it is more compact—as it is six bit long

II. RELATED WORK

Current methods have several shortcomings. For example, LBP encodes the local neighborhood intensity by using the center pixel as a threshold for a sparse sample of the neighboring pixels[10]. The few number of pixels used in this method introduce several problems. First, it limits the accuracy of the method. Second, the method discards most of the information in the neighborhood. Finally, it makes the method very sensitive to noise. Moreover, these drawbacks are more evident for bigger neighborhoods. To avoid these problems more information from the neighborhood can be used, as other methods do[7].

Although the use of more information makes these methods more stable, they still encode the information in a similar way as LBP: by marking certain characteristics in a bit string[11]. And despite the simplicity of the bit string coding strategy, it discards most information of the neighborhood. For example, the directional (LDiP) and derivative (LDeP) methods miss some directional information by treating all directions equally. Also, they are sensitive to illumination changes and noise[14], as the bits in the code will flip and the code will represent a totally different characteristic.

Thus, the use of the directional numbers produces a more robust code than a simple bit string. Moreover, the use of principal directions may be similar to a weighted coding scheme, in the sense that not all directions have the same importance. It uses the two principal directional numbers of each neighborhood instead of assigning weights to them. It picks the prominent information of each pixel's neighborhood. Therefore, the proposed method filters and gives more importance to the local information before coding it, while other methods weight the grouped information.

III. PROPOSED WORK

Each face is represented by a LDN histogram (LH). The LH contains fine to coarse information of an image, such as edges, spots, corners and other local texture features. Given that the histogram only encodes the occurrence of certain micro-patterns without location information, to aggregate the location information to the descriptor the face image is divided into small regions, $\{R1, \ldots, RN\}$, and extract a histogram Hi from each region Ri . We create the histogram, Hi , using each code as a bin, and then accumulate all the codes in the region in their respective bin by:

$$\mathrm{H}i\,(c) = \sum_{\substack{(x,y)\in Ri \\ LDN(x,y)=c}} \upsilon\ , \forall c$$

where c is a LDN code, and (x, y) is a pixel position in the region Ri, LDN(x, y) is the LDN code for the position (x, y), and v is the accumulation value—commonly the accumulation value is one. Finally, the LH is computed by concatenating those histograms:

$$\mathrm{LH} = \prod_{i=1}^{N} H^{i}$$

where π is the concatenation operation, and N is the number of regions of the divided face. The spatially combined LH plays the role of a global face feature for the given face. The use of the derivative-Gaussian mask allows us to freely vary the size of the mask. The change in the size allows the coding scheme, LDNG, to capture different characteristics of the face. Hence, a fine to coarse representation is achieved by computing the LDNG σ code at n different σi (which we represent by LDNG $\sigma 1,\ldots,\sigma n$), and by concatenating the histogram of each σi, Hi σi , which is computed in the same way as Eq. (*1*) by using LDNG σ , we can merge the characteristics at different resolutions. We call this mixture of resolutions a multi-LDN histogram (MLH), and it is computed by:

$$MLH_{\sigma 1,\ldots,\sigma N} = \prod_{j=1}^{N} \prod_{i=1}^{n} H_{\sigma i}^{j}$$

where π is the concatenation operation, H$j\sigma\, i$ is the histogram of the LDNG σi code at the $R\,j$ region, and n is the number of σ's used—in this experiments it limit ourselves to three. The change in the mask's size allows our method to capture features in the face that otherwise may be overlooked. It is vital to provide descriptive features for long range pixel interaction. However, previous works do not take into account the long range pixel interaction that takes place outside the coverage of their neighborhood system. We find that combining the local shape information, the relation between the edge responses, and relating the information from different resolutions can better characterize the face's characteristics.

A. Face Recognition

The LH and MLH are used during the face recognition process. The objective is to compare the encoded feature vector from one person with all other candidate's feature vector with the Chi-Square dissimilarity measure. This measure between two feature vectors, F1 and F2, of length N is defined as:

$$\chi 2(\mathrm{F1}, \mathrm{F2}) = \sum_{i=1}^{N} \frac{(\mathrm{F1}(i) - \mathrm{F2}(i))^2}{\mathrm{F1}(i) + \mathrm{F2}(i)}$$

The corresponding face of the feature vector with the lowest measured value indicates the match found.

B. Expression Recognition

Facial expression recognition is performed by using a Support Vector Machine (SVM) to evaluate the performance of the proposed method. SVM is a supervised machine learning technique that implicitly maps the data into a higher dimensional feature space. Consequently, it finds a linear hyper plane, with a maximal margin, to separate the data in different classes in this higher dimensional space. Given a training set of M labeled examples $T = \{(xi , yi) \mid i = 1, . . . , M\}$, where $xi \in \mathrm{R}n$ and $yi \in \{-1, 1\}$, the test data is classified by:

$$f(x) = \mathrm{sign}(\sum_{i=1}^{M} \alpha i\ yi\ K(xi , x) + b)$$

where αi are Lagrange multipliers of dual optimization problem, b is a bias, and $K(\cdot, \cdot)$ is a kernel function. Note that SVM allows domain-specific selection of the kernel function. Although many kernels have been proposed, the most frequently used kernel functions are the linear, polynomial, and Radial Basis Function (RBF) kernels. Given that SVM makes binary decisions, multi-class classification can be achieved by adopting the one-against-one or one against-all techniques. In our work, we opt for one against- one technique, which constructs $k(k-1)/2$ classifiers, that are trained with data from two classes. We perform a grid-search on the hyper-parameters in a 10-fold cross validation scheme for parameter selection, as suggested by Hsu *et al.*. The parameter setting producing the best cross validation accuracy was picked.

IV. EXPERIMENTAL RESULTS

Regarding the length of the proposed descriptor, the basic LDN has 56 different values, and the length of the final descriptor will be a multiple of this length. Consequently, LDNK has a length of 56, and the LDNG codes have a length of 56 n, where n is the number of sigmas used (in our experiments we set $n = 3$). Note that similar methods have descriptors with greater lengths. For example, the basic length of: LBP (in the uniform case) is 59, LDiP is 56, LDeP is 1024, LPQ is 256, LTP (coded as two uniform LBP codes) is 128, and general LTP is 38. However, multi-scale codes, like HGPP, have huge lengths, as the global version (GGPP) length is 256 ns , and the local version (LGPP) length is 256 $nsno$ (where $ns = 5$ is the number of scales, and $no = 8$ is the number of orientations). Furthermore, HGPP is a combination of the local and global versions, which will combine its lengths (note that the use of real and imaginary values will double the length). Due to the length of the HGPP descriptor, we will not compare against it in the following section—see the end of next section for more details on the differences with the literature. Additionally, all these lengths should be multiplied by the grid size. In comparison, our multi-scale descriptor is extremely compact, and the single scale is more compact than other descriptors. Moreover, the execution time of our code is, in average, 37 ms on images of size 100×100 for one mask—we produced this time using a Dual-Core CPU with 2.5 GHz, using un-optimized MATLAB code. Moreover, the codes that use *n* different sigmas will take, in average, *n* times more.

This method is tested for face recognition in several databases: FERET, Yale B, Extended Yale B, LFW , and CAS-PEAL. Moreover, we cropped and normalized all images to 100×100 pixels, based on the ground truth positions of the two eyes and mouth when available, or used a face detector to crop the face. In our experiments, every image is partitioned into 10 ×10 regions for all the methods.

1) FERET Results: We tested the performance of the methods, for the face recognition problem, in accordance to the CSU Face Identification Evaluation System with images from the FERET database. In this problem, given a gallery containing labeled face images of several individuals (one or more face images for each person), we classify a new set of probe images. Thus, we used *fa* image set as gallery and the other four sets as probe images. These sets are *fb*, for expression variation, *fc*, for illumination variation, *dupI* and *dupII*, for time lapse variation.

Given that LDN*G* σ depends on its parameter σ, we test different σ's to analyze the performance of the code when varying this parameter. As Fig. 5 reveals, the code presents an increment in the interval $0.5 \leq \sigma \leq 1.5$. Thus, we test the multi-LDN code, LDN*G* $\sigma1,\sigma2,\sigma3$, for different combinations in this interval. We choose to investigate the combination of the small neighborhoods (3 × 3, 5 × 5, 7 × 7) in LDN*G* 0.3,0.6,0.9, medium neighborhoods (5×5, 7×7, 9×9) in LDN*G* 0.5,1.0,1.5, and large neighborhoods (7×7, 9×9, 11×11) in LDN*G* 1.0,1.3,1.6. The LDN*G* $\sigma1,\sigma2,\sigma3$ codes outperform the results of LDN*K*, and other methods in the expression and time lapse variation data sets (fb, dupI, and dupII). For the intensity variation data set (fc), LBP has the same accuracy as the best LDN*G* code, but not as good as the LDN*K* code. However, for extreme illumination variation LBP's and LDN*K*'s performance considerably drop in comparison to LDN*G* codes—*c.f*. Moreover, GGPP and LPQ produce the best results in fc data set, because they do not rely on intensity. Instead, GGPP and LPQ use phase as main feature to build their code, which makes them more robust to illumination. Regarding the LDN*G* combined codes, the medium neighborhood combination, LDN*G* 0.5,1.0,1.5, performs better than the other two. This high accuracy is due to the σ combination that recovers small to large characteristics, instead of picking only small or large characteristics. Therefore, we can say that the improvement of this assemble is due to the balance of
its masks sizes that range from small to large regions. This behavior is also supported by the high performance of these middle σ's.

2) Noise Evaluation: To evaluate the robustness of the proposed method against noise, we corrupted the probe face images, in the FERET database, with white Gaussian noise, and then try to identify them using the same process as described before Perform this experiment with different levels of noise. The robustness of LDN, against noise, is notable as it outperforms the other methods for every level of noise in every data set. LDiP and LBP have problems overcoming the errors introduced by the noise. However, LDN, due to the use of the directional numbers, has a higher recognition rate. Among the different LDN*G* schemes that we tested, the combination of the medium neighborhoods (LDN*G* 0.5,1.0,1.5) has a higher average recognition rate than the other two. As this combination includes characteristics of several resolutions that are different enough, yet consistent with each other, to represent the face's textures. This balance of the size combination is not outstanding in the other two LDN*G* schemes. Nevertheless, these other LDN*G* schemes produce better results than previous methods. Moreover, each LDN*G* scheme has different characteristics that can be exploited in certain conditions. A high presence of noise in the face makes the textures on the face more difficult to detect, thus the use of the Kirsch and derivative-Gaussian masks stabilize the texture codes.

3) (Extended) Yale B Results: Furthermore, we used the Yale B and the Extended Yale B, which is an
improvement over the former, databases for illumination variation evaluation. The former contains images of ten subjects, and the latter contains images of 38 subjects, both with nine poses and 64 illuminations per pose. And we used the frontal face images of all subjects, each with 64 different illuminations. The faces are divided into five subsets based on the angle of the light source directions. The five subsets are: *Sub 1* (0° to 12°), *Sub 2* (13° to 25°), *Sub 3* (26° to 50°), *Sub 4* (51° to 77°), and *Sub 5* (above 78°). We used *Sub 1* image set as gallery and the other four sets as probe images.

The difficulty of these databases increases for the subsets four and five, due to the illumination angles that cover half of the face with shadows. Fig. 8 shows that the LDN coded faces reveal the facial features in presence of shadows in the face. Moreover present results without pre-processing to evaluate the robustness of the descriptors alone. Nevertheless, this method recovers face features in the dark areas, as it does not rely on intensity, like LBP or LTP. This method is evaluated against other methods: Gradient faces, LDiP, LBP, LBP*w*, LTP, LQP, and GGPP. Most methods perform well in the normal Yale B database, except in the Sub 5 data set, in which the proposed method, Gradient faces, and GGPP outperform the other methods.

All methods, except for LTP, are flawless in the first two sets in the Extended Yale B, which have minor illumination changes. On the other hand, LDN takes advantage of its compass masks, which are more robust against illumination changes, and uses the directional number encoding scheme to produce a more discriminative code. The proposed LDN performs better than LPQ, which use phase information. Additionally, GGPP has two more scales, and a greater length than our method to build its descriptor.

The recognition rate difference of LDN*G*, in average on the last two data sets of the normal Yale B, between Gradient faces and GGPP is 0.7% and 1.3%, respectively. Additionally, on the extended database, LDN*G* is 1.1% better than GGPP, in average on the last two data sets. The LDN*K* code is better than previous codes, but not as good as its Gaussian counterpart. Although Gradient faces has a higher accuracy than LDN codes in the illumination variation problem (in the normal Yale B), it is not robust against expression and time lapse variation. Gradient faces has a non-acceptable recognition rate of 7% in fb, and 1% in dupI and dupII in the FERET database. However, LDN codes showed to be more reliable in different variation conditions.

4) LFW Results: Additionally, we evaluated the proposed method using the Labeled Faces in the Wild (LFW) database, which comprises a collection of annotated faces captured from news articles on the web. This database was published with a specific benchmark, which focuses on the face recognition task of pair matching. In this task, given two face images, the goal is to decide whether the two pictures are of the same individual. This is a binary classification problem, in which the two possible outcomes are same or not-same. We used a straightforward approach for pair matching, in which we considered the distance between two image descriptors, and we learned a threshold that classifies whether the distances correspond to a matching pair. This method can be generalized by using a binary SVM.

5) CAS-PEAL Results: Also, we tested the gallery/probe sets methodology in the CAS-PEAL face database, which contains 99594 images of 1040 individuals (595 males and 445 females) with varying pose, expression, accessory, and lighting (PEAL). The CAS-PEAL-R1, a subset of the CAS-PEAL face database, has been released for the purpose of research, which contains 9060 images of 1040 persons.

The CAS-PEAL database provides several comparison environments. The proposed method LDN gives good results for most of the data sets, and it is close to the best methods. Our proposed methods are outperform by the LDiP, LPQ, and GGPP methods in the Accessory, Age, Background, and Expression data sets by 0.6%, 1.5%, 0.2% and 1.2%, respectively. Note that the Lighting data set in this database is really challenging. In contrast to the Yale B database which contains only images with shadows, the CASPEAL Lighting data set has dark (with shadows) and bright (with flashes) images. This combination of images makes this data set more challenging. As we see in the results, the methods can detect less than half of the data set. Nevertheless, the proposed method still outperforms other methods that use intensity, and even outperforms the GGPP and LPQ methods that use phase information as feature. However, the use of phase information makes GGPP to produce really good results.

Furthermore, note that there are other versions of the algorithm proposed by Zhang *et al.* (HGPP and its weighted variations) from which the authors reported better results (produced in their environment) than those presented here—HGPP has an average accuracy of 87.57%, and 91.01% in FERET and CAS-PEAL databases, respectively. Moreover, the difference in the detection accuracy that we presented and the one shown by Zhang *et al.*, may due to the face detector we used our deliberated lack of pre-processing and the parameters used to re-size and divide the face. The recognition conducted using SVM with different kernels to classify the facial expressions. For this task it is analyzed the four variations of LDN that we presented before: LDN*K* and the three variations of LDN*G*.

6) (Extended) Cohn-Kanade Results: The Cohn-Kanade Facial Expression (CK) database consists of 100 university students. Subjects were instructed to perform a series of 23 facial displays, six of which were based on description of prototype emotions. Image sequences from neutral to target display were digitized. In this setup, it is selected 408 image sequences from 96 subjects, each of which was labeled as one of the six basic emotions. For 6-class prototypical expression recognition, the three most expressive image frames were taken from each sequence that resulted into 1224 expression images. In order to build the neutral expression set, the first frame (neutral expression) from all 408 sequences was selected to make the 7-class expression data set (1632 images). Furthermore, we used the extended Cohn-Kanade database (CK+), which includes 593 sequences for seven basic expressions (happiness, sadness, surprise, anger, disgust, fear, and contempt). In our experiments, we selected the most expressive image frame from 325 sequences from 118 subjects from the database for evaluation.

The best recognition rates of the proposed methods, in comparison with other methods, in which the LDN*G* codes perform better in the 6- and 7-class problem on the CK database. To obtain a better picture of the recognition accuracy of individual expression types, we present the confusion matrices for 6- and 7-class expression recognition using the CK database for the best LDN codes. However, as we include the neutral expression in the 7-class recognition problem, the accuracy of other five expressions decreases because some facial expression samples are confused with a neutral expression.

V. CONCLUSION

This paper introduce a novel encoding scheme, LDN, that takes advantage of the structure of the face's textures and that encodes it efficiently into a compact code. LDN uses directional information that is more stable against noise than intensity, to code the different patterns from the face's textures. Additionally, we analyzed the use of two different compass masks (a derivative-Gaussian and Kirsch) to extract this directional information, and their performance on different applications. In general, LDN, implicitly, uses the sign information of the directional numbers which allows it to distinguish similar texture's structures with different intensity transitions—*e.g.*, from dark to bright and vice versa.

Also LDN evaluated under expression, time lapse and illumination variations, and found that it is reliable and robust throughout all these conditions, unlike other methods. For example, Gradient faces had excellent results under illumination variation but failed with expression and time lapse variation. Also, LBP and LDiP recognition rate deteriorates faster than LDN in presence of noise and illumination changes.

VI. REFERENCES

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, Dec. 2003.

[2] S. Z. Li, A. K. Jain, Y. L. Tian, T. Kanade, and J. F. Cohn, "Facial expression analysis," in *Handbook of Face Recognition*. New York: Springer-Verlag, 2005, pp. 247–275.

[3] H. Hong, H. Neven, and C. von der Malsburg, "Online facial expression recognition based on personalized galleries," in *Proc. 3rd IEEE Int. Conf. Automatic Face Gesture Recognition.*, Apr. 1998, pp. 354–359.

[4] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 172–187, Jan. 2007.

[5] N. G. Bourbakis and P. Kakumanu, "Skin-based face detection-extraction and recognition of facial expressions," in *Applied Pattern Recognition*. New York: Springer-Verlag, 2008, pp. 3–27.

[6] N. Bourbakis, A. Esposito, and D. Kavraki, "Extracting and associating meta-features for understanding people's emotional behaviour: Face and speech," *Cognit. Comput.*, vol. 3, no. 3, pp. 436–448, 2011.

[7] P. Kakumanu and N. Bourbakis, "A local-global graph approach for facial expression recognition," in *Proc. 18th IEEE Int. Conf. Tools Artif. Intell.*, Nov. 2006, pp. 685–692.

[8] A. Cheddad, D. Mohamad, and A. A. Manaf, "Exploiting voronoi diagram

properties in face segmentation and feature extraction," *Pattern Recognit.*, vol. 41, no. 12, pp. 3842–3859, 2008.

[9] X. Xie and K.-M. Lam, "Facial expression recognition based on shape and texture," *Pattern Recognit.*, vol. 42, no. 5, pp. 1003–1011, 2009.

[10] P. Ekman, *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*. New York: Times Books, 2003.

[11] M. Kabir, T. Jabid, and O. Chae, "A local directional pattern variance (LDPv) based face descriptor for human facial expression recognition,"
in *Proc. 7th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Sep. 2010, pp. 526–532.

[12] 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 Trans. Image Process.*, vol. 19, no. 2, pp. 533–544, Feb. 2010.

[13] B. Zhang, L. Zhang, D. Zhang, and L. Shen, "Directional binary code with application to PolyU near-infrared face database," *Pattern Recognit. Lett.*, vol. 31, no. 14, pp. 2337–2344, 2010.

[14] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang, and X. Liu, "Face recognition under varying illumination using gradientfaces," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2599–2606, Nov. 2009.

[15] Z. Xie and G. Liu, "Weighted local binary pattern infrared face recognition based on Weber's law," in *Proc. 6th Int. Conf. Image Graph.*, Aug. 2011, pp. 429–433.