Stratified Multiscale Local Binary Patterns: Application to Face Recognition

Udhayakumar.M, Gayathri.S, Mary Jeya Priya.R, Arunkumar.M
PG Scholars of Dept of ECE[123], Faculty of Dept of ECE[4]
Bannari Amman Institute of Technology, Sathyamangalam, Erode.

Abstract: - This paper presents a efficient facial image recognition based on multi scale local binary pattern (LBP) texture features .It's a fast and simple for implementation, has shown its superiority in face recognition. To extract representative features, "uniform" LBP was proposed and its effectiveness has been validated. However, all "non-uniform" patterns are clustered into one pattern, so a lot of useful information is lost. In this study, propose to build a hybrid multiscale LBP histogram for an image. The face image is divided into several regions from which the LBP feature distributions are extracted and concatenated into an enhanced feature vector to be used as a face descriptor. The useful information of "non-uniform" patterns at large scale is dug out from its counterpart of small scale, The performance of the proposed method is that it can fully utilize LBP information while it does not need any training step, which may be sensitive to training samples assessed in the face recognition problem under different challenges. Other applications and several extensions are also discussed.

Index Terms: - *Facial image representation, local binary pattern, multiscale, component based face recognition, texture features, face misalignment*

I. INTRODUCTION

THE face has its own advantages over other biometrics for people identification and verification related applications, since it is natural, nonintrusive, contactless etc. Unfortunately, all human faces are similar to each other in their configurations and hence offer low distinctiveness, unlike other biometrics, e.g. the iris and finger print, palm print. [1]. Furthermore, infraclass variations due to factors as diverse as pose variation and facial expression etc. They can get high recognition rate and are user friendly.

How to extract discriminate information from an image is one of the key components for biometrics system. There are many different algorithms proposed in the past, such as principal component analysis (PCA) [2], Gabor phase encoding [3], local ternary pattern(LTP) and local binary pattern (LBP) [4-9] for feature extraction. Among them, LBP based method has shown its superiority in face recognition. LBP was originally proposed as a texture descriptor. It owns many advantages, such as it is simple to implement and fast to compute.

In that texture descriptor has been validated that "uniform" patterns play an important role in texture classification [10]. "Uniform" patterns also showed its superiority in face recognition [4-5, 8-9]. Incorporating "uniform" idea, many patterns, which are not "uniform" patterns, are clustered into one "non-uniform" pattern. By this way, many discriminate but "non-uniform" patterns fail to provide useful features. And, the percentage of "non-uniform" patterns increases as the radius increases, so much information is lost. Recently, some works were proposed to address this issue. Many "non-uniform" patterns are isolated from the "non-uniform" cluster [6-7]. However, such methods are learning based algorithms, which require some training samples to discover useful "non-uniform" patterns. Thus, the recognition performance may be related with the training samples.

In this paper, we propose a hybrid multiscale LBP algorithm for face recognition. The LBPs for biggest radius is firstly extracted. Then, for those "non- uniform" patterns, the counterpart LBPs of smaller radius is extracted. Among the new LBPs, those "non-uniform" patterns is further proceeded to extract "uniform" patterns in even smaller radius. The procedure is iterated until the smallest radius is proceeded. The proposed scheme could fully utilize the information of "non-uniform" LBPs of bigger radius. Furthermore, this hybrid scheme is totally training free which are not sensitive to the training samples.

The rest of the paper is organized as follows. Section 2 reviews the PCA, LDA, LBP. Section 3 presents the proposed hybrid multiscale LBP method. Section 4 gives the conclusion and future work

II. FACE DESCRIPTION

2.1 Principal Component Analysis (PCA):

PCA is a statistical dimensionality reduction method, which produces the optimal linear least-square decomposition of a training set. Kirby and Sirovich [3] applied PCA for representing faces and Turk and

Pentland [7] extended PCA for recognizing faces. PCA is a useful statistical technique that has found application in fields such as face recognition and image compression, and is a common technique for finding patterns in data of high dimension. PCA commonly referred to as the use of Eigen faces [3]. With PCA, the probe and gallery images must be the same size and must first be normalized to line up the eyes and mouth of the subjects within the images. The PCA approach is then applied to reduce the dimension of the data by means of data compression, and reveals the most effective low dimensional structure of facial

Fig 1: Eigen faces

The feature vectors are derived using Eigen faces. shows a typical example of PCA. It can be easily spotted that the majority of 2D points are located closer to the 1st PC line, which means that the projection can be performed with these points on the 1st PC line without losing essential 2D information.

PCA Method uses Eigenvectors and Eigenvalues for representing face images. These Eigenvectors can be thought of as a set of features which together characterize the variation between face images [7]. Each image location contributes more or less to each Eigenvector, so that we can display the Eigenvector as a sort of ghostly face which is called as an Eigen face. face space, and then perform patterns. This reduction in dimensions removes information that is not useful and precisely decomposes the face structure into orthogonal (uncorrelated) components known as Eigen faces [6]. Each face image may be represented as a weighted sum (feature vector) of the Eigen faces, which are stored in a 1D array A probe image is compared against a gallery image by measuring the distance between their respective feature vectors. The PCA approach typically requires the full frontal face to be presented each time; otherwise the image results in poor performance in shown fig 2, The primary advantage of this technique is that it can reduce the data needed to identify the individual to 1/1000th of the data presented

2.2 Linear Discriminant Analysis:

Linear Discriminant Analysis (LDA) approach outperforms the Principal Com- ponent Analysis (PCA) approach in face recognition tasks. Due to the high dimensionality of a image space, many LDA based approaches, however, first use the PCA to project an image into a lower dimensional space or so-called the LDA to maximize the discriminatory power. LDA can be used not only for classification, but also for dimensionality reduction. For example, the LDA has been widely used for dimensionality reduction in speech recognition. LDA algorithm offers many advantages in other pattern recognition tasks, and we would like to make use of these features with respect to face recognition as well. The basic idea of LDA is to find a linear transformation such that feature clusters are most separable after the transformation.

LDA is a data separation technique. The objective of LDA is to find the directions that can well separate the different classes of the data once projected upon. The set of human faces is represented as a data matrix X, where each row corresponds to a different human face. Each image X, represented by a (n, m) matrix of pixels, is represented by a high dimensional vector of n*m pixels. Turk and Pentland [8] were among the first who used this representation for face recognition. 2-dimensional principal component analysis (2dPCA) [9] was proposed which directly computes the Eigenvectors of the covariance matrix without matrix to vector conversion.2-dimensional LDA [10,11] computes directly the directions which will separate the classes without matrix to vector conversion as well. Higher recognition rate was reported for both cases. Both of these algorithms work in batch mode. The PCA/LDA-based face recognition systems suffer from the scalability problem.

2.3 Local binary patterns :
The original LBP operator labels the pixels of an image by thresholding a 3×3 neighborhood of each pixel with the center value and considering the results as a binary code. The LBP code of the center pixel in the neighborhood is obtained by converting the binary code into a decimal one. Figure 1 gives an illustration for the basic LBP operator. Based on the operator, each pixel of an image is labeled with an LBP code. The 256-bin histogram of the labels contains the density of each label and can be used as a texture descriptor of the considered region. The procedure of extracting LBP features for facial representations is implemented as follows: First, a face image is divided into several non overlapping blocks. Second, LBP histograms are

computed for each block. Finally, the block LBP histograms are concatenated into a single vector. As a result, the face image is represented by the LBP code.
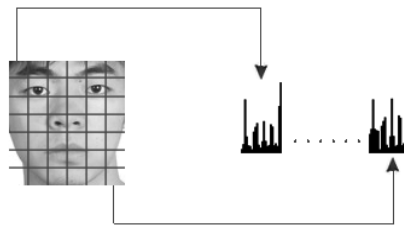

Fig.2.Basic LBP Features.

Local binary patterns The original LBP operator [10] labels the pixels of an image by thresholding a 3 × 3 neighborhood of each pixel with the center value and considering the results as a binary code. The LBP code of the center pixel in the neighborhood is obtained by

operator. converting the binary code into a decimal one. Figure 1 gives an illustration for the basic LBP Based on the operator, each pixel of an image is labeled with an LBP code. The 256-bin histogram of the labels contains the density of each label and can be used as a texture descriptor of the considered region. The procedure of extracting LBP features for facial LBP approach can obtain the relationship among pixels of a facial image in a larger scale, which can contain more face features with the cost of increasing data redundancy. Binary pattern: 11010101
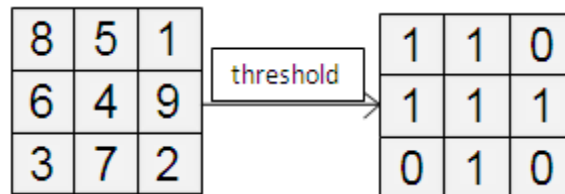

Fig.3.Fundamental LBP operator.

$$LBP_{P,R} = \sum_{p=1}^{p} s(g_p - g_c)2^{p-1} \text{ --------------- (1)}$$

$$s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$$

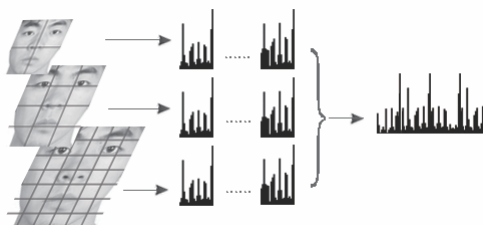**2.4 Multi-Resolution with Multi-Scale LBP:**

Fig.4.basic MLBP Features.

Most LBP operators describe the texture distribution of each pixel with its neighborhood only within the original face image. However, the differences between two faces can be demonstrated not only by the texture distribution of each pixel with its neighborhood, but also by the relative connection with further pixels,

III.   PROPOSED STRATIFIED MULTISCALE LBP

The performance of single LBP operator is limited. Multi scale or multi resolution could represent more image feature under different settings. Traditionally, LBP features of different scale are extracted first, and then the histograms are concatenated into a long feature. Joint distribution could contain more information, but

it suffers from huge feature dimension. As shown in Section 2, (2P -P*(P-1)-2) "non-uniform" patterns are clustered into one "non-uniform" pattern. By applying this scheme, much information is lost. And, as the radius increases, the percentage of "non-uniform" pattern increases. For example, Table I shows the percentage of "non-uniform" patterns in images.
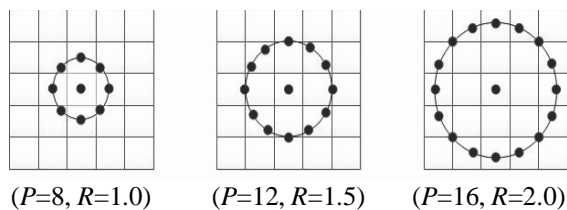

(*P*=8, *R*=1.0)   (*P*=12, *R*=1.5)   (*P*=16, *R*=2.0)

Fig.5: Example of proposed stratified multiscale LBP

	R=1	*R*=2	*R*=3
P=8	15.82	23.68	29.86

Table I: Percentage(%) of "non-uniform" patterns in AR database [11].

be estimated by interpolation where gc is the gray value of the central pixel, gp is the value of its neighbors, P is the total number of involved neighbors and R is the radius of the neighborhood. Suppose the coordinate of gc is (0, 0), then the coordinates of gp are (R*cos(2πp/P),R*sin(2πp/P)). Fig. 1 gives examples of circularly symmetric neighbor sets for different configurations of (P,R). The gray values of neighbours that are not in the center of grids can be estimated by interpolation.

As shown in Table I, around one third information is wasted by using previous method. To extract more useful feature from the image, some works were proposed to dig out information from these "non-uniform" patterns [6-7]. However, such methods require a training step to learn which patterns are useful. The recognition accuracy may be dependent on the training samples. Fig. 2 shows an example. The pattern of a bigger radius is "non-uniform", but its counterpart in a smaller radius is "uniform". Thus, it is possible to classify the "non-uniform" patterns according to their counterpart of smaller radius.
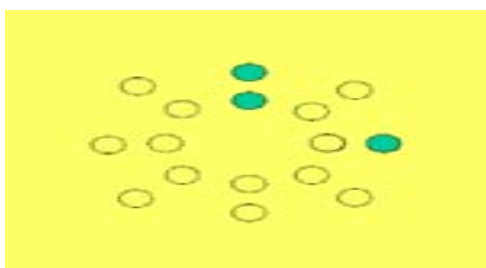

Fig.6.: Binary Patterns of different radius of a local region. Solid circles represent 1 while hollow circles mean 0.
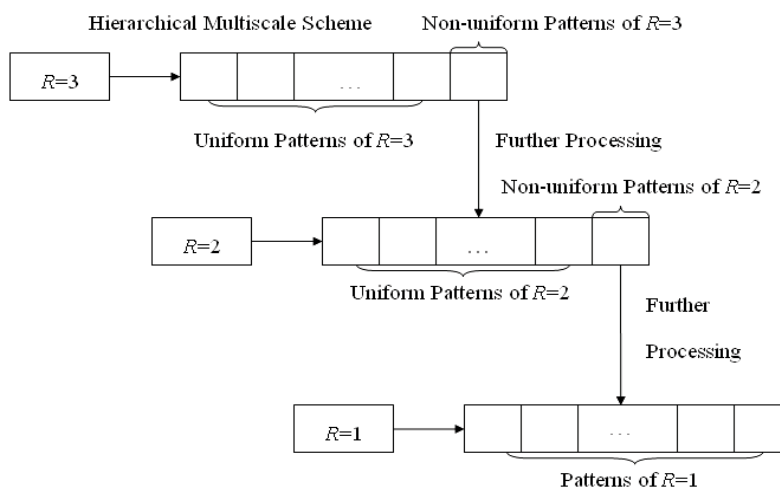

Fig.6.Multi Scale LBP

Suppose the texture image is of size N×M. After identifying the LBP pattern of each pixel (i, j), a histogram is built to represent the whole texture image:

$$H(k) = \sum_{i=1}^{N} \sum_{j=1}^{M} f(LBP_{P,R}(i,j),k), k \in [0,k]$$

$$f(x,y) = \begin{cases} 1, x=y \\ 0, otherwise \end{cases}$$

where K is the maximal LBP pattern value. The U value of an LBP pattern is defined as the number of spatial transitions (bitwise 0/1 changes) in that pattern

$$U(LBP_{P,R}) = \left| s\langle g_{p-1} - g_c \rangle - s\langle g_0 - g_c \rangle \right| +$$

$$\sum_{p=1}^{p-1} \left| s\langle g_p - g_c \rangle - s\langle g_{p-1} - g_c \rangle \right|$$

For example, the LBP pattern 00000000 has a U value of 0 and 01000000 has a U value of 2. The uniform LBP patterns refer to the patterns which have limited transition or discontinuities (U≤2) in the circular binary presentation [10]. It was verified that only those "uniform" patterns are fundamental patterns of local image texture [10]. In practice, the mapping from $LBP_{P,R}$ to $LBP_{P,R}^{u2}$ (superscript "u2" means that the uniform patterns have a U value of at most 2), which has P*(P-1)+3 distinct output values, is implemented with a lookup table of 2^P elements. The dissimilarity of sample and model histograms is a test of goodness-of-fit, which could be measured with a nonparametric statistic test. In this study, the dissimilarity between a test sample S and a class model T is measured by the chi-square distance:

$$D(S,T) = \sum_{n=1}^{N} \frac{(S_n - T_n)^2}{(S_n + T_n)}$$

where N is the number of bins, S_n and T_n are the values of the sample and model images at the nth bin.

Here we propose to build multi scale LBP histogram from big radius to small radius, then that LBP map for each pixel in biggest radius, That pixels are divided into two types of patterns like Uniform pattern and Non-Uniform pattern ,A Sub histogram is build for uniform pattern. Those pixels in Non-uniform pattern are further processed at another small radius, When the new pattern contains full of uniform that process will be stop, and remaining pixels are continued to extract LBP pattern at smaller radius.

Fig. 3 shows an example of the proposed hierarchical multi scale LBP scheme. The LBP histogram for *R*=3 is first built. For those "non-uniform" patterns by *R*=3 operator, a new histogram is built by *R*=2 operator. Then, the "non- uniform" patterns of *R*=2 are further proceeded to build a histogram R=1operator .Finally ,three histograms are concatenated into one multi scale histogram.

There are mainly two differences compared with traditional multi scale LBP. Suppose the number of scale is *S*, the dimension of the proposed scheme is smaller than traditionally scheme by *S*-1.

Second, sum of frequencies of the proposed histogram is 1/*S* of traditional one.

IV.   RESULTS AND DISCUSSION

4.1 Face database:

AR database was used to analysis the performance of proposed method ,That contain the images of 120 individual were taken two different sessions, and each session contain 13 images, facial images are normally cropped into size of 100*80 and that are further reduced to 50*40 for PCA and LDA to speed up the process. Here the image is divided into 5*3window because it to get good balance between feature size and recognition accuracy

The proposed scheme could not get better result than LDA in the first experiment, but, it can get around 1% improvement in the second experiment. This is mainly because LDA is a training based method in feature extraction, thus it requires many training samples to learn representative features. It can get good results when the training sample is enough, however, it may fail to get good enough features when fewer training samples are provided. This is why LDA get around 4% drop from 4 training samples per subject to 3 samples per subject. As a training free method in feature extraction, the proposed method is more robust to the variation of training sample. As shown in Table III, it only gets around 2% drop from the first experiment to the second experiment. It is a very good property in many applications, as getting enough training samples is not a trivial issue.

Table II. Recognition accuracy (%) of different methods.

Method	Accuracy (1st Experiment)	Accuracy (2nd Experiment)
PCA	68.89	79.79
LDA	99.72	95.83
Single scale LBP (R=2, P=8)	87.5	81.46
Traditional multiscale LBP (R={2,3,4}, P=8)	97.78	94.79
Proposed multiscale LBP (R={2,3,4}, P=8)	98.89	96.67

Similar findings as Table I could be found from Table II, such as both multiscale schemes could significantly improve the recognition accuracy and the proposed multiscale scheme is better than traditional one.

However, there is one disadvantage of the proposed scheme. The feature dimension of multiscale LBP is a little higher. For example, in AR database, the feature sizes of the proposed scheme and traditional multiscale are 2,625 ((3*59-2)*5*3) and 2,655 (3*59*5*3), respectively. It is much larger than PCA (418 for the first experiment and 100 for the second experiment) and LDA (96 for the first experiment and 72 for the second experiment) [11]. Fortunately, 2,622 is not a big issue for current PC.

V. CONCLUSION

In this paper, to fully extract useful feature from an image, a stratified mutliscale LBP is proposed. It could dig out useful information from those "non-uniform" patterns. The main advantage of the proposed method could maintain the training free property during feature extraction, which is very important for some applications. Its effectiveness is shown in one palmprint and one face database. Compared with traditionally multiscale LBP, the proposed method could get more than 1% improvement. It could also get better result than those training based methods, when the training samples are not enough.
The feature size of multiscale LBP is a little high.How to reduce the feature size but get good performance in recognition will be our future work

REFERENCES

[1] Jain, A., Bolle, R., and Pankanti, S., (eds), *Biometrics: Personal Identification in Network Society*, Boston, Kluwer Academic, 1999.

[2] M. Turk, and A. Pentland, "Face recognition using eigenfaces,"*IEEE Conference on Computer Vision and Pattern Recognition*, pp.586-591, 1991.

[3] D. Zhang, A. Kong, J. You and M. Wong, "Online palmprint identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp.1041-1050, 2003.

[4] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.28, no. 12, pp.2037-2041, 2006.

[5] G. Zhang, Z. Huang, and S. Z. Li, "Boosting local binary pattern (LBP)-based face recognition," SinoBiometrics, pp. 179-186, 2004.

[6] Y. Raja, and S. Gong, "Sparse multiscale local binary patterns,"*Proceeding of 17th British Machine Vision Conference*, 2006.

[7] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li, "Learning multi-scale block local binary patterns for face recognition," *International Conference on Biometrics*, pp. 828-837, 2007.

[8] G. K. O. Michael, T. Connie, and A. B. J. Teoh, "Touch-less palm print biometrics: novel design and implementation," *Image and Vision Computing*, vol. 26, no. 1, pp. 1551-1560, 2008.

[9] T. Ojala, M. Pietikäinen, and T. Mäenpää, "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] L. Zhang, Q. Gao, and D. Zhang, "Directional independent component analysis with tensor representation," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-7, 2008.

[11] R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates", IEEE Trans.on PAMI,vol.15no10, pp.1042-1052, 1993.

[12] R. Brunelli, "Template Matching Techniques in Computer Vision: Theory and Practice", Wiley publications, ISBN 978-0-470-51706-2, 2009.