A NEW PROBABILISTIC LOCAL BINARY PATTERN FOR FACE VERIFICATION

Nutao Tan[1,2], *Lei Huang*[1] , *Changping Liu*[1]
[1]Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[2]Graduate University of Chinese Academy of Sciences, Beijing 100190, China
{nutao.tan, lei.huang, changping.liu}@ia.ac.cn

ABSTRACT

This paper presents a new Probabilistic Local Binary Pattern (PLBP), an extension of existing Local Binary Pattern (LBP), for face verification. Unlike LBP employing the sign of the difference to express the result of comparing two pixels, PLBP employs probability to express it. The advantage is that it can encode the magnitude of the difference, which is useful for face verification but is ignored by the original LBP. Besides, PLBP is combined with Linear Discriminant Analysis (LDA) to extract its most discriminant information. The experimental results show that PLBP can achieve a better performance and is more robust to noise than the original LBP when combined with LDA.

Index Terms— Face Verification, Local Binary Pattern, LBP, Probabilistic Local Binary Pattern, LDA

1. INTRODUCTION

Automatic human face recognition has gained a lot of attention. It has many practical applications, such as bankcard identification, access control, mug shots searching, security monitoring, and surveillance systems.

Many methods for face recognition have been proposed within the last two decades [1]. Among these methods, Linear transform based methods such as Eigenface (PCA)[2], Fisherface(LDA)[3], and Independent Component Analysis (ICA) [4] have had a significant influence within the face recognition community for a considerable time. The LDA, also called Fisher linear discriminant(FLD), defines a projection that makes the within-class scatter small and the between-scatter large. However, it is well-known that classical LDA requires that one of the scatter matrices is nonsingular. For the face recognition application, all scatter matrices can be singular since the feature dimension, in general, exceeds the number of sample size. This is known as the singularity, or Small Sample Size(SSS) problem. Many LDA extensions were proposed to overcome the singularity problem in face recognition, such as PCA+LDA [3], Regularized LDA [5], Direct LDA[6]. Besides, in recent years, tensor learning [7-9] has gained development, because tensor not only can preserve the spatial structure of image and lessen the arguments to learn, but also just need a few samples for training. Gang Hua[9] has proposed a face recognition algorithm using discriminatively trained orthogonal rank one tensor projections and it performs better than LDA. However, he didn't suggest a good method for initializing the subspaces.

In recent years, local matching approaches show inspiring results. The most representative is Local Binary Pattern (LBP)[10-11]. It is a significant breakthrough for face representation, outperforming earlier methods such as PCA, LDA and EBGM [12]. LBP was first proposed for texture classification and later applied to face recognition. Ahonen[11] proposed to divide the image into many non-overlapped blocks, then extract LBP histograms from each block and concatenate them into an enhanced one. This algorithm has gotten good performance on FERET database.

Although the original LBP is a good texture descriptor for face verification, it has some disadvantages. Firstly, it is sensitive to noise. When noises exist, a larger pixel may be changed to a smaller one, especially for those whose values are very near to each other. Then the LBP code will also be changed. Secondly, the original LBP only considers the sign of the difference between two values, but doesn't consider the magnitude of the difference, which may be very useful.

Although an extended LBP[13] has been proposed to make full use of the magnitude of the difference, it is only fit to infrared images which are not sensitive to illumination, but isn't fit to general images which are often covered with illumination. In this paper, we propose a new LBP operator, which adopts the probability to express the result of comparison to overcome those drawbacks. We call it probability-LBP, denoted as PLBP. It uses the difference as an argument of the probability. For this, it not only can encode the magnitude of the difference, but also is robust to the noise. Besides, like the original LBP, PLBP has some redundant information which is harmful to verification. So, we combine it with LDA to extract the discriminant information. The experiment results show that our proposed method is effective.

The remainder of the paper is organized as follows: Sec. 2 introduces the basic Local binary pattern. Sec.3 introduces the probabilistic LBP. Sec.4 presents extensive experimental results and discussions. Finally we conclude in Sec.5.

2. BASIC LOCAL BINARY PATTERN OPERATOR

The original LBP operator, introduced by Ojala [11], is a powerful method of texture description. The operator labels the pixels of an image by thresholding the 3×3 neighborhood of each pixel with the center value and considering the result as a binary number. Then the histogram of the labels can be used as a texture descriptor. An illustration of the basic LBP operator is shown in Fig.1a. The decimal form of the resulting 8-bit word (LBP code) can be expressed as following:

$$\mathrm{LBP}_{\mathrm{P,R}}(x_c, y_c) = \sum_{p=0}^{\mathrm{P}-1} s(g_p - g_c)2^p \tag{1}$$

Where g_c corresponds to the gray value of the center pixel (x_c, y_c), g_p to the gray values of the 8 surrounding pixels, and function $s(x)$ is defined as:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{2}$$

Later the operator was extended to use neighborhoods of different sizes. Using circular neighborhoods and bilinearly interpolating the pixel values allow any radius and number of pixels in the neighborhood. Let $\mathrm{LBP}_{\mathrm{P,R}}$ denotes p equally spaced pixels on a circle of radius R. Fig.1b shows an $\mathrm{LBP}_{8,2}$ operator. Another extension to the original operator uses so called uniform patterns (ULBP). A Local Binary Pattern is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular. For example, 00000000, 00011110 and 10000011 are uniform patterns. Uniform patterns mainly depict the local micropatterns, such as edges, spots and flat areas, over the whole image. Ojala et al. noticed that in their experiments with texture images, uniform patterns account for a bit less than 90% of all patterns when using $\mathrm{LBP}_{8,1}$. Another contribution of ULBP is that it can reduce the dimension of $\mathrm{LBP}_{8,1}$ from 256 to 59, which is very important to face recognition. Some examples of ULBP are shown in Fig.2.

Recently, Ahonen etc.[11] have proposed a face recognition system based on LBP descriptor. They first divided the face image into R non-overlap regions, then calculated the LBP histograms $\{H^r \mid r \in (0,....R-1)\}$ from each region and concatenated them into a single spatially enhanced feature histogram efficiently representing the face image. The Histogram of a labeled image $f(x, y)$ can be defined as

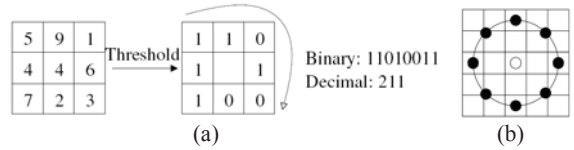


Fig. 1. (a)The basic LBP operator (b) The circular LBP operator.

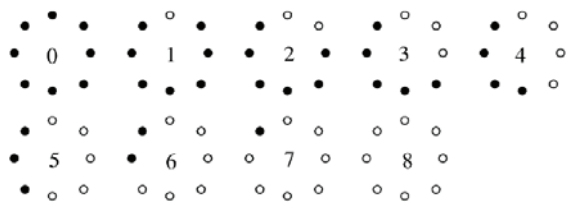


Fig. 2. Samples of ULBP

$$H_i^r = \sum_{x,y \in block_r} I(f(x,y) = i), \quad i = 1,...,L \tag{3}$$

Where L is the number of different labels produced by the LBP operator and

$$I\{A\} = \begin{cases} 1 & A \text{ is true} \\ 0 & A \text{ is false} \end{cases} \tag{4}$$

Then, the concatenated histogram is defined as

$$H = [H^0 \; H^1 \; ... \; H^{R-1}] \tag{5}$$

It is used to face recognition or verification.

3. PROBABILISTIC LOCAL BINARY PATTERN

Local Binary Pattern is a very useful texture descriptor, but it only employs the sign of the difference when compares the center pixel with its neighbors, and doesn't make full use of the magnitude of the difference. So, it is sensitive to noise. For example, as shown in Fig.3, *a*, *b* and *c* are different just in the right-top position. Apparently, *a* is more similar to *b* than to *c*, but from last section, we know that the LBP code of *a* is the same as *c* and different to *b*, which doesn't coincide with the truth. In this case the original LBP will result in some error. So, we can see that LBP sometimes can't capture the local detail of an image completely. Accordingly, in order to make LBP robust to noise, we propose probabilistic-LBP, denoted as PLBP, which expresses the result of the comparison with probability rather than binary mode.

As known that the center may be greater or smaller than its neighbors, so we use two different probabilities to describe these cases. One describes that the center is greater than the neighbor, denoted as p_g, and the other describes that the center is smaller than the neighbor, denoted as p_s. The sum of p_g and p_s is 1. Here, we suppose that the probability density can be molded as Gaussian, then the probabilities between the center pixel U and its neighbor V can be denoted as:

$$p_g(u,v)=\begin{cases}1.0-0.5*e^{-\frac{x^2}{\sigma^2}} & if\ \ I_u \ge I_v \\ 0.5*e^{-\frac{x^2}{\sigma^2}} & if\ \ I_u < I_v\end{cases} \tag{6}$$

$$p_s(u,v)=\begin{cases}0.5*e^{-\frac{x^2}{\sigma^2}} & if\ \ I_u \ge I_v \\ 1.0-0.5*e^{-\frac{x^2}{\sigma^2}} & if\ \ I_u < I_v\end{cases} \tag{7}$$

Where σ is the standard variance, x is defined as

$$x=\frac{|I_u-I_v|}{I_u} \tag{8}$$

I_u, I_v denote the values of U and V in the gray image. It is obvious that x can encode the magnitude of the difference.

From the formula (6), we can see that if U is far greater than V, then p_g is near to 1, if U is far smaller than V, then p_g is near to 0, and if U is approximately equal to V, then p_g is near to 0.5. Now, let us review the Fig.3, apparently, the probabilities that the center is greater than the right-top position are almost the same in *a* and *b*, but different to *c*, which coincides with the truth. So adopting probability for LBP can not only encode the magnitude of difference, but also improve its robustness to noise.

As we know that only one code corresponds to each pixel for original LBP, so when calculating the histogram, each pixel need only add one to the bin that corresponds to its code. This method doesn't work for PLBP, because PLBP allow each possible code be occurred at a certain probability. For example, there are 256 possible codes for each pixel if use PLBP(8,2) as the texture descriptor. The probability for each code is equal to the product of all of the neighbor's probability, and can be denoted as

$$P_c(u)=\prod_{i=1}^{8} p_i \tag{9}$$

Where p_i is defined as

$$p_i=\begin{cases}p_g(u,x_i) & if\ c_i=0 \\ p_s(u,x_i) & if\ c_i=1\end{cases} \tag{10}$$

u is the center pixel, x_i is the *i-th* neighbor, c is the LBP code, c_i denotes the value of position i in c, it values 0 (means the center is greater than the neighbor) or 1 (means the center is smaller than the neighbor). Therefore, the histogram of PLBP can be denoted as

$$H_c^r=\sum_{u\in block_r} P_c(u), \quad c=1,...,L \tag{11}$$

Apparently, if the probability is only valued 0 or 1, then PLBP histogram is the same as LBP histogram. So, LBP is just a special case of PLBP. Like ULBP derived from LBP, we can also derived PULBP from PLBP.

5	101	99
12	100	102
97	173	243

(a)

5	101	101
12	100	102
97	173	243

(b)

5	101	3
12	100	102
97	173	243

(c)

Fig. 3. Local Binary Pattern samples

Table 1
Verification performance on the PIE database

method	0.01%	0.1%	1%	ERR
LDA	3.65	8.44	21.71	20.10
ULBP+LDA	35.84	58.88	80.99	6.39
PULBP+LDA	54.57	73.91	89.01	4.49

Table 2
Verification performance on the AR database

method	0.01%	0.1%	1%	ERR
LDA	9.79	19.52	38.15	17.97
ULBP+LDA	28.57	56.46	82.70	6.08
PULBP+LDA	44.48	69.24	86.58	4.89

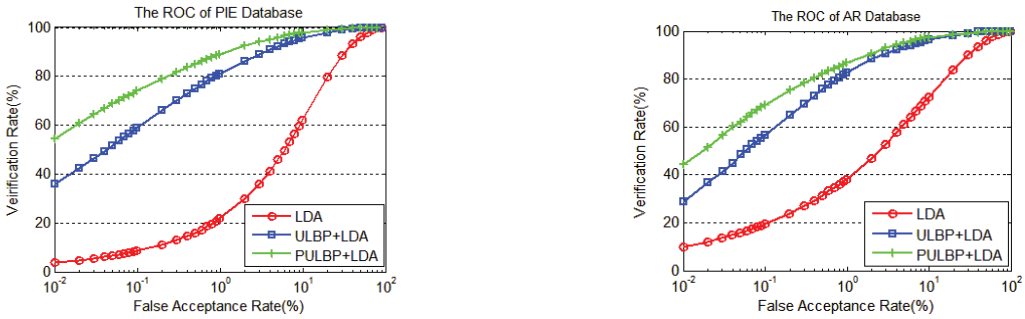
4. EXPERIMENTAL RESULTS

In this section, we will evaluate the performances of the proposed algorithm for face verification based on different face databases. The databases used include the PIE database and the AR database.

The PIE dataset contains 41368 images of 68 people (13 poses, 43 illumination conditions and 4 expressions). We used the images of the nearly frontal poses under all illumination conditions and expressions. Finally, a subset of 10667 face images with 150 to 167 images per person was used. We randomly select 680 images that include 10 images per person for training, and the rest for testing.

The Purdue AR dataset contains 3247 face images of 126 persons (70 man and 56 women) under different expression (nature, smile, angry, outcry), illumination (left, right, and two sides), wears (sun glass, scarf). The images were taken under the same light and pose by the same camera in two sessions. In the experiment, we selected 117 persons with 14 images that without wears per person. The images from session 1 are used for training, and those from session 2 are used for testing.

On all datasets, the gray-scale face images are cropped and aligned by fixing the eye locations, and then resized to 64×64. No other pre-processing is performed. Eye positions are marked manually. Each image is divided into 8×8 regions and Euclidean distance is used for verification. To get the best performances, the argument σ is set to 0.025 for PIE database and 0.0125 for AR database.



(a)ROC curve on PIE database (b) ROC curve on AR database

Fig. 4. The ROC curve on PIE and AR database

In order to compare the proposed method with other algorithms, three different methods are given: (1) LDA; (2) ULBP histogram based on LDA, denoted as ULBP+LDA; (3) PULBP histogram based on LDA, denoted as PULBP+LDA. The experimental results are shown in three forms: ROC curve, the verification rate (VR) under different false acceptance rates (FAR), such as 0.01%, 0.1%, 1%, and the Equal Error Rate (EER). For PIE database, the argument σ is set to 0.025 and it is set to 0.0125 for AR database.

Fig.4 shows the ROC curves on PIE and AR database. The X-axis is in log scale. It can be seen that under the whole false acceptance rate, PULBP+LDA can get the highest verification rate on both databases, followed by ULBP+LDA. LDA performs the poorest. This indicts that PLBP is more robust to noise than LBP.

The verification rate under different false acceptance rate and the EER value of PIE and AR database are shown in Table.1 and Table.2 respectively. From Table.1, we can see that PULBP+LDA can get as high as 54.57% VR under 0.01% FAR on the PIE database, which is higher than ULBP+LDA and LDA by 18.73% and 50.92% respectively. It is a great improvement. Besides, the EER value of PULBP+LDA is 4.49%, which is also lower than ULBP +LDA(6.08%) and LDA (20.1%). The verification performance on AR database is shown in Table.2. It also shows that PULBP+LDA performs the best. The verification rate under 0.01% FAR is reached to 44.48% for PULBP+LDA, while ULBP+LDA is only at 28.57% and LDA is only at 9.79%. Compared to ULBP+LDA(6.08%) and LDA (17.97%), the EER value of PULBP+LDA(4.89%) is also the lowest.

5. CONCLUSION

In this paper we have presented a new Local Binary Pattern which we refer to as PLBP. PLBP is able to encode the magnitude of the difference between two compared pixels, which is ignored by the original LBP, by employing probability. The experimental results on PIE and AR database show that PLBP can get higher verification rate and lower ERR value than the original LBP for face verification when combined with linear discriminant analysis, and it is more robust to noise.

6. REFERENCES

[1] A.F. Abate, M. Nappi, D. Riccio, G. Sabatino, “2D and 3D face recognition: A survey,” *Pattern Recognition Letters*, vol.28, pp.1885-1906, 2007.

[2] M. Turk, A. Pentland, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience*, vol.3, no.1, pp.71-86, 1991.

[3] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, “Eigenfaces vs. Fisherfaces: recognition using class specific linear projection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.19, no.7, pp.711-720, 1997.

[4] M.S. Bartlett, J.R. Movellan, T.J. Sejnowski, “Face recognition by independent component analysis”, *IEEE Transactions on Neural Networks*, vol.13, no.6, pp.1450-1464, 2002.

[5] J. Lu, K. Plataniotis, A.Venetsanopoulos, “Regularization Studies of Linear Discriminant Analysis in Small Sample Size Scenarios with Application to Face Recognition,” *Pattern Recognition Letter*, vol. 26, no. 2, pp.181-191, 2005.

[6] H. Yu, J. Yang, “A Direct LDA Algorithm for High-dimensional Data with Application to Face Recognition,” *Pattern Recognition*, vol.34, pp.2067-2070, 2001.

[7] X. He, D. Cai, P. Niyogi, “Tensor subspace analysis,” *In Advances in Neural Information Processing Systems*, Vancouver, Canada, 2005.18.

[8] D. Xu, S. Lin, S. Yan, X. Tang, “Rank-one projections with adaptive margins for face recognition,” I*n Proc. of IEEE Conf. on Computer Vision and Patter Recognition*, NewYork City, NY, pp.175~181, 2006.

[9] G. Hua, P.A. Viola, S.M. Drucker. “Face Recognition using Discriminatively Trained Orthogonal Rank One Tensor Projections,” *Proceeding of CVPR*, pp.1-8, 2007.

[10] T. Ojala, M. Pietikainen, T. Maenpaa, “Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns,” *PAMI*,vol.24, no.7, pp.971-987, 2002.

[11] T. Ahonen, A. Hadid, M. Pietikainen, “Face Recognition with Local Binary Patterns”, *Proceeding of ECCV*, pp.469-481, 2004.

[12] L. Wiskott, J.M. Fellous, N. Kruger, C. Von der Malsburg, “Face recognition by elastic bunch graph matching,” *PAMI*, vol.19, no.7, pp.775-779, 1997.

[13] D. Huang, Y.H. Wang, Y.D. Wang, “A Robust Method for Near Infrared Face Recognition Based on Extended Local Binary Pattern,” *Advances in Visual Computing*, pp.437-446, 2007.