IWSSIP 2010 - 17th International Conference on Systems, Signals and Image Processing

A Linear Discriminant Analysis approach for Local Binary Pattern Based Face Recognition

Dário Augusto Borges Oliveira, Raphael Pithan, Raul Queiroz Feitosa, Alvaro de Lima Veiga Filho
Department of Electrical Engineering
Pontifical Catholic University of Rio de Janeiro
Rio de Janeiro, Brazil
dario@ele.puc-rio.br, raphaelpithan@yahoo.com.br, raul@ele.puc-rio.br, alvf@ele.puc-rio.br

Antonio Carlos Censi, José Luiz Buonomo de Pinho
Montreal Informática
Rio de Janeiro, Brazil
accensi@montreal.com.br, pinho@montreal.com.br

Abstract— This paper proposes a linear discriminant analysis (LDA) approach for LBP based face recognition. The method estimates the LBP weights directly from the discriminant axis based on the chi square distances between the tiles of a pair of images. It is also able to handle with some characteristics of the face database, such as non-symmetric illumination, since it does not have a symmetry constraint. The proposed method is evaluated by experiments on the FERET face database, and the results are compared with reports of other related works. In these experiments the solution provided by the proposed method brought about the best recognition performance.

Keywords- Face Recognition, Local Binary Pattern, Linear Discriminant Analysis

I. INTRODUCTION

Automatic Face Recognition (AFR) has been a very active research topic for more than 3 decades [1]. In spite of the remarkable progress achieved so far, AFR systems are still distant from the ability of the human vision system to cope with facial image variations in pose, expression, illumination and so on.

To solve this task, new image representations are being proposed which are less sensitive to variations in the facial appearance. A recent and robust algorithm for face recognition, introduced in [2], which is particularly effective in identification of frontal faces in galleries consisting of a single sample per subject, is gaining increasing attention in the AFR research community [3] [4] [5]. It is based on texture descriptors called Local Binary Patterns (LBP). The main advantage of this approach lies in the robustness of LBPs to changes in illumination and in its computational simplicity.

A LBP is a binary code representing the local intensity arrangement within a pixel neighborhood. These codes are computed for each pixel, bringing about a new image representation. The recognition method partitions the face image into a set of non-overlapping blocks and assigns a different weight to each block that expresses the importance of the corresponding face region in the recognition process. Since the type and amount of variations in face images is characteristic of each particular AFR application, the set of optimal weights is application dependent.

In fact, the relative importance of each facial region will depend on the variations the AFR system has to deal with. For instance, the region around the mouth may be more or less discriminative depending on the expected amount of facial expression variation. Since the type and amount of variations in face images is characteristic of each particular AFR application, the set of optimal weights is application dependent.

However, in most LBP based methods proposed thus far [6] [2], the choice of weight values is carried out empirically and guided by the human visual perception without concern for the kind of variations present in the working datasets. In a recent paper [10] we proposed a new method that fits the weights to each particular application, as long as a representative set of training images can be provided. However, this method is fairly complex and entails a large number of training images per weight being estimated.

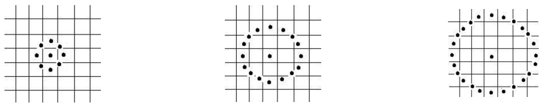
This paper addresses these issues and proposes a new method that defines the optimal set of weights as the discriminant axis produced by Linear Discriminant Analysis (LDA) [11]. The method was implemented upon the algorithm proposed in [2] and tested on the FERET database having achieved a better performance when compared to other weighting methods proposed in the literature. The experiments also indicated that the method is able to capture the characteristic variations of each application and requires comparatively few training samples per weight being estimated.

The remainder of this paper is organized as follows: the next section briefly describes the computation of LBP image representation and how it is used for recognition. Section III presents the proposed LDA approach. Experimental set up and results are presented in section IV and the main conclusions are summarized in section V.

II. LBP APPLIED TO FACE RECOGNITION

Local Binary Patterns were originally designed for texture description [7] in gray scale images. The original method and some extensions have been applied to face recognition [5]. The description below is a summary of the so called uniform LBP [7].

In the first step the LBP image representation is computed, assigning to each pixel the corresponding LBP code. This is given by the sign of the difference between the central pixel intensity and the intensity of its m neighbors arranged over a circle of radius R, as illustrated in figure 1. The binary values "0" and "1" are assigned respectively to a negative and to a positive difference. A bilinear interpolation is used whenever the sampling point does not fall in the center of a pixel.



m-8, R=1 *m=16, R=2* *m=24, R=3*

Figure 1. Sets of neighboring pixels for different values of m and R.

In the recognition step the LBP image is divided in equal-sized non-overlapping blocks numbered from 1 to B, as illustrated in figure 2.

(a) (b)

Figure 2. A sample image (a) and its LBP representation (b) divided in 10×8 blocks.

The histogram ${}^{b}H$ of LBP codes inside the block b is compute for $b=1,2,\ldots,B$. In uniform LBP approach, only codes with two or less "0" to "1" or "1" to "0" transitions are taken in the computation, considering the LBP code to be circular.

The dissimilarity between corresponding blocks of two LBP images X and Y may be computed in different ways [8], being the weighted Chi-square distance the most cited one in the literature. This is given by

$$\chi^2\left({}^{b}H_X, {}^{b}H_Y\right) = \sum_c \frac{\left[{}^{b}H_X(c) - {}^{b}H_Y(c)\right]^2}{{}^{b}H_X(c) + {}^{b}H_Y(c)} \tag{1}$$

where ${}^{b}H_X(c)$ and ${}^{b}H_Y(c)$ are the cth histogram counts for block b respectively of images X and Y.

The overall Chi-square distance between X and Y is computed by summing up the block distances, i.e.,

$$\chi^2(H_X, H_Y) = \sum_b w_b \chi^2\left({}^{b}H_X, {}^{b}H_Y\right). \tag{2}$$

where w_b is the weight associated to bth block.

The decision whether or not X and Y are facial images of the same subject will be based upon the overall Chi-square distance between them.

III. LDA Based Recognition Procedure

A. Method Description

We assume that the face database consists exclusively of well-framed images with a constant intraocular distance and the eyes imaged at the same pixel coordinates for all samples.

Let's denote with $\underline{S}$ the available face data base and with $\mathbf{S}_i$ the set of face images of the ith subject represented in $\underline{S}$. Thus

$$\underline{\mathbf{S}} = \bigcup_{i=1}^{K} \mathbf{S}_i \tag{3}$$

where K is the number of subjects with images in $\underline{S}$. Let's further denote with S_{ij} the jth image in $\mathbf{S}_i$, i.e.,

$$\mathbf{S}_i = \left\{S_{i1}, S_{i2}, \ldots, S_{ik_i}\right\} \tag{4}$$

where k_i is the number of faces of the ith subject represented in $\underline{S}$. We further denote with the symbol ${}^{b}H_{ij}$, the LBP histograms computed within the bth block of image S_{ij}.

The Chi-square distance ${}^{b}D_{irjt}$ between the histograms ${}^{b}H_{ir}$ and ${}^{b}H_{jt}$ of the bth blocks respectively of face images S_{ir} and S_{jt} will be denoted by

$${}^{b}D_{irjt} = \chi^2\left({}^{b}H_{ir}, {}^{b}H_{jt}\right) \tag{5}$$

It is expected that pairs of images from the same individual (i.e., $i=j$) have lower distance values, when compared with pairs of images from different individuals (i.e., $i \neq j$).

Considering a pair of images divided in B blocks, it is possible to define a B dimensional distance vector $\mathbf{D}_{irjt}$ between the images S_{ir} and S_{jt} as

$$\mathbf{D}_{irjt} = \left\{{}^{1}D_{irjt}, {}^{2}D_{irjt}, \ldots, {}^{B}D_{irjt}\right\} \tag{6}$$

The LDA procedure finds the axis in the space of distance vectors that better discriminates the two classes of our problem, i.e., pairs of images from the same individual, and pairs of images from different individuals. Along this axis the variability between classes relative to the common variability within classes is maximum.

It can be determined in the following way. First, collect a set of distance vectors $\mathbf{D}_{irjt}$ from each class. Then, compute the within class and the between class covariance matrices, denoted respectively $\mathbf{\Sigma}_W$ and $\mathbf{\Sigma}_B$, upon the collected vectors. Finally, solve the eigenvalue/eigenvector problem stated below

$$\Sigma_W^{-1} \Sigma_B \mathbf{x} = \lambda \mathbf{x} \tag{7}$$

The vector $\mathbf{x}$ corresponding to highest scalar λ among the solutions of the problem above is the discriminant axis we are after. Further details of the LDA technique including the estimation of the within and between classes covariance matrices can be found in most text books on Multivariate Statistical Analysis (e.g. [11]).

The dissimilarity between a pair of face images will be expressed in this proposal by the projection of the distance vector relative to these two images upon the discriminant axis **x**. Clearly, for a pair of equal images the distance vector as well as its projection will be equal to zero.

It is important to point out that the weights delivered by the proposed LDA based method generally do not follow the expected left-to-right face symmetry. Unless the images are strictly symmetric, which occurs in virtually no real database, the components of the vector that defines the discriminant axis may be far from symmetric. The symmetry constraint may be imposed to reduce the problem complexity and to improve generalization. This can be done, by computing a reduced version (**Dr**) of each distance vector used for training, as defined by

$$\mathbf{Dr}_{irjt} = \left\{ \left({}^{1}D_{irjt} + {}^{1}Ds_{irjt} \right), \ldots, \left({}^{B/2}D_{irjt} + {}^{B/2}Ds_{irjt} \right) \right\} \qquad (8)$$

where ${}^{b}Ds_{irjt}$ is the element of $\mathbf{D}_{irjt}$ corresponding to the block symmetric to block b in the face image, for $b=1,2,\ldots, B/2$ and B even. The LDA procedure applied to the reduced distance vectors will produce a $B/2$-dimensional vector, which refers to the left (or right) face half. The components related to the other face half can be obtained by mirroring. The derivation for B odd is similar.

The symmetric LDA procedure may be advantageous when training samples are scarce. On the other hand, the non-symmetric LDA procedure may be preferable when the work images have some relevant asymmetric property, such as a non uniform illumination pattern.

It is further worth mentioning that the vector defining the discriminant axis has often negative components. This may be counter-intuitive if one thinks on these components as weights that express the relative importance of face regions in the recognition process. As a matter of fact, humans tend to regard as little important the face regions that do not change significantly from subject to subject. Thus, the weights associated to these regions are expected to be comparatively small.

This analogy applies only partially to the components of the LDA discriminant vector. Besides considering the interclass variation expressed in $\mathbf{\Sigma}_B$, the LDA procedure also favors directions with low within class variability as captured in $\mathbf{\Sigma}_W$. Therefore, face regions with low subject-to-subject variation may be regard as significant by the LDA approach. Therefore, the parallel between human perception about the relevance of face regions and the components of the discriminant vector does not apply perfectly.

The inclusion of the within class variance in the problem model is also responsible for bringing about negative components of the discriminant vectors.

IV. Experiments

A. Experiment Setup

The experiments for performance assessment used a data set containing 1062 images of 531 subjects, 2 images per subject, from the *fa* and *fb* sets of the FERET database [9]. This set consists of frontal images with slight variation in facial expressions.

The images were resized to 80×64 pixel resolution, which was defined experimentally in a previous work [10]. The center of the right eye is located at pixels coordinates (*row, column*) = (20, 14). Figure 2a shows a typical image sample used through all experiments.

16	16	16	1	1	16	16	16
2	11	10	15	15	10	11	2
11	-6	7	6	6	7	-6	11
2	16	11	18	18	11	16	2
4	-2	2	8	8	2	-2	4
-13	-5	-1	21	21	-1	-5	-13
3	14	5	2	2	5	14	3
-8	4	10	1	1	10	4	-8
21	8	15	-1	-1	15	8	21
11	19	2	-4	-4	2	19	11

symmetric LDA

15	1	20	4	0	12	18	16
1	3	22	9	2	9	7	26
11	1	-2	0	-2	5	11	-12
1	5	16	30	6	10	11	2
2	-4	-8	-9	12	-18	-5	-9
-12	-2	16	19	19	7	1	1
1	6	1	-1	-2	1	15	-1
-1	4	5	-1	9	9	-5	10
10	7	14	-5	11	1	6	0
0	8	-3	-4	-2	5	15	8

non-symmetric LDA

21	11	11	11	11	11	11	21
21	22	22	11	11	22	22	21
21	22	22	11	11	22	22	21
5	5	5	15	15	5	5	5
0	5	5	15	15	5	5	0
0	5	5	15	15	5	5	0
0	5	5	5	5	5	5	0
0	5	5	8	8	5	5	0
0	5	5	8	8	5	5	0
0	5	5	5	5	5	5	0

Feitosa et al.

16	10	9	8	8	9	10	16
17	21	22	14	14	22	21	17
15	24	26	15	15	26	24	15
9	10	10	5	5	10	10	9
3	7	8	4	4	8	7	3
0	6	8	5	5	8	6	0
0	6	8	8	8	8	6	0
0	6	8	12	12	8	6	0
0	6	8	11	11	8	6	0
0	6	8	8	8	8	6	0

Ahonen et al.

Figure 3. Weighting delivered by the proposed method with and without the symmetry constraint, by Feitosa et al. and by Ahonen et al.

We selected in our experiments the uniform LBP variant with 8 sampling points ($m=8$), over a circle of radius 2 pixels ($R=2$), and 10×8 non-overlapping blocks of size 8×8.

We separated the subjects in two randomly selected groups. One group was used for computing the best discriminating axis or weighting estimation and the other group was used for test. The experiment was carried out twice by switching the estimation and test groups. The results reported henceforth are the average computed in both experiments for each database.

B. Weights Estimates

Figure 3 shows the average weights obtained by our method with (symmetric LDA) and without (non-symmetric LDA) the symmetry constraint, as well as the weights computed by the method proposed in [10] and the weights proposed in [4]. In all cases the weights were scaled so as to form a vector with magnitude equal to 100, for interpretation easiness. The weight matrix proposed by Ahonen et al. in [4] was resampled so as to fit the 10×8 grid.

With the method of Feitosa et al. we imposed as in [10] the restriction that certain groups of blocks should have equal weights. This was mandatory because the method did not generalize well for more weights with the available face. Ahonen et al. proposed in [2] a 7×7 weight matrix that we resized to a 10×8 arrangement to keep consistency for all methods. In the LDA method we estimated 40 and 80 weights respectively for the symmetric and for the non-symmetric version. So, we had in these cases a much higher degree of freedom than we had with Feitosa's method.

The weights provided by the non-symmetric LDA method show no perceptible structure. This suggests that there is an unbalance between model complexity and the number of available training images.

C. *Performance Results*

Figure 4 shows the average rank recognition rates achieved with the weightings presented in figure 3. For clarity, the performance obtained with uniform weighting is also included. Recall that these rates refer to the average performance achieved by two experiments using weights estimated in one group and tested in another group of the same database, which are then switched.

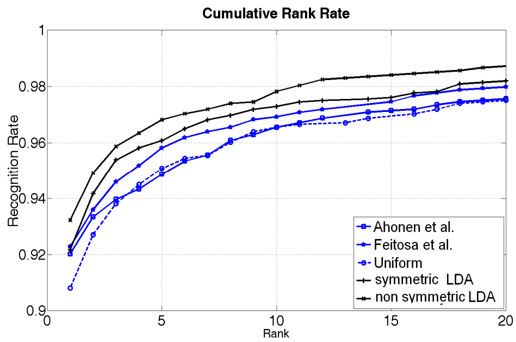


Figure 4. Cumulative rank rates for different weightings

Among the alternatives shown in figure 5, our method was able to consistently deliver the highest performance among the other methods. The method was able to provide considerably different weights, which result in better recognition rates when compared with the ones proposed in the literature.

It is also worth mentioning that all rates are fairly close to 100%. Under these conditions the absolute performance improvements can not be high, and should not be underestimated. In our experiments one single image corresponds to less than 0.2% in the recognition rate.

As mentioned before, the symmetric version of the LDA method is expected to present better generalization characteristics, i.e., to perform better on images not used in the computation of the discriminant axis. On contrary, our experiments indicated the non-symmetric version as the best among all considered alternatives. This suggests that the work images in the set used for estimation as well as in the test set have some asymmetric properties, which were captured by the non-symmetric LDA method. These results show the ability of the LDA method to capture specific variability in the images of each particular application. Yet, the choice between the symmetric or non-symmetric version should necessarily take into account the image characteristics of the target application.

V. Conclusions and Future Work

In this work we presented a novel method to estimate optimal facial region weights for LBP based face recognition. More than proposing a standard general weighting for most applications, the present work proposes a method to determine a set of weights that best fits the image characteristics within a particular application. The method was presented in two versions. The first version explores the horizontal symmetry of face images so as to simplify the estimation problem and to provide improved generalization. The second version disregards the face symmetry and has the potential to capture asymmetric properties of the work images.

Experiments based on FERET database have shown the superiority of the values obtained by the proposed method in comparison to other weightings and weighting methods proposed in the literature.

In spite of the good rates achieved in our experiments the highly uneven weight distribution obtained with the non-symmetric LDA version suggests that there is still room for improvements, if we manage to reduce the model complexity by exploring some form of prior knowledge related to the face structure. It is further subject of a future investigation the performance assessment of our method for larger databases containing types of facial image variations other than the ones present in the currently used databases.

References

[1] S. Z. Li and A. K. Jain, Eds., Handbook of Face Recognition. Springer, 2005.

[2] T. Ahonen, A. Hadid and M. Pietikäinen, Face recognition with local binary patterns, in Proc European Conference on Computer Vision, Prague, pp. 469–481, 2004.

[3] G. Heusch, Y. Rodriguez and S. Marcel, Local Binary Patterns as an Image Preprocessing for Face Authentication, in Proc. FG' 2006, pp. 9-14, April 2006.

[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] S. Marcel, Y. Rodriguez and G. Heusch, On the Recent Use of Local Binary Patterns for Face Authentication, International Journal of Image and Video Processing , Special Issue on Facial Image Processing, pp.1-9, 2007.

[6] S. Liao, W. Fan, A. C. S. Chung and D.Y. Yeung, Facial expression recognition using advanced local binary patterns, Tsallis entropies and global appearance features, in Proc. ICIP'2006, pp.665-668, 2006.

[7] 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, Jul 2002.

[8] H.K. Ekenel, M. Fischer, E. Tekeli, R. Stiefelhagen, A. Ercil, Local Binary Pattern Domain Local Appearance Face Recognition, in Proc SIU 2008, pp. 1 – 4, April 2008.

[9] W. Zhao, R. Chellappa, P.J. Phillips and A. Rosenfeld, Face Recognition: A literature Survey, ACM Computing Surveys, vol. 35, no. 3, pp. 399–458, 2003.

[10] R.Q. Feitosa, A.C. Censi, R. Pithan, J.L.B. Pinho, Optimal Region Weighting for Local Binary Pattern Based Face Recognition, in Proc. ISSNIP Biosignals and Biorobotics Conference, 2010.

[11] R.A. Johnson, D.W., Wichern, Applied Multivariate Statistical Analysis, Prentice Hall, New Jersey, 1998.