Neural network & LBP-Two dimesional subspace  algorithm based face recognition system

Mr.BENOUIES MOHAMED¹
Department informatique & mathematique
¹Oran University
Oran,Algeria
mhbenouis@yahoo.com

Mr.BELKACEM AMAR²
Department of informatique
²University Kasdi merbah
Ourgla,Algeria
am.belgacem@gmail.com

Abstract—in this paper proposes a novel approach to face recognition based on the combination of feature extraction methods, such as two-dimensional LBP-2DPCA and LBP-2DLDA, with a neural network (MLP). This later is used to classify the features matrix extracts for space data created by Two-dimensional Subspace Analysis. The proposed approach is tested on ORL and FEI face databases. Experimental results on these databases demonstrated the effectiveness of the proposed approach for face recognition with high accuracy compared with previous methods.

Keywords—biometric, face recognition, 2DPCA, 2DLDA, LBP, MLP

I. INTRODUCTION

The security of persons, goods or information is one the major concerns of the modern societies. Face recognition is one of the most commonly used solutions to perform automatic identification of persons. However, automatic face recognition should consider several factors that contribute to the complexity of this task such as the occultation, changes in lighting, pose, expression and structural components (hair, beard, glasses, etc.).

Several techniques have been proposed in the past in order to solve face recognition problems. Each of them evidently has its strengths and weaknesses which, in most of the cases, depend on the conditions of acquiring information[1]. One of the crucial aspects in automatic facial recognition is facial representation, which derives a set of features from original face images to effectively represent faces. The major approaches developed for face recognition are principal component analysis, analysis discriminate linear, independent component analysis which provides an optimal linear transformation from the original image space to an orthogonal eigenspace with reduced dimensionality in the sense of least mean squared reconstruction error (MSE)[14]. LDA seeks to find a linear transformation by maximizing the between-class variance and minimizing the within-class variance. Both of them represent a face with holistic facial features. The local binary pattern (LBP) has been widely applied to texture classification and face recognition as an effective arithmetic operator in texture description, with rotational and gray scale invariance.

Recently, several efforts and research in this domain have been done in order to increase the performance of the recognition, such as support vector machine (SVM), Markov hidden model (HMM), probabilistic methods (Bayesian networks) and artificial neural networks[2]. This latter has attracted researchers because of its effectiveness in detection and classification of shapes which has been adopted in new face recognition systems.

II. FACE RECOGNITION SYSTEM

A face recognition system is a system used for the identification and verification of individuals, which checks if a person belongs to the system’s database, and identifies him/her if this is the case.

The methods used in face recognition based on 2D images are divided into three categories: global, local and hybrid methods.

- Local or analytical facial features approaches. These type consists on applying transformations in specific locations of the image, most frequently around the features points (corners of the eyes, mouth, nose,...). They therefore require a prior knowledge of the images...
- Global approaches use the entire surface of the face as a source of information without considering the local characteristics such as eyes, mouth, etc..
- Hybrid methods associate the advantages of global and local methods by combining the detection of geometrical characteristics (or structural) with the extraction of local appearance characteristics.

In our work, we use Local Binary Pattern (LBP) to present a face. This method can encode each pixel to integral, which contains the information of gray value of this pixel and its neighborhood. Then, we can define some regions, and count
the histogram of those integral. The histograms are the final presentation of each face image. This method is not easily affected by global change of illuminations and slight rotation of the face. After, we are interested in two dimensional features extraction methods of the face which preserves the original shape of the face in a two-dimensional array (matrix) compared to the one-dimensional methods which represent a face by a one-dimensional array (vector), this passage from a matrix to a vector loses some geometric and temporal information related to the pixels of the image.

A. Two-dimensional principal component approach analysis (2DPCA)

Proposed by Yang in 2004 [3][4], 2DPCA is a method of feature extraction and dimensionality reduction based on Principal Component Analysis (PCA) that deals directly with face images as matrices without having to turn them into vectors like as the traditional global approach.

1) The steps of face recognition by 2DPCA

Considering a training set \( S \) of \( N \) face images, the idea of this technique is to project a matrix \( X \) of size \((n \times m)\) via a linear transformation such that:

\[
Y_i = X.R_i
\]

Where \( Y_i \) is the principal component vector of size \((n \times 1)\), and \( R_i \) is the base projection vector of size \((m \times 1)\). The optimal vector \( R_i \) of the projection is obtained by maximizing the total generalized variance criterion

\[
J(R) = R^T.G.R
\]

Where \( G \) is the covariance matrix of size \((m \times m)\) given by:

\[
G = \frac{1}{M} \sum_{j=1}^{M} (X_j - \bar{X})^T(X_j - \bar{X})
\]

With \( X_j \): the \( j \)th image of the training set

\[
\bar{X} \text{ : The average image of all the images in the training set}
\]

\[
\bar{X} = \frac{1}{M} \sum_{j=1}^{M} X_j
\]

In general, one optimal projection axis is not enough. We must select a set of projection axes like:

\[
\{R_1, R_2, \ldots, R_d\} = \text{arg} \max J(R)
\]

These axes are the eigenvectors of the covariance matrix corresponding to the largest “d” eigenvalues. The extraction of characteristics of an image using 2DPCA is as follows

\[
Y_k = X.R_k \quad ; k = 1, \ldots, d
\]

Where \([R_1, R_2, \ldots, R_d]\) is the projection matrix and \([Y_1, Y_2, \ldots, Y_d]\) is the features matrix of the image \( X \).

B. the 2DLDA approach

In 2004, Li and Yuan [5] have proposed a new two-dimensional LDA approach. The main difference between 2DLDA and the classic LDA is in the data representation model. Classic LDA is based on the analysis of vectors, while the 2DLDA algorithm is based on the analysis of matrices.

1) Face recognition using 2D LDA

Let \( X \) be a vector of the \( n \)-dimensional unitary columns. The main idea of this approach is to project the random image matrix of size \((m \times n)\) on \( X \) by the following linear transformation:

\[
Y_i = A_jX
\]

\( Y \) : the \( m \)-dimensional feature vector of the projected image

2DLDA approach

In 2004, Li and Yuan [5] have proposed a new two-dimensional LDA approach. The main difference between 2DLDA and the classic LDA is in the data representation model. Classic LDA is based on the analysis of vectors, while the 2DLDA algorithm is based on the analysis of matrices.

1) Face recognition using 2D LDA

Let \( X \) be a vector of the \( n \)-dimensional unitary columns. The main idea of this approach is to project the random image matrix of size \((m \times n)\) on \( X \) by the following linear transformation:

\[
Y_i = A_jX
\]

\( Y \) : the \( m \)-dimensional feature vector of the projected image

A. Let us suppose \( L \) : class numbers.

\( \mathbf{M} \) : the total number of training images

The training image is represented by a matrix \( m \times n \)

\[
A_j(j = 1, \ldots, M)
\]

\( \bar{A}_i \) : the mean of all classes

\( N_i \) : Number of samples in each class

The optimal vector projection is selected as a matrix with orthonormal columns that maximizes the ratio of the determinant of the dispersion matrix of the projected inter-class images to the determinant of the dispersion matrix of the projected intra-class images:

\[
J_{FLD}(W_{opt}) = \arg \max W \frac{|W^T S_b W|}{|W^T S_w W|}
\]

\( P_b = \text{trace} (S_b) \)

\( P_w = \text{trace} (S_w) \)

Where, \( S_b \) : The inter-class dispersion matrix,

\( S_w \) : The intra-class dispersion matrix.

\[
S_b = \sum_{i=1}^{L} N_i [(\bar{A}_i - \bar{A})X][((\bar{A}_i - \bar{A})X)^T
\]

\[
S_w = \sum_{i=1}^{L} \sum_{x \in \mathcal{C}_i} [X - (\bar{A}_i - \bar{A})X][X - (\bar{A}_i - \bar{A})X]^T
\]

The criterion can be expressed by:

\[
J(X) = \frac{X^T S_b X}{X^T S_w X}
\]

Where \( X \) : unitary column vector.

The unitary vector \( X \) maximizing \( J(X) \) is called the optimal projection axis. The optimal projection is chosen when \( X_{opt} \) maximizes the criterion, as the following equation:

\[
X_{opt} = \arg \max X J(X)
\]

If \( S_w \) is invertible, the solution of optimization is to solve the generalized eigenvalue problem.

\[
S_b X_{opt} = \lambda S_w X_{opt}
\]

Such that \( \lambda \) is the maximum eigenvalue of \( S_w^{-1} S_b \)
In general, it is not enough to have only one optimal projection axis. We need to select a set of projection axes \( x_1, x_2, \ldots, x_d \) under the following constraints:

\[
\{x_1, x_2, \ldots, x_d\} = \operatorname{argmax}_X J(X)
\]

\[
X^T X_j = 0, i \neq j, i, j = 1, 2, \ldots, d
\]

Indeed, the optimal projection axes \( x_1, x_2, \ldots, x_d \) are orthonormal eigenvectors of \( S_W^{-1} S_B \) corresponding to the best first \( d \) eigenvalues permitting to create a new projection matrix \( X \), which is a matrix of size \( n \times d \):

\[
X = [x_1, x_2, \ldots, x_d]
\]

2) Feature extraction:

We will use the 2DLDA optimal projection vectors \( x_1, x_2, \ldots, x_d \) to extract the image features; we use the following equation:

\[
y_k = AX_k, \quad k = 1, 2, \ldots, d
\]

Then, we have a family of feature vectors \( y_1, y_2, \ldots, y_d \) which form the matrix \( y = [y_1, y_2, \ldots, y_d] \) of size \( M \times d \) called image characteristics matrix \( A \).

III. LOCAL BINARY PATTERN (LBP)

LBP descriptor computed using LBP operator introduced by Ojala et al. [9] is one of the widely used texture descriptors that have shown promising results in many applications [13],[14], Ahonen et al. [12] used it for face recognition, Lian and Lu [10] and Sun et al. [11] employed it for gender recognition. The initial LBP operator associates a label with each pixel of an image; the label is obtained by converting each pixel value in the 3x3-neighbourhood of a pixel into a binary digit (0 or 1) using the center value as a threshold and concatenating the bits, as shown in Figure 1. Later the operator was extended to general neighborhood sizes, and its rotation invariant and uniform versions were introduced [9].

![Figure 1: LBP Operator](image)

The general LBP operator is denoted by \( LBP_{P,R} \) and is defined as follows:

\[
LBP_{P,R} = \sum_{i=1}^{P} 2^S(p_i - p_c)
\]

where \( P \) is the total number of pixels in the neighborhood and \( R \) is its radius, \( p_c \) is the center pixel and the thresholding operation is defined as follows:

\[
S(p_i - p_c) = \begin{cases} 
1 & p_i - p_c \geq 0 \\
0 & p_i - p_c < 0.
\end{cases}
\]

Commonly used neighborhoods are \((8, 1), (8, 2), \text{ and } (16, 2)\).

The histogram of the labels is used as a texture descriptor. The histogram of labeled image \( f_i(x, y) \) is defined as:

\[
H(i) = \sum_{x,y} I(f_i(x,y) = i), \quad i = 0, 1, \ldots, n - 1
\]

where \( n \) is the number of different labels produced by the LBP operator and

\[
I(x) = \begin{cases} 
1 & x \text{ is true} \\
0 & x \text{ is false}.
\end{cases}
\]

Figure 2 shows the histogram extracted from an image with LBP operator.

![Figure 2: LBP Histogram Calculation for Full image](image)

To overcome this issue, spatially enhanced LBP histogram is calculated. Figure.1 shows the process of computing spatially enhanced LBP histogram. An image is divided into blocks; LBP histogram is calculated from each block and concatenated.

General LBP operator has three parameters: circular neighborhood \((P, R)\), rotation invariance \((ri)\) and uniformity \((u2)\). For a particular application, it is necessary to explore this parameter space to come up with the best combination of these parameters. In this Paper we will explore Uniform version of LBP with \( P \) and \( R \) as 8 and 1.

The idea behind using the LBP features is that the face images can be seen as composition of micro-patterns which are invariant with respect to monotonic grey scale transformations.

IV. FACE CLASSIFICATION USING NEURAL NETWORKS

Several studies have shown improved face recognition systems using a neural classification compared to classification based on Euclidean distance measure.

A. The neural network

An artificial neural network is a computational model whose design is inspired by the schematic of biological neurons[15].

Our neural network is a multi-layer using a supervised back propagation algorithm for training, has a single hidden layer, the input number N depends directly on the size of the image.

a) The parameters MLP: To adjust the parameters of the network, using the standard technique Retro-propagation method is a type of gradient descent, looking for a local minimum acceptable to obtain a minimum error. The error is defined as the root mean square differences between the real and desired outputs of the neural network.
c) The Training phase of MLP: The training phase then returns to find the topology (Number of neurons in the hidden layer) and the optimal parameters of the MLP (Multi-Layer Perceptron). As for the test phase, it projects images of faces to be recognized in the new subspace using the projection matrices, determined in the training phase, and then perform the classification using the neural network MLP[16][17].

Face recognition system based on LBP-2DPCA/2DLDA consists of two processes, training and recognition processes. In the training process: The face image is first divided into small regions from which the Local Binary Pattern (LBP) features [8,9] are extracted and represented into matrix feature histogram efficiently representing the face image. The textures of the facial regions are locally encoded by the LBP patterns while the whole shape of the face is recovered by the construction of the face feature histogram. We applied Two-dimensional Subspace Analysis (2DPCA,2DLDA) which are capable of creating the face subspace, defined as orthogonal basis of vectors that contain the most relevant information about a face[12]. These vectors are the eigenvectors of the covariance matrix of the distribution. The recognition process is achieved by forwarding feature matrix which must be transformed into vectors before providing them to the MLP classifier to perform classification and decision.

V. RESULTS AND DISCUSSION

In order to evaluate and test our approach described for face recognition system, we chose three databases: ORL, FEI[19][20] and our database of our laboratory. All experiences were performed in Matlab installed on a laptop with a dual core processor T5870 with 2.03 GHz and 2 GB of RAM.

A) OWN database:

We have collected face images at different moments using a capture device (webcam) to form our own database. The database includes 100 face images in a JPG taken on 10 different subjects (N = 10), each been registered under 10 different views.

All experiences have been performed using the ORL and the FEI databases with 5 images for training and 5 images for the test per person for a total of 200 images for each phase (training and test).

B) First experience

In this experience, we compare the performance of the 2DPCA and PCA approaches in terms of recognition rate and execution time. We use the ORL database.

To better illustrate the results of the PCA and 2DPCA approaches, we fix the number of dimensions of both PCA and 2DPCA eigenfaces that give better recognition rate.

After a series of experiences, we choose the best values of parameters in order to fix the choice of eigenvectors which give a better recognition rate.

<table>
<thead>
<tr>
<th></th>
<th>2DPCA</th>
<th>PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recognition rate</td>
<td>96.00</td>
<td>94.00</td>
</tr>
<tr>
<td>Running time (s)</td>
<td>33.718743</td>
<td>55.442794</td>
</tr>
</tbody>
</table>

Tab 1: comparison between PCA and 2D-PCA based on ORL

Discussion

- Table (1) illustrates a comparison between the PCA and 2DPCA approaches in terms of recognition rate and running time, we find that the recognition rate obtained by 2DPCA is better compared to that obtained with PCA.
- We notice that the running time of 2DPCA is less than that of PCA.
- According to the above figures, we noticed that the image reconstruction using 2D-PCA approach is relatively better than the PCA in the matter of the projection error (1.43% for 2D-PCA versus 22.95 for PCA)

Test image

\[ \|e_{projection}\| = 1.43\% \]

figure 4: The reconstruction of a test image using 2DPCA

Test image

\[ \|e_{projection}\| = 22.95\% \]

figure 5: The reconstruction of a test image using PCA

C) SECOND EXPERIENCE

To evaluate the performance of our proposed approach, we chose two test databases: ORL and FEI.

1) Database: The global performance of algorithms tested on the FEI database is not as good as that of the ORL database. There are two main reasons:

- The image quality of the ORL database is better than that of the FEI database.
The FEI database is more complex due to variations in the face details and head orientations.

In this part, several LBP based algorithms are tested in order to demonstrate the superiority of the proposed approach. Conventionally, selecting the image division strategy for LBP based approaches is heuristical or empirical. Results in previous publications are often obtained by division strategies which maximize the recognition rates. Here we tested various algorithms with different division strategies. In our test, an image is divided into 4 × 4, 7 × 7 and 16 × 16 sub-blocks respectively.

The LBP based face recognition approach consists of extracting the LBP features histograms from the whole face (59-bin histogram), divide the face image into 7x7 blocks and concatenate the blocks histograms into a unique vector (7x7 blocks × 59-bin/block = 2891 features). These features are calculated for each individual sample and a mean model is computed for each ID subset. In face identification step, given a new input face image, the 2891 features vector is extracted and compared to all available models by means of a dissimilarity metric. The input face will be identified with the minimum ID model distance.

3) Feature extraction using 2DPCA/2DLDA: After reducing the dimensional of the face images using LBP descriptor, we used the 2DPCA and 2DLDA feature extraction approaches in order to extract the weight images (Features images in the new space) which must be converted into vectors before implementing the MLP network.

4) Choice of the number of eigenvalues: Two dimensional methods do not escape this problem, and the choice of the appropriate number depends on the used method and faces database. In our experiences, we have selected the best eigenvalues corresponding to the best variance values (eigenvectors)

e) Selection parameters training: MLP To get a higher recognition rate, we have made a series of experiments to choose the best topology MLP based on parameters as follows:

Example: the parameters training MLP

<table>
<thead>
<tr>
<th>performance function</th>
<th>MSE</th>
<th>Activation function</th>
<th>Tansig</th>
<th>desired error</th>
<th>0.001</th>
<th>Hidden layer number</th>
<th>100</th>
<th>Number epochs</th>
<th>600</th>
<th>Layer input</th>
<th>Feature vector</th>
<th>Layer output</th>
<th>Each output for a face</th>
</tr>
</thead>
</table>

In this experiment, we compared the different methods described in this paper and evaluated on two databases, ORL and FEI. We followed the same protocol testing previous experiences.

we encoded a faces images based technique LBP to reduce the affect of illumination ,noise and size. These databases in other hand reduce the memory and compute of our neural network training algorithm (MLP)

6) Adding some effects: In this work, it is wanted to test our system with and without added noisy in the two data base in order to evaluate robustness of these approaches namely 2DPCA, 2DLDA, LBP-2DPCA, LBP-2DLDA combined by using two classifier : KNN (Euclidean distance) and MLP

7) Noise: Two types of noise are used in this simulation: the Saltand Pepper type noise with a noise density a=0.06 (Figure 5 (a)) and Gaussian noise with mean m=0, variance v=0.04 Figure 6 illustrates these effects which are obtained as follows.

(a)Salt&pepper Noise (b) Gaussian Noise (e) Gaussian Noise m=0, v=0.01

Figure 6: Adding Noise(database face ORL &FEI)

Table 2. The recognition rate obtained by different methods on the database ORL

<table>
<thead>
<tr>
<th>Type of classifier</th>
<th>2DPCA</th>
<th>2DLDA</th>
<th>LBP-2DPCA</th>
<th>LBP-2DLDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euclidean</td>
<td>94%</td>
<td>94.8%</td>
<td>94%</td>
<td>95%</td>
</tr>
<tr>
<td>MLP</td>
<td>95.8</td>
<td>96%</td>
<td>97%</td>
<td>98%</td>
</tr>
<tr>
<td>Running time</td>
<td>35 s</td>
<td>38 s</td>
<td>22 s</td>
<td>20 s</td>
</tr>
</tbody>
</table>

Table 3. The recognition rate obtained by different methods on the database ORL with added noise

<table>
<thead>
<tr>
<th>Type of classifier</th>
<th>2DPCA</th>
<th>2DLDA</th>
<th>LBP-2DPCA</th>
<th>LBP-2DLDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euclidean</td>
<td>90%</td>
<td>91%</td>
<td>90%</td>
<td>92%</td>
</tr>
<tr>
<td>MLP</td>
<td>92%</td>
<td>92%</td>
<td>94%</td>
<td>95%</td>
</tr>
<tr>
<td>Running time</td>
<td>41 s</td>
<td>45 s</td>
<td>25 s</td>
<td>27 s</td>
</tr>
<tr>
<td>Type of classifier</td>
<td>2DPCA</td>
<td>2DLDA</td>
<td>LBP-2DPCA</td>
<td>LBP-2DLDAs</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------</td>
<td>-------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Distance Euclidean</td>
<td>80%</td>
<td>82.8%</td>
<td>90%</td>
<td>94%</td>
</tr>
<tr>
<td>MLP</td>
<td>88%</td>
<td>90%</td>
<td>93%</td>
<td>96%</td>
</tr>
<tr>
<td>Running time</td>
<td>42 s</td>
<td>41 s</td>
<td>34 s</td>
<td>33 s</td>
</tr>
</tbody>
</table>

Table 5. The recognition rate obtained by different methods on the database FEI with added noisy

<table>
<thead>
<tr>
<th>Type of classifier</th>
<th>2DPCA</th>
<th>2DLDA</th>
<th>LBP-2DPCA</th>
<th>LBP-2DLDAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance Euclidean</td>
<td>80%</td>
<td>82.8%</td>
<td>85%</td>
<td>90%</td>
</tr>
<tr>
<td>MLP</td>
<td>85.5%</td>
<td>88%</td>
<td>90%</td>
<td>92%</td>
</tr>
<tr>
<td>Running time</td>
<td>45 s</td>
<td>48 s</td>
<td>38 s</td>
<td>37 s</td>
</tr>
</tbody>
</table>

Discussion:

After these series of experiments we clearly see the superiority of LBP-2DPCA & LBP-2DLDAs methods combined with a neural network classifier combining those of a Euclidean distance classifier. We also note that the choice of optimal component and the choice smoothing parameter which represents a better recognition rate for both methods, 2DPCA and 2DLDA and accuracy of classification MLP.

The results obtained after using our approach shows that it reduces the computation time of training and improves the recognition rate.

IV. CONCLUSION

In this paper, an efficient facial expression representation and classification methodology is proposed. Features extraction based on LBP and 2DPCA called LBP-2DPCA. Feature vectors are computed from histograms of LBP. These features are classified using the two layer feed forward neural network. For training this network multi layer perceptron (MLP), back propagation optimization algorithm is used.

In this paper, we propose an approach for face recognition based on the combination of two approaches, one used for the reduction of space and feature extractions in two dimensions and the other for classification and decision.

Our choice of using LBP techniques as a preprocessing stage and is applied to obtain low dimensional feature vectors is demonstrated by improved performance of our system in terms of recognition rate, speed of calculation and reduce memory computation of MLP.

As a perspective, we propose to use this approach in an uncontrolled environment (video surveillance) based on video sequences (dynamic images) in order to make the task of face recognition more robust.

REFERENCES

[1] Handbook of Biometrics edited by Anil K. Jain Michigan State University, USA Patrick Flynn University of Notre Dame, USA Arun A. Ross West Virginia University, USA © 2008 Springer Science+Business Media, LLC.

[2] biometric recognition challenges and opportunities Joseph N. Pato and Lynette I. Millett, Editors Whither Biometrics Committee Computer Science and Telecommunications Board Division on Engineering and Physical Sciences Copyright 2010 by the National Academy of Sciences,


