ISSN 2229-5518

# Face Recognition Using LBP, FLD and SVM with Single Training Sample Per Person

Mustafa Zuhaer Nayef Al-Dabagh

**Abstract**— in face recognition system, many of methods have good results if there were sufficient number of representative training samples per person. But, few of them give good results if only single training sample per person is available. In this paper, a face recognition system using local binary pattern (LBP) for pre-processing, Fisher's linear discriminant (FLD) for features extraction and support vector machine (SVM) for classification. These methods are proposed to solve the one training sample problem. The performance of the proposed method was evaluated on the Yale face database and the experimental results showed that these present method give good recognition rat.

**Index Terms**— Face recognition, local binary pattern; Fisher's linear discriminant; support vector machine.

## 1 INTRODUCTION

Face recognition from still images and video sequence has been an active research topic due to its scientific challenges and wide range of potential applications, such as biometric identity authentication, human-computer interaction, and video surveillance. The challenges of face recognition mainly come from the large variations in the visual stimulus due to illumination conditions, viewing directions, facial expressions, aging, and disguises. Within the past two decades, numerous face recognition methods have been proposed to deal with these challenging problems, as reviewed in the literature survey [1].

In the last decade, Fisher linear discriminant analysis (LDA) has been demonstrated to be a successful discriminant analysis algorithm in face recognition. It performs dimensionality reduction by trying to find a mapping from originally high-dimensional space to a low-dimensional space in which the most discriminant features are preserved. As LDA has been broadly applied and well-studied in recent years, a series of LDA algorithms have been developed, the most famous method of which is Fisherface .It uses a PCA plus LDA as a two-phase framework. Its recognition effectiveness has been widely proved [2]. The PCA approach, also known as eigenface method, is a popular unsupervised statistical technique that supports finding useful image representations. It also exhibits optimality when it comes to dimensionality reduction. However, the PCA is not ideal for classification purposes mainly because of the fact it retains unwanted variations occurring due to lighting and facial expression. There are numerous extensions to the standard PCA method. Meanwhile, the LDA method, also known as fisherface method, is a supervised learning approach whose functioning depends on class-specific information. This statistically motivated method maximizes the ratio of between-class scatter and within-class scatter and is also an example of a class-specific learning method. Again, there are various enhancements made to the LDA [3].

LBP-based facial image analysis has been one of the most popular and successful applications in recent years. Facial image analysis is an active research topic in computer vision, with a wide range of important applications, e.g., human–computer interaction, biometric identification, surveillance and security, and computer animation. LBP has been exploited for facial representation in different tasks, which include face detection, face recognition, facial expression analysis, demographic (gender, race, age, etc.) classification, and other related applications [4].

This kind of realistic "one sample per person problem" severely challenges existing face recognition techniques, especially their robustness performances under possible variations and has rapidly emerged as an active research sub-area in recent years. Although several methods have been proposed dealing with the one sample problem such as (PCA, FLDA and LBP) the variation issue is far from solved. Recent surveys of face recognition techniques employing one training image can be found in literatures [5].

Support vector machines (SVMs) provide efficient and powerful classification algorithms that are capable of dealing with high-dimensional input features and with theoretical bounds on the generalization error and sparseness of the solution provided by statistical learning theory. Classifiers based on SVMs have few free parameters requiring tuning, are simple to implement, and are trained through optimization of a convex quadratic cost function, which ensures the uniqueness of the SVM solution. Furthermore, SVM-based solutions are sparse in the training data and are defined only by the most "informative" training points [6].

In this paper, a face recognition system using LBP, FLD and SVM are applied to give solution for the single training sample problem. The rest of the paper is organized as follows: LBP, FLD and SVM are introduced in Section 2; Finally, Sections 3 and 4 present the experimental results, discussions and conclusions.

## 2 FACE RECOGNITION SYSTEM

The recognition system consists of three main stages: image pre-processing, features extraction and classification. In which LBP technique is used in pre-processing to improve the face image and FLD is used for extraction features from face image. Finally, the SVM is applied for classify the features that extract. Fig.1 Describe the block diagram of proposed recogni-

ISSN 2229-5518

tion system.

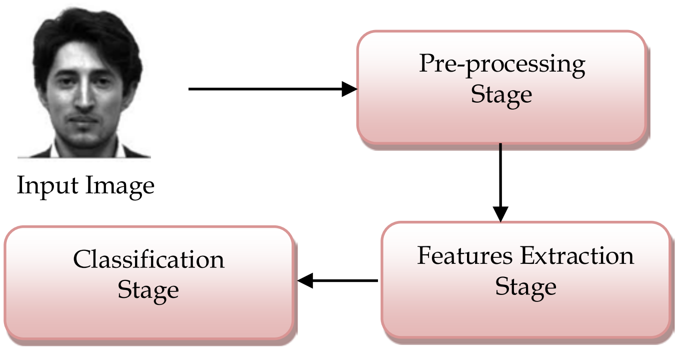


Fig 1: The block diagram of proposed recognition system

## 2.1 Pre-processing

Local binary pattern (LBP) is a popular technique used for image/face representation and classification. LBP has been widely applied in various applications due to its high discriminative power and tolerance against illumination changes such as texture analysis and object recognition. It was originally introduced by Ojala et al. [7] as gray-scale and rotation invariant texture classification. Basically, LBP is invariant to monotonic gray-scale transformations. The basic idea is that each 3x3-neighborhood in an image is threshold by the value of its center pixel and a decimal representation is then obtained by taking the binary sequence (Fig 2.) as a binary number such that LBP ∈[0, 255].

| | | |
|---|---|---|
| $i_0$ | $i_1$ | $i_2$ |
| $i_7$ | $i_c$ | $i_3$ |
| $i_6$ | $i_5$ | $i_4$ |

| | | |
|---|---|---|
| 1 | 2 | 4 |
| 128 | 0 | 8 |
| 64 | 32 | 16 |

Fig. 2. LBP operator: (left) the binary sequence (8 bits) and (right) the weighted threshold.

For each pixel, LBP accounts only for its relative relationship with its neighbors, while discarding the information of amplitude, and this makes the resulting LBP values very insensitive to illumination intensities. LBP is originally described as:

$$LBP(x_c, y_c) = \sum_{n=0}^{7} s(i_n - i_c) \cdot 2^n, \qquad (1)$$

• *Mustafa Zuhaer Nayef Al-Dabagh received the B.Eng. and M.Tech. degrees in computer Technology Engineering from Technical College of Mosul, Mosul, Iraq, in 2008 and 2013. E-mail: mustafazuhaer@yahoo.com. His current research interests include robotic and face analysis/recogni rion, image processing, and pattern recognition. Mr. Mustafa is a Member of the IEEE Computer Society.*

Where ic corresponds to the grey value of the centre pixel (xc, yc), in the gray values of the 8 surrounding pixels. s(x) is defined as:

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

The original LBP is later extended to be multi-scale LBP [11] which uses a circular neighborhood of different radius sizes using bilinearly interpolating. LBPP, R indicates P sampling pixels on a circle of radius of R. The example of multiscale LBP operator is illustrated in Fig. 2. An another extension called uniform patterns [8] which contain at most two bit-wise 0 to 1 or 1 to 0 transitions (circular binary code). For example the patterns 11111111 (0 transition), 00000110 (2 transitions), and 10000111 (2 transitions) are uniform whereas the pattern 11001001 (4 transitions) is not. These uniform LBPs represent the micro-features such as lines, edges and corners [9].

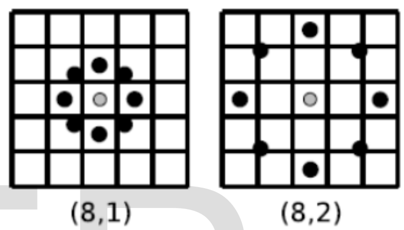


Fig. 3. The multi-scale LBP operator with (8,1) and (8,2) neighbourhoods. Pixel values are bilinearly interpolated for points which are not in the centre pixel.

## 2.2 Features Extraction

FLD is a popular discriminant criterion that measures the between-class scatter normalized by the within-class scatter [10]. Let $w_1$, $w_2$, ..., $w_L$ and $N_1$,$N_2$, ...,$N_L$ denote the classes and the number of images within each class, respectively. Let $M_1$,$M_2$, ...,$M_L$ *and* $M$ be the means of the classes and the grand mean. The within- and between class scatter matrices, ∑W and∑b, are defined as follows:

$$\Sigma_\omega = \sum_{i=1}^{L} P(\omega_i)\varepsilon\{(\gamma^\rho - M_i)(\gamma^\rho - M_i)^t|\omega_i\} \qquad (3)$$

And

$$\Sigma_b = \sum_{i=1}^{L} P(\omega_i)(M_i - M)(M_i - M)^t \qquad (4)$$

Where $P(\omega_i)$ is a priori probability, $\sum\omega,\sum b \in R^{mxm}$ , and L denotes the number of classes. FLD derives a projection matrix Ψ that maximizes the ratio $|\Psi^t \sum b\,\psi| \;/\; |\Psi^t \sum \omega\,\psi|$ [13]. This ratio is maximized when consists of the eigenvectors of the matrix $\sum_w^{-l} \sum_b$[11] :

$$\Sigma_\omega^{-l}\Sigma_b\Psi = \Psi\Delta \qquad (5)$$

ISSN 2229-5518

Where $\Psi,\ \Delta \in R^{mxm}$ are the eigenvector and eigenvalue matrices of $\sum_{\omega}^{-l}\sum b$, respectively. Concatenating 2D matrices into 1D vectors leads to very high dimensional nature of image vector, where it is difficult to evaluate the scatter matrices accurately due to its large size. Furthermore, the within-class scatter matrix is always singular, making the direct implementation of FLD algorithm an intractable task. In order to make FLD approach more efficient in this study, we have preceded the face images locally using a block-based SP feature extraction which renders small the size of the feature vectors [12].

### 2.3 Classification

In this stage of the system, Support vector machines which are one of famous classification methods are applied to find the best separating hyper-plane between features that belong to different classes. It may be applied to binary classification, using the v-SV procedure. Systematic analysis and discussion on SVM can be found in [13]. Consider points N that belong to two different classes:

$$\{(x_i, y_i)\}_{i=1}^{N} \quad \text{and} \quad y_i = \{+1, -1\} \tag{6}$$

Where $x_i$ is an -dimension vector and yi is the label of the class that the vector belongs to. SVM separates the two classes of points by a hyperplane:

$$w^T x + b = 0 \tag{7}$$

Where xi is an input vector, w is an adaptive weight vector, and b is a bias. The goal of SVM is to find the optimal separating hyperplane, to maximize the margin (i.e., the distance between the hyperplane and the closest point of both classes). By Lagrangian formulation, the prediction of the SVM is given by:

$$f_l(x) = \sum_{i=1}^{m} y_i \alpha_i \langle x, x_{si} \rangle + b \tag{8}$$

Where m is the number of support vectors, each xsi representing a support vector and α_i is the corresponding Lagrange multiplier. Each test vector is then classified by the sign of f(x). The solution can be extended to the case of nonlinear separating hyperplanes by a mapping of the input space into a high dimensional space x→Ø(x). The key property of this mapping is that the function is subject to the condition that the dot product of the two functions Φ(xi) • Φ(yi) can be rewritten as a kernel function . The decision function in (8) then becomes [14]:

$$f(x) = \sum_{i=1}^{m} y_i \alpha_i K(x, x_{si}) + b. \tag{9}$$

There are different types of SVM kernel functions, such as (Gaussian, linear, polynomials, Multi- Layer Perception, and Radial Basis Function) that can be applied. Some of these Kernels are defined in the equation below:

Linear kernel function:

$$K(x,x_j) = (x \bullet xj) \tag{10}$$

Polynomial kernel function:

$$K(x,x_j) = [(x \bullet xj)+1]^q \tag{11}$$

Radial Base Kernel function:

$$k(x, x_j) = \left\{-\frac{|x - x_j|^2}{\sigma^2}\right\} \tag{12}$$

In this paper compression between different SVM kernels function are applied.

## 3 Experimnets Result and Analysis

To evaluate the performed of the proposed method that using in this paper, experiments on Yale database [15] is used. This database is freely distributed on the Internet and contained 15 distinct subjects with 11 different images for each subject. For some subjects, the images were captured at various times, under different lighting conditions, facial details and facial expressions. All the images were taken with a white homogeneous background and its resolution of each image is 243 x 320 pixels with 265 gray levels per pixel. For example 11 sample of one person of Yale database are shown in Fig. 4.

Fig.4. Samples from the datasets

At the experiment, one sample of each subject is employed as test sample while the others are constructed the training set with different kernels functions. The experimental results proved show that linear kernel function gives higher result compare than Multilayer Perceptron, Quadratic, Polynomial and Gaussian Radial Basis Function (RBF). Table 1 shows the recognition accuracy among Linear, Polynomial, Radial Basis and Multi- Layer Perception (mlp) Function (RBF) SVMs. The degree d =3 in the case of the polynomial and the γ = 1 value in the case of the RBF kernel has been used in the experiment.

TABLE I
ACCURACY RATE OF DIFFERENT KERNEL

| Kernel type | Number of training sample | Number of training sample | Accuracy Rate (%) |
|---|---|---|---|
| Linear | 150 | 15 | 92.6667 |
| Quadratic | 150 | 15 | 84.667 |
| Polynomial | 150 | 15 | 72.333 |
| Gaussian Radial BasisFunction | 150 | 15 | 66 |

ISSN 2229-5518

| | | | |
|---|---|---|---|
| Multilayer Perceptron | 150 | 15 | 75 |

## 4 CONCLUSIONS

The single sample per person is the problem that most of face recognition systems often suffer from its and many of supervised learning methods fail to solve it. In this paper, combination between two methods (LBP and FLD) is proposed to deal with and overcome to this problem as well as used SVM to give good separated solutions. Experimental results on the Yale database shows the effectiveness of proposed method where the recognition rate reaches to 92.6667%.

## REFERENCES


[1] Yu Su, Shiguang Shan, Xilin Chen and Wen Gao, "Adaptive Generic Learning for Face Recognition from a Single Sample per Person", IEEE Computer Society Conference on Computer Vision and Pattern, pp. 2699-2706, 2010.

[2] Fei Ye, Zhiping Shi and Zhongzhi Shi," A Comparative Study of PCA, LDA and Kernel LDA for Image Classification", International Symposium on Ubiquitous Virtual Reality, pp.51-54, 2009.

[3] Kyu-Dae Ban, Keun-Chang Kwak, Su-Young Chi, and Yun-Koo Chung," Appearance-based Face Recognition from robot camera images with Illumination and Distance Variations", SICE-ICASE International Joint Conference, pp. 321-325,2006.

[4] Di Huang, Caifeng Shan, Mohsen Ardebilian, Yunhong Wang, and Liming Chen," Local Binary Patterns and Its Application to Facial Image Analysis: A Survey ",IEEE Transactions on Systems, Man, and Cybernetics, vol. 41, pp. 765-781, 2011.

[5] Guan-Chun Luh," Face Recognition Using PCA Based Immune Networks With Single Training Sample Per person", International Conference on Machine Learning and Cybernetics, vol. 4, pp. 1773-1779, 2011.

[6] Manuel Davy, Arthur Gretton, Arnaud Doucet, and Peter J. W. Rayner," Optimized Support Vector Machines for Nonstationary Signal Classification", IEEE Signal Processing Letters, vol. 9, pp. 442-445, 2002.

[7] T. Ojala, M. Pietik¨ainen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. Pattern Recognition, pp.51–59, 1996.

[8] T. Ojala, M. Pietik¨ainen, and T. M¨aenp¨a¨a. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, pp.971–987, 2002.

[9] Amnart Petpon and Sanun Srisuk," Face Recognition with Local Line Binary Pattern", Fifth International Conference on Image and Graphics, pp. 533-539, 2009.

[10] K. Fukunaga, Introduction to Statistical Pattern Recognition, 2nd ed. New York: Academic, 1991.

[11] D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," IEEE Transaction Pattern Analysis Machine Intelligence, vol. 18, p. 831836, 1996.

[12] Mohamed El Aroussi, Sanaa Ghouzali, Mohammed Rziza and Driss Aboutajdine," Face Recognition using enhanced Fisher Linear Discriminant", Fifth International Conference on Signal Image Technology and Internet Based Systems,pp. 48-53,2009.

[13] V. N. Vapnik, Statistical Learning Theory. New York: Wiley, 1998.

[14] Zhifeng Li and Xiaoou Tang, " Using Support Vector Machines to Enhance the Performance of Bayesian Face Recognition", IEEE Transactions on Information Forensics and Security, vol. 2, no. 2, June 2007.

[15] http://cvc.yale.edu/projects/yalefaces/yalefaces.html