

Face Recognition Using Two-Dimensional Principle Component Analysis And Neural Classifier

Ketan Patel, Dr. Hitesh Shah, Prof. Rahul Kher

PG Student, Professor, Associate Professor

Electronics and Communication Engineering Department

G.H.Patel College of Engineering & Technology, Vallabh Vidyanagar – 388 120

Abstract

With the growth of information technology coupled with the need for high security, the application of biometric as identification and recognition process has received special attention. The biometric authentication systems are gaining importance and in particular, face biometric is more preferred for person authentication because of its easy and non-intrusive method during acquisition procedure. Face recognition is considered to be one of the most reliable biometric, when security issues are taken into concern. Various methods are used for face recognition. To recognize the face, feature extraction becomes a critical problem. In this paper, a face recognition system with two-dimensional principle component analysis as feature Extraction method and Neural Network based classifier for Classification is presented.

1. Introduction

Biometrics or **biometric authentication** is the identification of humans by their characteristics or traits. Biometrics is used to describe a characteristics or a process. Various biometric traits can be utilized for the purpose of human recognition like fingerprint, palm print, hand geometry, iris, face, speech, gaits, signature, key strokes etc. The problem with finger print, iris palm print, speech, gaits are they need active co-operation of person while face recognition is a process does not require active co-operation of a person so without instructing the person one can recognize the person. So face recognition is much more advantageous compared to the other biometrics.

The face is the primary focus of attention in the society, playing a major role in conveying identity and emotion. Although the ability to infer intelligence or character from facial appearance is suspect, the human ability to recognize faces is remarkable. A human can recognize thousands of faces learned throughout the lifetime and identify familiar faces at a glance even after years of separation. This skill is quite robust,

despite of large changes in the visual stimulus due to viewing conditions, expression, aging, and distractions such as glasses, beards or changes in hair style. Face recognition has become an important issue in many applications such as security systems, credit card verification, criminal identification etc. Face recognition system consist of mainly two parts: feature extraction and classification. This paper presents a technique that utilizes two-dimensional principle component analysis and neural network based classification.

The most important problem in face recognition is the curse of dimensionality problem. The face image having very high dimension and it contain so much irrelevant or non-informative data and that makes much difficult for making decision. Feature extraction can act as a powerful dimension reduction agent. So, it is desirable to select smaller number of relevant and important feature with the help of dimension reduction techniques. High dimension also pose problem in computation, so it's also desirable to reduce the dimension. There are several methods are available for dimensionality reduction, that may be linear or non linear. We used linear method of feature extraction.

Principle Component Analysis (PCA), also known as Karhunen-Loeve expansion, is a classical feature extraction and data representation technique widely used in the areas of pattern recognition and computer vision. Sirovich and Kirby [1], [2] first used PCA to efficiently represent pictures of human faces. They argued that any face image could be reconstructed approximately as a weighted sum of a small collection of images that define a facial basis (eigen images), and a mean image of the face. Within this context, Turk and Pentland [3] presented the well-known Eigenfaces method for face recognition in 1991. Since then, PCA has been widely investigated and has become one of the most successful approaches in face recognition [4,

5, 6]. Sirovich also discussed the problem of the dimensionality of the “face space” when eigenfaces are used for Representation.

In the PCA-based face recognition technique, the 2D face image matrices must be previously transformed into 1D image vectors. The resulting image vectors of faces usually lead to a high dimensional image vector space, where it is difficult to evaluate the covariance matrix accurately due to its large size and the relatively small number of training samples. However, this does not imply that the eigenvectors can be evaluated accurately in this way since the eigenvectors are statistically determined by the covariance matrix, no matter what method is adopted for obtaining them. As opposed to conventional PCA, 2DPCA is based on 2D matrices rather than 1D vector. That is, the image matrix does not need to be previously transformed into a vector. Instead, an image covariance matrix can be constructed directly using the original image matrices. In contrast to the covariance matrix of PCA, the size of the image covariance matrix using 2DPCA is much smaller. As a result, 2DPCA has two important advantages over PCA. First, it is easier to evaluate the covariance matrix accurately. Second, less time is required to determine the corresponding eigenvectors. The remainder of this paper is organized as follows: In Section 2 2DPCA method has been discussed and in section 3 back propagation algorithm has been discussed; section 4 discusses our experiment and results. Final section discusses the conclusion.

2. Two- Dimensional Principle Component Analysis

Image representation and feature extraction are pervasive techniques that are commonly used for face recognition process. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant then the input data will be transformed into a reduced representation set of features. Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input.

Yang et al.[7] proposed a new approach- two-dimensional PCA for face recognition. We used same algorithm for feature extraction of face database. In two dimensional PCA, the image matrix A is transform on to the X matrix using linear transformation given by:

$$Y = AX \quad (1)$$

So, Y is a projected vector of image A. Y is also called as a projected feature vector. The total scatter of the projected samples can be characterized by the trace of the covariance matrix of the projected feature vectors. So the idea is to maximize the following:

$$J(X) = \text{tr}(S_x) \quad (2)$$

Where, S_x denotes the covariance matrix of the projected feature vectors of the training samples and $\text{tr}(S_x)$ denotes the trace of S_x . The physical significance of maximizing the criterion in Eq. (2) is to find a projection direction X, onto which all samples are projected, so that the total scatter of the resulting projected samples is maximized. The covariance matrix S_x is given by

$$\begin{aligned} S_x &= E(Y - EY) * E(Y - EY)^T \\ &= E[(A - EA)X] * E[(A - EA)X]^T \end{aligned} \quad (3)$$

Hence

$$J(x) = X^T E[(A - EA) * (A - EA)^T] X \quad (4)$$

For given a set of training image $A(1), A(2), \dots, A(M)$.

$$J(X) = X^T \left[\sum_{n=1}^M (A(i) - \bar{A})(A(i) - \bar{A})^T \right] X \quad (5)$$

Where \bar{A} is the average of the training images. Now, let

$$G_t = E \left[(A(i) - \bar{A})^T (A(i) - \bar{A}) \right] \quad (6)$$

The matrix G_t is called the image covariance (scatter) matrix. It is easy to verify that G_t is an $n * n$ nonnegative definite matrix from its definition, here G_t is computed directly from the training image. Suppose that there are M training image samples in total, the jth training image is denoted by an $m * n$ matrix $A_j(j=1,2,3, \dots, M)$ and the average image of all training samples is denoted by \bar{A} . Then, G_t can be evaluated by

$$G_t = \frac{1}{M} \sum_{j=1}^M (A(j) - \bar{A})^T (A(j) - \bar{A}) \quad (7)$$

Alternatively (2) can be written as

$$J(X) = X^T G_t X \quad (8)$$

Where X is a unitary column vector. This criterion is called the generalized total scatter criterion. The unitary vector X that maximizes the criterion is called the optimal projection axis. Intuitively, this means that the total scatter of the projected samples is maximized after the projection of an image matrix onto X. The optimal projection axis X_{opt} is the unitary vector that maximizes $J(X)$ i.e., the eigenvector of G_t corresponding to the largest eigenvalue. In general, it is not enough to have only one optimal projection axis. We usually need to select a set of projection axis, X_1 ;

... ; X_d , subject to the orthonormal constraints and maximizing the criterion $J(x)$, that is,

$$\{X_1, X_2, \dots, X_d\} = \arg \max (J(X)) \quad (9)$$

And

$$X_i^T X_j = 0 \quad ; i \neq j, i, j = 1, 2, \dots, d. \quad (10)$$

In fact, the optimal projection axis, $\{X_1, X_2, \dots, X_d\}$ are the orthonormal eigenvectors of G_t corresponding to the first d largest eigenvalues. The optimal projection vectors of 2DPCA, $\{X_1, X_2, \dots, X_d\}$ are used for feature extraction. For a given image sample A , let

$$Y_k = AX_k \quad (11)$$

Then, we obtain a family of projected feature vectors, $Y_1; \dots; Y_d$, which are called the principle component of the sample image A . The principle component vectors obtained are used to form an $m \times d$ matrix $B = [Y_1; \dots; Y_d]$, which is called the feature matrix or feature image of the image sample A . The feature vector generated is used to form a Feature matrix and same will be used for Classification.

3. Neural classifiers

Neural Network is a machine learning algorithm that has been used for various pattern classification problems such as gender classification, face recognition, and classification of facial expression. Neural Network classifier has advantages for classification such as incredible generalization and good learning ability.

A Neural Network is made up of neurons residing in various layers of network. These neurons of different layers are connected with each other via links and those links have some values called weights. These weights store the information. Basically the neural network is composed of 3 types of layers: first is Input layer, which is responsible for inserting the information to the network. Second is Hidden layer. It may consist of one or more layers as needed but it has been observed that one or two hidden layers are sufficient to solve difficult problems. The hidden layer is responsible for processing the data and training of the network. Last layer is the output layer which is used to give the network's output to a comparator which compares the output with predefined target value neural networks requires training. We give some input patterns for training and some target values and the weights of neural networks get adjusted. A Neural network is said to be good and efficient if it requires less training patterns, takes less time for Training and is able to recognize more unseen patterns.

We have used back propagation feed forward neural network (BPNN). Back propagation is a multi-layer feed forward, supervised learning network based on gradient descent learning rule [8]. This BPNN provides a computationally efficient method for changing the

weights in feed forward network, with differentiable activation function units, to learn a training set of input-output data. Being a gradient descent method it minimizes the total squared error of the output computed by the net. The aim is to train the network to achieve a balance between the ability to respond correctly to the input patterns that are used for training and the ability to provide good response to the input that are similar.

4. Experiments and Results

Our algorithm has been tested on well-known, publically available ORL database [7]. The database was generated in Olivetti Research laboratory in April 1992. The ORL or AT&T database contains 400 images of 40 person each person having 10 images. The images are of 112×92 pixels size and all are monochrome images. The face image having different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background. The images were taken with a tolerance for some tilting and rotation of the face of up to 20 degrees. Before applying our algorithm we have gone through pre-processing steps. The pre-processing step is required to increase efficiency and it can decrease the required time to execute algorithm and required storage memory.

4.1 Pre-processing step

In our experiments the input image has been resized. The image is of 112×92 pixel size in original and after pre-processing steps it is converted to 56×46 pixel size. The advantage of pre-processing is it consume less time for computation and increase speed of recognition, required less memory for storage.



Figure.1 Original Image and Resized Image.

Figure. 1 shows the pre-processing steps as our 112×92 pixel image has been converted into 56×46 pixel.

4.2 Feature Extraction

First, an experiment was performed using the first five image samples per class for training, and the remaining images for test. Thus, the total number of training samples and testing samples were both 200. The 2DPCA algorithm was first used for feature extraction. Here the covariance matrix size is 46×46 . We choose 3 eigen vector corresponding to 3 largest eigen value. After having 3 projection axis we project

our image on to the projection axis and we have obtained three principle component vectors. This has been used to form feature matrix. The size of the feature matrix is of 200*168. This matrix consists of feature matrix of 200 images.

4.3 Recognition

Feed Forward Back propagation Neural Network has been used for the classification. The neural network consists of 168 input node 10 hidden layer and 1 output layers.

The parameter set for experiments are

1. Training No of Epoch:-1000
2. Training Goal Error:-0.0001
3. Training Function:- TRAINLM
4. Adaption learning function:-LEARNGDM
5. Performance Function:-MSE
6. Number of Layers:-3
7. Numbers of Neuron:-10
8. Transfer Function:-TANSIG

In our first experiments we test our algorithm with different number of test image while keeping the number of feature for a single image is 168. Here 168 is the length of feature for a single image. From the Table 1 we can say that as the number of class (or number of image) in the algorithm increases the recognition rate is decreases. For 100 images we get 100% recognition rate. The database having total 400 images while applying the algorithm for all 400 image we are able to get up to 92.661% of recognition rate. While experimenting we divide our database in to two parts one is for training purpose and another for testing purpose. We are getting best result of 100% rate for 100 images of 10 classes. For 200 image 96.6% of recognition rate.

Table 1: Recognition Rate with different number of class

No of Class	No of train image	No of test image	Recognition Rate(%)
1	5	5	100
2	10	10	100
4	20	20	100
5	25	25	100
10	50	50	100
15	75	75	98.5
20	100	100	96.2
25	125	125	95.5
30	150	150	94.0
35	175	175	93.56
40	200	200	92.661

In our second experiments we use different length of feature vector for 400 images. From the Table 2 we can say that as the number of feature for a single image increases the recognition rate increases. We test our

algorithm with 23, 56, 112, 135 and 168 features. We are able to get maximum of 92.661% of recognition rate with 168 feature of a single image.

Table 2. 40 class with different size of feature vector

Feature Vector size	No of Class	No of Train Image	No of Test Image	Recognition Rate(%)
23	40	200	200	84.796
56	40	200	200	86.281
112	40	200	200	88.085
135	40	200	200	88.940
168	40	200	200	92.661

5. Conclusion

This paper presents a face recognition method based on 2DPCA and BPF neural networks. Facial features are extracted by the 2DPCA method, which reduces the dimensionality of the original face images. The extracted feature is given as input to the neural network classifiers.

Reference

- [1] L. Sirovich and M. Kirby, "Low-Dimensional Procedure for Characterization of Human Faces", *J. Optical Soc. Am.*, Vol. 4, pp. 519-524, 1987.
- [2] M. Kirby and L. Sirovich, "Application of the KL Procedure for the Characterization of Human Faces", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 12(1), pp. 103-108, 1990.
- [3] M. Turk and A. Pentland, "Eigenfaces for Recognition", *J. cognitive Neuroscience*, Vol. 3(1), pp. 71-86, 1991.
- [4] A. Pentland, "Looking at People: Sensing for Ubiquitous and Wearable Computing," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22(1), pp. 107-119, Jan. 2000.
- [5] M. A. Grudin, "On Internal Representations in Face Recognition Systems", *Pattern Recognition*, Vol. 33(7), pp. 1161-1177, 2000.
- [6] D. Valentin, H. Abdi, A.J. O'Toole, and G.W. Cottrell, "Connectionist Models of Face Processing: a Survey", *Pattern Recognition*, Vol. 27(9), pp. 1209-1230, 1994.
- [7] J. Yang, D. Zhang, A. F. Frangi, and J. Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Machine Intell*, Vol.26 (1), pp.131-137, 2004.
- [8] S.Lawrence, C.L.Giles, A.C.Tsoi, and A.d.Back, "Face recognition: A convolutional neural network approach", *IEEE Transactions of Neural Networks*, Vol.8 (1), pp.98-113., 1993.