

# **Design and Implementation of Resampling Techniques for Face Recognition using Classical LDA Algorithm in MATLAB**

**Madhavi R. Bichwe**  
Dept of Computer Science &  
Technology  
Kavikulguru Institute of  
Technology & Science,  
Ramtek, Maharashtra

**S. R Bichwe**  
Dept. of Electronics &  
Communication  
Kavikulguru Institute of  
Technology & Science,  
Ramtek, Maharashtra

**Sugandha Satija**  
Dept. of Information  
Technology  
Kavikulguru Institute of  
Technology & Science,  
Ramtek, Maharashtra

## **ABSTRACT**

A number of applications require robust human face recognition under varying environmental lighting conditions and different facial expressions, which considerably vary the appearance of human face. However, in many face recognition applications, only a small number of training samples for each subject are available; these samples are not able to capture all the facial appearance variations, to utilize the resampling techniques to generate several subsets of samples from the original training dataset. A classic appearance-based recognizer, LDA-based classifier, is applied to each of the generated subsets to construct a LDA representation for face recognition.

The discrimination power of various human facial features is studied and a new scheme for automatic face recognition (AFR) is proposed. This paper focuses on the linear discriminant analysis (LDA) of different aspects of human faces in the spatial as well as in the wavelet domain. The LDA of faces also provides us with a small set of features that carry the most relevant information for classification purposes. The features are obtained through eigenvector analysis of scatter matrices with the objective of maximizing between-class variations and minimizing within-class variations. The result is an efficient projection-based feature-extraction and classification scheme for AFR. Multisource data analysis are used to provide more reliable recognition results. For a medium-sized database of human faces, excellent classification accuracy is achieved with the use of very-low-dimensional feature vectors. Moreover, the method is applicable to many other image-recognition tasks.

## **Keywords**

Face recognition, feature extraction, linear discriminant analysis (LDA).

## **1. INTRODUCTION**

Human face recognition has been drawing a lot of attention in the past decade. A number of face recognition algorithms have been investigated and several commercial face recognition products are available. However, robust face recognition in unconstrained environments is still a very challenging problem. A face recognition system has two stages, training and test. In real applications, current face recognition systems encounter difficulties due to the small number of available training face images and complicated facial variations during the testing stage. Human face appearance has a lot of variations resulting from varying lighting conditions, different head poses and facial expressions. Among the face recognition algorithms,

appearance-based approaches [2], which utilize the intensity or intensity-derived features of original images, have been successfully developed [13]. The dimensionality of the feature vector used by these methods is often very high while the training sample size is relatively small. The classifier based on such training data may be biased and have a large variance, resulting in a poor performance [10]. To improve the performance of the weak classifiers, a number of approaches have been presented [4]. The training set is randomly resampled with replacement to generate independent bootstrap replicates. Finally, the majority voting is applied to integrate results of all the classifiers. The final decision is made by majority voting. In the random subspace method [8], classifiers are constructed in the random subspaces of the feature space. Simple majority voting is used as the final decision rule. Bolle et al. [3] used the bootstrap techniques for evaluating authentication systems.

A resampling technique is utilized to generate a number of subsets from the original training dataset. A classic appearance-based face recognizer based on the LDA representation is constructed on each of the generated subsets. Two integration strategies, majority voting and the sum rule, are used to combine the classification results to arrive at the final decision.

## **2. REVIEW OF LITERATURE**

### **2.1 History of Face Recognition**

Pioneers of Automated Facial Recognition include: Woody Bledsoe, Helen Chan Wolf, and Charles Bisson.

During 1964 and 1965, Bledsoe, along with Helen Chan and Charles Bisson, worked on using the computer to recognize human faces (Bledsoe 1966a, 1966b; Bledsoe and Chan 1965). Given a large database of images (in effect, a book of mug shots) and a photograph, the problem was to select from the database a small set of records such that one of the image records matched the photograph.

After Bledsoe left PRI in 1966, this work was continued at the Stanford Research Institute, primarily by Peter Hart. In experiments performed on a database of over 2000 photographs, the computer consistently outperformed humans when presented with the same recognition tasks (Bledsoe 1968). Peter Hart (1996) enthusiastically recalled the project with the exclamation, "It really worked!"

By about 1997, the system developed by Christoph von der Malsburg and graduate students of the University of Bochum in Germany and the University of Southern California in the United States outperformed most systems with those of

Massachusetts Institute of Technology and the University of Maryland rated next. The Bochum system was developed through funding by the United States Army Research Laboratory.

In 2006, the performances of the latest face recognition algorithms were evaluated in the Face Recognition Grand Challenge (FRGC). High-resolution face images, 3-D face scans, and iris images were used in the tests. The results indicated that the new algorithms are 10 times more accurate than the face recognition algorithms of 2002 and 100 times more accurate than those of 1995. Some of the algorithms were able to outperform human participants in recognizing faces and could uniquely identify identical twins. In about January 2007, image searches were "based on the text surrounding a photo," for example, if text nearby mentions the image content.

## 2.2 Face Recognition –Different Approaches

Generally, face identification technique can be considered as image based or feature based. The image based methods uses predefined standard face patterns where as feature based techniques concentrate on extracted features such as distance between eyes, skin color, eye socket depth etc. More specifically, face recognition techniques fall in three categories, holistic, feature based, model based and hybrid approaches.

### A) Holistic Approach

In holistic or appearance based approach, the whole face region is considered as the input data to the face recognition system. Examples are Eigen faces, fisher faces, probabilistic eigen face etc. These techniques help to lower the dimensions of the dataset without tampering the key characteristics. Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Independent Component Analysis (ICA) are the most widely used holistic methods.

#### i. Principal Component Analysis

Principal Component Analysis (PCA) is a technique to analyse and dimensionally reduce the data, highlighting their similarities and differences. The task of facial recognition is discriminating input signals (image data) into several classes (persons). The input signals are highly noisy (e.g. the noise is caused by differing lighting conditions, pose etc.), yet the input images are not completely random and in spite of their differences there are patterns which occur in any input signal (eyes, nose, mouth and the relative distance between these objects) These characteristic features are called eigen faces.

#### ii. Linear Discriminant Analysis

Linear Discriminant Analysis (Fisher's LDA) is efficient in encoding discriminatory information. The basic idea of LDA is grouping of similar classes of data where as PCA works directly on data. It seeks to find directions along which the classes are best separated. It does so by taking into consideration the scatter within-classes but also the scatter between classes. It is also more capable of distinguishing faces even in the presence of variations in images. Linear Discriminant Analysis easily handles the case where the within-class frequencies are unequal and their performances have been examined on randomly generated test data. This method maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability.

### iii. Independent Component Analysis

Independent Component Analysis is a method to transform the observed multidimensional vector into its components which are maintained statically as independent as possible. The fundamental restriction is that the independent components must be non-Gaussian for ICA to be possible. Independent Component Analysis (ICA) minimizes both second-order and higher-order dependencies in the input data and attempts to find the basis along which the data (when projected onto them) are *statistically independent*. Bartlett et al. provided two architectures of ICA for face recognition task: *Architecture I* - statistically independent basis images, and *Architecture II* - factorial code representation.

### B) Feature Based Approach

The heart of any feature based algorithm is the localization and extraction of features on the face. Elastic Bunch Graph Matching and Dynamic link methods belong to this category.

#### i. Elastic Bunch Graph Matching

Elastic bunch graph matching uses model graph as a general representation of human face and encodes local appearance using 'wavelet jets'. The facial features are captured and presented using a special stack like structure called face bunch graph[23]. All human faces share a similar topological structure. Faces are represented as graphs, with nodes positioned at fiducial points. (eyes, nose...) and edges labeled with 2-D distance vectors. Each node contains a set of 40 complex Gabor wavelet coefficients at different scales and orientations (phase, amplitude). They are called "jets". Recognition is based on labeled graphs. A labeled graph is a set of nodes connected by edges, nodes are labeled with jets, and edges are labeled with distances.

#### ii. Dynamic Link Matching

In dynamic link architecture, the images and all models are represented by layers of neuron, which are labeled by jets as local features. Jets are wavelet components describing grey level distribution. In short dynamic link structure encodes information using wavelet transformations. DLM establishes one-one mapping between initially all-to-all connected layers and hence reduces distortion. The process uses a winner-take-all strategy once a correct mapping is obtained to choose the apt model.

### C) Model-Based Methods

#### i. Active Appearance Model:

Active appearance models (AAM) are very useful in interpreting images of deformable objects and require a combination of statistical shape and texture models to form a combined appearance model. An Active Appearance Model (AAM) is an integrated statistical model which combines a model of shape variation with a model of the appearance variations in a shape-normalized frame [7]. An AAM contains a statistical model of the shape and gray-level appearance of the object of interest which can generalize to almost any valid example. Matching to an image involves finding model parameters which minimize the difference between the image and a synthesized model example projected into the image.

#### ii. 3D Morphable Model

Human face is a surface lying in the 3-D space intrinsically. Therefore the 3-D model should be better for representing faces, especially to handle facial variations, such as pose, illumination etc. Blantz et al. proposed a method based on a 3-D morphable face model that encodes shape and texture in terms of model parameters, and algorithm that recovers these parameters from a single image of a face. The goal of 3D morphable models is to create any arbitrary face in

correspondence to face images already present in the database. In most of the real time applications acquisition of training images under varying conditions is cumbersome. The emergence of 3D morphable models is a remedy to this problem since it reduces the number of training samples needed. A very significant issue in the implementation of 3D face modeling is the 3D dense correspondence problem. The 3D dense morphable face shape models [5] resolve this issue successfully using dense correspondence algorithm.

#### D) Hybrid Approach

Human perception system recognizes faces using local facial features and the whole face region information. The hybrid method is more akin to human perception system since it is influenced by both feature based and holistic methods. This approach is most effective and efficient in the presence of irrelevant data. The key factors affecting performance depend on the selected features and the techniques used to combine them[].

#### i. Hidden Markov Model Method

Hidden Markov Models (HMM) are a set of statistical models used to characterize the statistical properties of a signal. HMM consists of two interrelated processes: (1) an underlying, unobservable Markov chain with a finite number of states, a state transition probability matrix and an initial state probability distribution and (2) a set of probability density functions associated with each state. Hidden Markov model (HMM) is a promising method that works well for images with variations in lighting, facial expression, and orientation. To process images using HMM, the temporal or space sequences are to be considered. In simple terms HMM can be defined as set of finite states with associated probability distributions. Only the outcome is visible to the external user not the states and hence the name Hidden Markov Model.

### 3. PROPOSED APPROACH – RESAMPLING AND INTEGRATION

A resampling technique is utilized to generate a number of subsets from the original training dataset. A classic appearance-based face recognizer based on the LDA representation is constructed on each of the generated subsets.

#### 3.1 System Overview

Resampling-integration scheme is illustrated in Fig. 3.1.  $S_1$  to  $S_K$  are the subsets resampled from the original training dataset.  $C_1$  to  $C_K$  are classifiers trained using the corresponding subsets. Here,  $K$  is the total number of subsets. The training dataset contains a small number of sample face images. A number of subsets are generated by resampling the training set. Each subset  $S_i$  is used to train a classifier  $C_i$ . In the test stage, the test image is loaded into each component classifier. Two strategies, (i) simple majority voting, and (ii) the sum rule, are used to integrate the outputs of component classifiers; the classifier outputs can be either the classification labels or the matching scores. Currently, the face recognizer based on LDA representation is used as the component classifier, but this framework does not limit the component classifiers to be of the same type.

#### 3.2 LDA-based Face Classifier

A two-dimensional face image is considered as a vector, by concatenating each row (or column) of the image. Let  $X = (x_1; x_2; \dots; x_i; \dots; x_N)$  denote the data matrix, where  $N$  is the number of face images in the training set. Each  $x_i$  is a face vector of dimension  $n$ , concatenated from a  $p \times p$  face image, where  $n$  represents the total number of pixels in the face image and  $n = p \times p$ . The Linear Discriminant Analysis (LDA) representation is a linear transformation from the original image vector to a projection feature vector, i.e.

$$Y = W_{LDA}^T X,$$

where  $Y$  is the  $d \times N$  feature vector matrix,  $d$  is the dimension of the feature vector,  $d \leq n$  and  $LDA$  is the transformation matrix, derived by

$$W_{LDA} = \arg \max_W \frac{W^T S_B W}{W^T S_W W},$$

where  $S_B$  is the between-class scatter matrix and  $S_W$  is the within-class scatter matrix,

$$S_B = \sum_{i=1}^c N_i (x_i - m)(x_i - m)^T,$$

$$S_W = \sum_{i=1}^c \sum_{x_k \in X_i} (x_k - m_i)(x_k - m_i)^T.$$

In the above expression,  $N_i$  is the number of training samples in class  $i$ ;  $c$  is the number of distinct classes;  $m$  is the mean vector of all the samples, i.e.,  $m = \sum_{i=1}^c x_i$ ;  $m_i$  is the mean vector of samples belonging to class  $i$  and  $X_i$  represents the set of samples belonging to class  $i$ .

In the face recognition problem, if the within-class scatter matrix  $S_W$  is singular, due to the facts that the rank of  $S_W$  is at most  $(N - c)$  and the number of training samples is generally less than the dimensionality of the face image (number of pixels), PCA [19] can be used to reduce the dimensionality of the original face image space [2] prior to applying LDA.

LDA derives a low dimensional representation of a high dimensional face feature vector space. The face vector is projected by the transformation matrix  $W_{LDA}$ . The projection coefficients are used as the feature representation of each face image. The matching score between the test face image and the training image is calculated as the cosine value of the angle between their coefficients vectors. A larger matching score means a better match.

#### 3.3 Resampling

The resampling module generates a number of subsets from the original training set. A number of resampling methods have been proposed in the literatures. In the LDA based face recognition, both intra- and inter-class information (between-class scatter matrix and within-class scatter matrix)

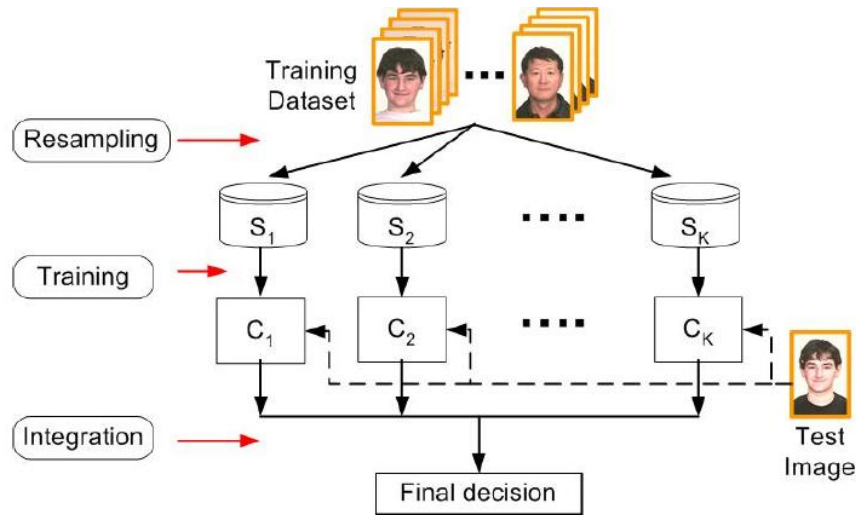


Fig. 3.1 The Resampling-Integration scheme for face recognition.

are utilized, so our sampling strategy does not randomly sample the whole training set, but does randomly sampling within each class (subject), subject to the following conditions:

1. The number of sample images for each subject in the subset is equal or as equal as possible.
2. Sampling within each class is achieved based on a uniform distribution.

The requirements listed above may not be the optimal ones, but work well as demonstrated by the empirical evaluation.

### 3.4 Integration

After resampling, several LDA-based classifiers are constructed. The matching scores between the test face image and the training images are computed by each component classifier. Let  $MS(i, j)$  be the matching score between the test image and the  $j^{th}$  training image, calculated by the  $i^{th}$  component classifier. For the  $i^{th}$  component classifier, the classification result for the test image is the subject label, denoted by  $Label(i)$ . This classification can be achieved by the nearest neighbor rule.

Two strategies for integration are applied, namely the simple majority voting and the sum rule.

1. Simple majority voting

Assign the test image with the label which appears most frequently in  $Label(i)$ ,

where  $i = 1, \dots, K$ .

2. The sum rule

Calculate  $MS_j = \sum_{i=1}^K MS(i, j)$ .

Assign the test image with the label of the  $J^{th}$  training image, such that

$$J = \arg \max_j MS_j.$$

The integration rules may not give desired results when the number of component classifiers ( $K$ ) is too small. But due to the resampling scheme presented, here  $K$  could be as large as needed.

## 4. RESULTS

### 4.1 Database description

The entire face database is divided into two parts. Nine images of each subject are used to construct the original training data and the remaining one is used for testing. This partition is repeated 10 different times so that every image of the subject can be used for testing. The recognition accuracy is the average of these ten different test sets. In resampling, 8 of 9 images for each subject are randomly selected according to the uniformly distributed seeds between 1 and 9. The sampling is without replacement. Each subject has different random generated seeds.



Fig 4.1 Face images in database

The component classifiers compute the cosine value of the angle between the two projection coefficients vectors (one from the test image and the other from the database image) as the matching score. Database image with the best match is used to determine the classification result of the input image from the component classifier. The recognition accuracy of different face recognition schemes is listed in table 1.

Table 1. Recognition accuracy (Number of resampled subsets, k=20)

Without resampling	Resampling + Majority Voting	Resampling + Sumrule
81.0%	88.7%	87.9%

The number of subsets,  $K$ , is decided empirically. In order to analyze the influence of  $K$  in our scheme, we conducted experiments with different settings of  $K$  values, from 1 to 20. Figure 4.2 demonstrates the recognition accuracy of the proposed recognition schemes as the number of subsets changes. These results show that the proposed resampling-integration scheme generally improves the performance of the LDA-based face classifier as  $K$  increases up to 20.

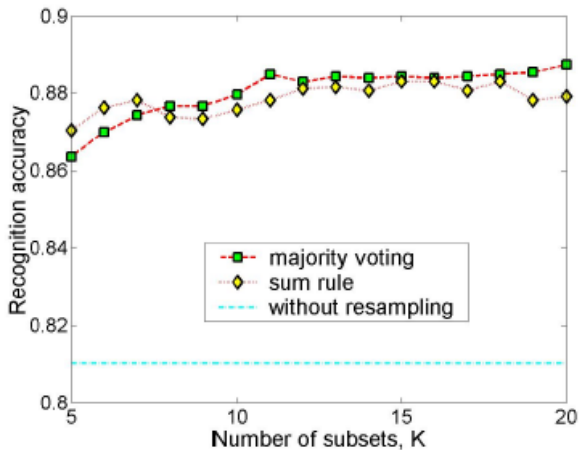


Fig. 4.2 Recognition accuracy with respect to the number of subsets

## 5. CONCLUSION

The resampling-integration scheme is proposed to improve the recognition accuracy of face classification. The resampling is designed to generate a number of subsets, which are used to train the component classifiers. The LDA-based face classifier is integrated into the scheme with the corresponding resampling design. The system framework is scalable in terms of the number of subsets, the type of component classifiers and resampling techniques. The recognition accuracy of the classical LDA-based face classifier is proposed.

## 6. FUTURE WORK

Different resampling techniques can be explored in this scheme. Since the subsets are resampled randomly, the resulting component classifiers may have different weights in the final decision. Some classifier selection technique can be applied. Although in this, all component classifiers are LDA based, the presented scheme does not limit the type of the component classifier. However, currently there is no guarantee that the proposed scheme always works for any type of classifiers. From the perspective of classifier combination, many other integration rules can be tried out.

## 7. REFERENCES

- [1] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman. Eigenfaces vs. Fisherfaces: "Recognition using class specific linear projection." IEEE Trans. Pattern Analysis and Machine Intelligence, 19(7):711-720, Jul. 1997
- [2] R. Bolle, N. Ratha, and S.Pankanti. Evaluating authentication systems using bootstrap confidence intervals. In proc. 1999 IEEE Workshop on Automatic Identification Advanced Technologies, pages 9-13, Morristown NJ, 1999.
- [3] Leo Breiman. Bagging predictors. Machine Learning, 24(2):123-140,1996.
- [4] R. O. Duda, P. E. Hart, and D. G. Stork. Pattern Classification. Wiley, New York. 2<sup>nd</sup> edition,2000.
- [5] Yoav Freund and Robert E. Schapire. Experiments with a new boosting algorithm. In Proc. International Conference on Machine Learning, pages 148-156,1996
- [6] Guo-Dong Guo and Hong-Jiang Zhang. Boosting for fast face recognition. In Proc. IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, pages 96-100, 2001.
- [7] T. K. Ho. The random subspace method for constructing decision forests. IEEE Trans. Pattern Analysis and Machine Intelligence, 20(8):832-844,1998.
- [8] A. K. Jain and B. Chandrasekaran. Dimensionality and sample size consideration in pattern recognition practice. Handbook of Statistics, P. R. Krishnaiah and L. N. Kanal (eds.),2:835-855,1987.
- [9] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. IEEE Trans. Pattern Analysis and Machine Intelligence, 20(3):226-239,1998.
- [10] P. Jonathon Phillips, Hyeonjoon Moon, Syed A Rizvi, and Patrick J. Rauss. The feret evaluation methodology for face-recognition algorithms IEEE Trans. Pattern Analysis and Machine Intelligence, 22(10):1090-1104,2000.
- [11] Ferdinando Samaria and Andy Harter. Parameterisation of a stochastic model for human face identification. In Proc. 2<sup>nd</sup> IEEE Workshop on Applications of Computer Vision, Sarasota FL, Dec. 1994.
- [12] R. E. Schapire and Y. Singer. Boostexter: A boosting-based System for text categorization. Machine Learning, 39(2-3):135-168,May/June 2000.
- [13] M. Skurichina and R.P.W. Duin. Bagging for linear classifiers. Pattern Recognition, 31(7):909-930, 1998.
- [14] M. Skurichina and R.P.W. Duin. Bagging, boosting and the random subspace method for linear classifiers. Pattern Analysis and Applications, 5(2):121-135,2002.
- [15] Feng Jiao, Wen Gao, Xilin Chen, Guoqin Cui, Shiguang Shan, "A Face Recognition Method Based on Local Feature Analysis", The 5th Asian Conference on Computer Vision, 23--25 January 2002.
- [16] Sanghoon Kim, Sun-Tae Chung, Souhwan Jung, Seoungseon Jeon, Jaemin Kim, Seongwon Cho, "Robust Face Recognition using AAM and Gabor Features", Proceedings Of World Academy Of Science, Engineering And Technology Volume 21 January 2007 Issn 1307-6884.
- [17] Yongsuk Jang Kim, Sun-Tae Chung, Boogyun Kim, Seongwon Cho, "3D Face Modeling based on 3D Dense Morphable Face Shape Model", International Journal of Computer Science and Engineering Volume 2 Number 3.