Appearance Based Facial Recognition System Using Dhmm with Linear Discriminant Analysis

Islam, Md. Rabiul

http://hdl.handle.net/20.500.11948/1487

Downloaded from http://dspace.library.daffodilvarsity.edu.bd, Copyright Daffodil International University Library
United States Technical University, Dhaka, Bangladesh

Abstract: The aim of this work is to enhance the performance of the face recognition system where appearance based facial features have been used. To locate and extract the facial features from the human face image, Active Shape Model (ASM) has been applied. To reduce the dimension of the feature vector, Linear Discriminant Analysis (LDA) based dimension reduction technique has been used. Finally, these reduced feature vectors are feed to the Discrete Hidden Markov Model (DHMM) using Baum-Welch algorithm and Baum’s Forward-Backward algorithm is used for the recognition model. Extended Yale face database has used to measure the performance of the proposed face recognition system with various dimensions. Finally proposed system performance has been reported according to LDA and existing PCA based facial feature reduction technique.

Keywords: Face Recognition, Appearance based Facial Feature, Active Shape Model, Linear Discriminant Analysis, Discrete Hidden Markov Model.

1. Introduction
Biometry is currently a very active area of research which is spanning several sub-disciplines such as image processing, pattern recognition, and computer vision. The main goal of biometry is to build systems that can identify people from some observable characteristics such as their face, fingerprints, iris, etc [1]. Face recognition is a dynamic biometric task where human can be identified by using facial image [2]. Among all of the biometric techniques, face recognition is one of the most popular methods of human identification. Face recognition has achieved mush popularity due to increasing demand in security and law enforcement applications.

Several excellent survey papers on face recognition techniques are available with a wide variety of methods [3, 4] that covers early face recognition approaches. While humans quickly and easily recognize faces under variable situations or even after several years of separation, the problem of machine face recognition is still a highly challenging task in pattern recognition and computer vision [5, 6].

Appearance and shape based face recognition are the two different approaches of face recognition system where appearance based face recognition is more popular and achieved great success [7]. Two dimensional facial image has been used to for the appearance based face recognition. High dimensional facial image has been used for the appearance based face recognition problem. But it is very difficult to process this high dimensional feature vector. As a result, it is required to reduce the dimensionality of the facial feature vector. Principal Component Analysis (PCA) has been used as the first reduction technique for face recognition problem [8] where PCA seeks for a set of projection vectors which project the image data into a subspace based on the variation in energy [7]. Eigenface method was introduced with PCA and performed enhanced result for face recognition [9]. Another well known method, fisherface was introduced into Linear Discriminant Analysis (LDA) to extract most important features with reduced form of face image. In general, LDA based methods outperform PCA based methods because LDA optimizes the low dimensional representation of face images with the focus on the most discriminant features extraction [7].
In this paper, we propose a method of face recognition where LDA based dimensionality recognition technique has been used to extract the most prominent features. Rabiul et. al. propose a technique of face recognition where appearance and shape base facial feature vectors are combined to enhance the recognition efficiency [10]. Here we propose a technique of appearance based face recognition which will further enhance the efficiency of the overall system performance.

2. Paradigm of the Proposed Face Recognition System

The overall system block diagram of the proposed work is shown in figure 1. At first facial images are captured from high resolution digital camera. Then appearance based facial features are extracted from various image pre-processing techniques which will be elaborated in section III. The dimensionality of extracted appearance based facial features is very large. To reduce the dimension of the extracted features, LDA based dimensionality reduction technique has been applied which is very effective to extract discriminant features with reduced form. Finally, these appearance based extracted reduced features are fed to the Discrete Hidden Markov Model (DHMM) based learning algorithm to create template. In classification stage, these learned templates are used for mapping with the unknown facial image to achieve the recognition performance.

3. Facial Image Pre-Processing and Feature Extraction

To extract the appearance based facial feature extraction, at first Stam’s Active Shape Model (ASM) [11] has used to detect the facial features from the face image. Figure 2(b) shows the detected face with facial features in red colors. Red colors are extracted in figure 2(c) as a binary image. Scaling has been done in figure 2(d) with the size of 200×200. Though appearance based facial feature is shown in figure 2(e), it has some noise between the region of check and neck. As a result, noises have been removed which is shown in figure 2(f).

4. Dimensionality Reduction of the Facial Feature Vector

After extracting the appearance based facial features, the 200×200 i.e., 40000 dimensional feature vector has produced which is very large to feed the classifier. As a result, Linear Discriminant Analysis technique has been imposed to reduce the dimension of the extracted facial feature vector. Though PCA can be used, LDA has been applied here to achieve the greater efficiency. Linear Discriminant Analysis can easily handles the case where the within-class frequencies are unequal. This method maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal classification. It has been decided to implement LDA in the expectation of providing better classification.
which is compared to Principal Components Analysis. The prime difference between LDA and PCA is that PCA does more of feature classification and LDA does data classification. In PCA, the shape and location of the original data sets changes when transformed to a different space whereas LDA doesn’t change the location but only tries to provide more class separability and draw a decision region between the given classes. In this proposed system, 40000 dimensional feature vector has been reduced to 200 dimension which is shown in figure 4.

The first step is to formulate the data sets and test sets, which are to be classified in the original space. The mean of each data set and mean of entire data set is to be calculated. After the mean image has been calculated, mean image is subtracted from each of those images of the training data set. Mean image is subtracted from each original image $F_i$ and stored in the variable $\Phi_i$. Each image in the data set differs from the average face by $t_i$, so, one scatter matrix is calculated for each person from its images.

$$\Phi_i = F_i - \mu$$  \hspace{1cm} (1)

Since LDA calculates the difference of features within all images of each person individually. So, one scatter matrix is calculated for each person from its images.

$$S_i = \sum_{j=1}^{L} F_j F_j^T$$  \hspace{1cm} (2)

Here,
- $S_i$ is the scatter matrix of $i^{th}$ person
- $L$ is the number of images of each person
- $F_j$ is the $j^{th}$ image of $i^{th}$ person

Summation of all scatter matrices is called within-class scatter matrix which represents variation among images of each persons.

Within-class scatter matrix,

$$S_w = \sum_{i=1}^{M} S_i$$  \hspace{1cm} (3)

Here,
- $M$ is the number of total persons
- $S_i$ is the $i^{th}$ scatter matrix

Between-class scatter matrix represents the variation among persons. For Between-class scatter matrix,

$$S_B = 2 \sum_{i=1}^{M} F_{mean} F_{mean}^T$$  \hspace{1cm} (4)

Here,
- $M$ is the number of total persons
- $S_i$ is the $i^{th}$ scatter matrix
- $F_{mean}$ represents mean image of $i^{th}$ person

Since LDA maximizes between-class scatter whereas minimizes the within-class scatter. To accomplish this, we must maximize $W$ matrix where,

$$J(W) = |W^T S_B W| / |W^T S_w W|$$ \hspace{1cm} (5)

From the matrix $W$, we will compute eigenvectors (Fisher vectors) which will represents linear discriminant features of each person. The steps to compute eigenvectors from $W$ matrix are given below:

1. Columns of $W$ are eigenvectors satisfying the equation,

$$S_i W = \lambda_i S_w W_i$$ \hspace{1cm} (6)

2. Eigenvalues are roots of the equation,

$$|S_w - \lambda_i S_w| = 0$$ \hspace{1cm} (7)

3. Calculation of eigenvectors by solving the equation,

$$(S_w - \lambda_i S_w) W = 0$$ \hspace{1cm} (8)

Eigenvectors of highest eigenvalues are selected and eigen vectors with lowest eigenvalues of the data set are ignored. Once eigenvectors are found, the next step is to order them by eigenvalue, highest to lowest. This gives the components in order of significance. Now those components having less eigenvalue can be ignored. If the eigenvalues are small, then it contains a less information about the data. To be precise, if original data have $n$ dimensions in data set and so, $n$ eigenvectors and eigenvalues are gained and then only the first $p$ eigenvectors are chosen then the final data set has only $p$ dimensions.

Now the feature vector is to be calculated. Taking the eigenvectors that we want to keep from the list of eigenvectors and forming a matrix with these eigenvectors in the columns construct this. At first eigenvectors are converted in column vector and then each of them are placed on a matrix in each row.

$$Feature \ vector = (eig_1 \ eig_2 \ eig_3 \ldots \ eig_p)$$ \hspace{1cm} (9)

Finally, we get the feature vector in reduced dimension which can be used in classification process. Linear Discriminant Analysis (LDA) searches for those vectors in the underlying
space that best discriminate among classes (rather than those that best describe the data). More formally, given a number of independent features $\theta$ relative to which the data is described, LDA creates a linear combination of these which yields the largest mean differences between the desired classes. Thus theoretically, LDA should give better performance than PCA [12, 13, 14] Figure 3 shows the process of LDA based dimension reduction technique of the proposed face recognition system.

![Figure 3: LDA based dimension reduction of the appearance based facial feature in the proposed system.](image)

5. DHMM based Learning and Recognition Model

In this proposed work, DHMM has been used for learning and recognition model for the face recognition system. Since DHMM can take only positive integer values as input, so it is required to transform the continuous valued features into discrete form. It has been performed by using vector quantization method. Vector quantization is a system for mapping a sequence of continuous or discrete vectors into a discrete codebook index.

In training phase, for each speaker $k$, an ergodic DHMM (Discrete HMM), $\theta_k$ has been built [15, 16, 17, 18]. The model parameters $(A, B, \theta)$ have been estimated to optimize the likelihood of the training set observation vector for the $k^{th}$ speaker by using Baum-Welch algorithm. The Baum-Welch re-estimation formula has been considered as follows [19]:

$$\Pi_{i} = \gamma_{1}(i)$$

$$a_{ij} = \frac{\sum_{t=1}^{T} \xi_{t}(i, j)}{\sum_{t=1}^{T} \gamma_{t}(i)}$$

where,

$$\bar{b}_{j}(k) = \frac{\sum_{t=1}^{T} \gamma_{t}(j)}{\sum_{t=1}^{T} \gamma_{t}(j)}$$

and $\gamma_{t}(i) = \sum_{j=1}^{N} \xi_{t}(i, j)$

In the testing phase, for each unknown face to be recognized, this procedure includes:

- Measurement of the observation sequence, $O = \{o_1, o_2, \ldots, o_N\}$, via a feature analysis of the person corresponding to a face.
- Transforming the continuous values of $O$ into integer values.
- Calculation of model likelihoods for all possible models, $P(O | \theta_k)$.
- Declaration of the face as $k^*$ person whose model likelihood is highest – that is,

$$k^* = \arg \max_{1 \leq k \leq K} \left[ P(O | \theta_k) \right]$$

In this proposed work, the probability computation step has been performed using the Baum’s Forward-Backward algorithm [19, 20].

6. Performance Analysis of the Proposed System

To measure the performance of the proposed system, the extended Yale face database [21] has been used which contains 5760 single light source images of 10 subjects each seen under 576 viewing conditions (9 poses x 64 illumination conditions). For every subject in a particular pose, an image with ambient (background) illumination was also captured. Hence, the total number of images is 5760+90=5850. The images in the database were captured using a purpose-built illumination rig. This rig is fitted with 64 computer controlled strobes. The 64 images of a subject in a particular pose were acquired at camera frame rate (30 frames/second) in about 2 seconds, so there is only small change in head pose and facial expression for those 64 (+1 ambient) images. The image with ambient
illumination was captured without a strobe going off.

Number of hidden state of DHMM affects the performance of the face recognition system. As a result, experiment is essential to select the point where the system can perform the highest accuracy. In the learning phase of DHMM, we have chosen the hidden states in the range from 5 to 25 with the difference of 2. The highest performance of 94% have been achieved at \( N_H = 20 \) which is shown in figure 4.

The proposed system performance has been counted with LDA based face recognition system. LDA based feature extraction technique is applied and Receiver Operating Characteristics (ROC) curve is populated where a trade off is made between security and user friendliness. Face recognition performance of the proposed system is shown in figure 5. A comparison has made of the proposed LDA based system with the PCA based system which was proposed by Rabiul et al. [10].

The proposed system perform better than PCA based feature for the DHMM based face recognition system. For example, at a FRR = 40%, the LDA based system FAR is 20% whereas PCA based system FAR is 30%. Therefore, this scenario represents that LDA based system can perform better than PCA based system.

7. Conclusion and Observation
This work shows the performance of the proposed appearance based face recognition with LDA based features. The results has been tested according to the existing PCA based face recognition and achieved promising result which can help further research and implementing practical applications. Though this system can not take any precautions for minimizing the environmental noises, it gives satisfactory results. Facial image preprocessing techniques may be introduced with this system to reduce the various lighting effect, different pose and angels which may be further work of this proposed system.

References
[1] Dominique Gin hac, Fan Yang, Xiaojian Liu, Jianwu Dang and Michel Paimdavonc, “Robust Face Recognition System Based on a Multi-Views Face Database”, in Recent Advances in Face Recognition, edited by Kresimir Delac, Mislav Grigic and Marian Stewart Bartlett, Published by In-The, 2008, pp. 3-38.


[14] Hongjun Jia and Aleix M. Martinez, “Support Vector Machines in Face Recognition with Occlusions”, The Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210, USA.


