Face Recognition Using Adaptive Margin Fisher's Criterion and Linear Discriminant Analysis (AMFC-LDA)

Marryam Murtaza, Muhammad Sharif, Mudassar Raza, and Jamal Hussain Shah
Department of Computer Science, COMSATS Institute of Information Technology, Pakistan

Abstract: *Selecting a low dimensional feature subspace from thousands of features is a key phenomenon for optimal classification. Linear Discriminant Analysis (LDA) is a basic well recognized supervised classifier that is effectively employed for classification. However, two problems arise in intra class during discriminant analysis. Firstly, in training phase the number of samples in intra class is smaller than the dimensionality of the sample which makes LDA unstable. The other is high computational cost due to redundant and irrelevant data points in intra class. An Adaptive Margin Fisher's Criterion Linear Discriminant Analysis (AMFC-LDA) is proposed that addresses these issues and overcomes the limitations of intra class problems. Small Sample Size (SSS) problem is resolved through modified Maximum Margin Criterion (MMC), which is a form of customized LDA and convex hull. Inter class is defined using LDA while intra class is formulated using quick hull respectively. Similarly, computational cost is reduced by reformulating within class scatter matrix through minimum Redundancy Maximum Relevance (mRMR) algorithm while preserving discriminant information. The proposed algorithm reveals encouraging performance. Finally, a comparison is made with existing approaches.*

Keywords: *AMFC, LDA, MMC, quick hull, mRMR.*

Received November 1, 2011; accepted May 22, 2012; published online January 29, 2013

1. Introduction

Face recognition is a multidisciplinary activity that contributes in many real world applications like in surveillance, criminal investigation, security, verification and authentication of a person [5]. Its applications gained much attention in recent years that flourished the locality of face recognition. With the evolution of face recognition, multiple issues seem to congregate like face expression, pose, occlusion and illumination variations [34]. However, in the presence of such issues, subspace classifiers selectively represent the features that minimize the processing area. Feature extraction plays a vital role to reduce the computational cost and progress the classification results because selecting a low dimensional feature subspace from bundle of features is very crucial for optimal classification. Wrong features selection degrades the performance of face recognition; even though superlative classifier may be used [15].

There are bunch of linear and non-linear classifiers that offer categorization between correlated and uncorrelated variables. Table 1 summarizes the algorithms that are exploited for feature extraction and classification. The two basic linear classification techniques are Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) [5].

Other classifiers like Independent Component Analysis (ICA), Support Vector Machine (SVM), Singular Value Decomposition (SVD) and Kernel Version Cassifiers like KPCA, KLDA, k-NN are Elastic Bunch Graph algorithms [11]. Active Appearance Model (AAM), Active Shape Model (ASM) and 3D morph able model based approaches are commonly used [11].

Table 1. Organization of classification algorithms.

Linear	Non-Linear	
PCA	Kernel versions	
LDA	2D	3D
ICA Others	Elastic bunch graph active appearance model	3D morph able model

LDA is the most important supervised batch classifier that can convert high dimensional input data to low dimensional feature subspace and outperforms simple PCA. LDA has a simple classification rule that maximizes the expanse of inter class variances and minimizes the ratio of intra class on the basis of calculating the dependencies between data points [13, 32].

In view of the fact that Small Sample Size (SSS) problem arises in within class when the amount of samples is smaller as compared to the dimensionality of the samples. Owing to small samples significant discriminative information is vanished subsequently we must consider this problem during training phase [3, 13]. SSS problem takes place because the intra class scatter matrix is singular. So, its direct inverse is not possible [9]. Another problem in LDA is the high computational cost in case of large amount of data [9]. These are the most important obstacles of classical LDA for face recognition.

In the last decade, lots of investigations were employed in order to solve 3S problem. Tian *et al*. [27] provided a method to classify images by changing S_w^{-1} with its pseudo inverse $S_w^{-1} = \Phi \Gamma^{-1} \Phi^{+}$ [27] where Φ are the Eigenvectors and Γ are the corresponding Eigenvalues which is computationally expensive. Zhao *et al.* [33] exploited both PCA and LDA and added a small matrix to the singular matrix that makes huge non-singular scatter matrix. Belhumeur *et al.* [2] primarily exploited PCA to convert into low dimensional subspace and then used LDA. The combination of two classifiers reduced the computational cost by eliminating the null space to minimize the SSS problem. However, due to the loss of null spaces, some important discriminative information may be lost.

Similarly, Chen *et al.* [4] encountered the same difficulty by finding an optimal solution to convert S_w^{-1} into zero matrices by projecting all image samples into the null space and then applied PCA which is equal to the Optimal Discriminant Vectors. Direct-LDA or D-LDA [10] is the direct method that discards the null space of S_b which holds zero discriminative information and used the discriminative information of S_w for classification. It has the same mechanism as PCA+LDA. Juwei *et al.* [12] used the regularized Fisher's Criterion. This method alters the regularization parameters and eradicates the null space of both S_w and S_b which is responsible to evaporate the important features that are helpful for classification.

Dai and Yuen [6] proposed three-parameter Regularized Discriminant Analysis (RDA), which though solves the 3S problem and works in the full space of S_w and S_b, which doesn't lose any significant information but it can increase the computational cost due to full features sub space. Generalized Singular Value Decomposition (GSVD) [32] is another development to diminish the inverse of intra class scatter matrix. The incremental LDA/GSVD processes the data in full space with low computational cost and memory utilization.

Similarly, Li *et al.* [16] proposed Maximum Margin Criterion (MMC) by focusing on various problems of LDA like singularity issue arrives in intra class and the need to find the inverse of singular matrix is diminished. Qiu and Wu [21] came across with discriminant information which is independent of nonsingular matrix due to nonparametric MMC and outperforms other related methods.

In [24, 29], the researchers have devised total bidirectional Laplacian matrix for between and within class where it is not necessary to map the original data on to the vector space and structural information is provided which is ignored in MMC.

On the other hand, Pannagadatta and Jebara in 2010 [20] revised the classification issues for large data sets. The proposed Relative Margin Machine (RMM) provides better classification distribution to the training examples.

Another major dilemma of simple LDA is the high computational cost. Many researchers exploit both PCA and LDA [2, 4, 33] in order to reduce their complexity. They applied the two way linear low dimensional feature classifiers for this purpose but still both techniques have their own complexity. The complexity of PCA is $O(m^3)$ [28], so it is right to say that complexity reduced for the test image but not during classification.

In this paper both aforementioned problems are addressed by decorating the intra class scatter matrix. Firstly, SSS problem is solved using modified MMC. The modified MMC equation is formulated with the help of LDA and Quick Hull. MMC is a blueprint for LDA that removes the headache of calculating the inverse of covariance matrix. Meanwhile, minimum Redundancy Maximum Relevance (mRMR) algorithm is used that eliminates the irrelevant variables from within class and selects highly correlated variables. It converts the sub space into smaller dimensional space in order to reduce the computational cost.

The rest of the paper is structured as follows: Section 2 provides the new Adaptive Margin Fisher's Criterion (AMFC-LDA). Experimental results are obtained in section 3. Comprehensive performance of proposed algorithm is evaluated in section 4. At the end, conclusion is formulated in section 5.

2. Proposed AMFC-LDA Algorithm

The notion of AMFC-LDA algorithm is to overcome the inadequacy of conventional LDA and MMC to make it possible to fight against the singularity of within class scatter matrix under the reasonable computational cost. As stated earlier that intra class scatter matrix of conventional LDA deals with two major problems:

1. SSS problem.
2. Boosted computational cost, so they will be discussed one by one.

This section effectively and efficiently reports the aforementioned problems using AMFC-LDA algorithm. To provide better understanding, Table 2 summarizes the notation used in this paper.

2.1. Small Sample Size Problem

LDA is a fundamental batch classifier that locates most favorable discriminant feature vectors in order to reduce high dimensional input data into low dimensional feature vectors. The performance of conventional LDA highly depends upon maximizing the Fisher's Criterion equation as in [14]:

$$J_{LDA}(w) = \frac{|w^T S_b w|}{|w^T S_w w|} \quad (1)$$

Table 2. Notations.

Notation	Description
C_i	Number of jth class
N_j	Number of jth samples in jth class
μ_i, μ_m	Class mean and overall mean
$data_{adj}$	Adjusted data at origin
S_w	Intra class scatter matrix
S_b	Inter class scatter matrix
$F'(\Psi)$	Transformed projected matrix
$\xi(x,y)$	Correlation between two variables
ϕ	Transformed subspace
Δ	Distance function

The corresponding projection matrix of Fisher's Criterion equation is the generalized form of $S_w^{-1}S_b$. However, the intra class scatter matrix is singular $|S_w=0|$ which means its inverse is not possible. This is the main difficulty of classical LDA in the computation of transformation matrix of largest eigenvalue matrix [9, 24]. 3S problem results in the degradation of conventional LDA due to small observable training sample size as compared to the dimensionality of feature vector [17]. In this paper, the main emphasis is to maximize the AMFC instead of isolated Fisher's Criterion (FC) and simple MMC to resolve aforementioned SSS catastrophe.

2.1.1. Fundamentals

The main ingredients for analyzing transformation matrix are the data set and test set in the original feature subspace. Let C is the class representation and $\{C_i|i=1,2,...,n_C\} \in R^D$ is the i^{th} class where n_C shows the total number of classes. Similarly, if N illustrates the samples then $\{N_{ij}|j=1,2,...,n_S\} \in R^D$ is the j^{th} sample in i^{th} class and n_S is the number of samples. The original data sets Set_{org} are formulated in the form of matrix as:

$$Set_{org}(C_i) = \begin{pmatrix} a_{11} & a_{12} & . & . & a_{1i} \\ a_{21} & a_{22} & . & . & a_{2i} \\ . & . & . & & . \\ . & . & & . & . \\ a_{j1} & a_{j2} & . & . & a_{ji} \end{pmatrix} \quad (2)$$

Note that D shows the higher dimensional original subspace and the ambition is to reduce D into d lower dimensional subspace such that $d<<D$ and provide linear classification across data. The performance of discriminant features of LDA highly depends on the differences of mean class. Hence, for largest variances in the data, so called Heteroscedasticity data compromises on the performance of LDA where it can lose the discriminative information in the class covariance differences [35]. For each $Set_{org}(C_i)$ compute the mean of each class $\{\mu_i|i=1,2,..., n_C\}$ as:

$$\mu_i = \frac{\sum_{i=1}^{n} x_i}{n}, \quad n > 0 \quad (3)$$

Where n is the total number of feature points in each class. The overall mean denoted by μ_m depicts overall scatter of the data and shows the behavior of variance in specified direction which can be estimated by the formula as:

$$\mu_m = \sum_{i=1}^{n_C} (p_i \mu_i) \quad (4)$$

p_i is the likelihood of entire data. The distribution of probability to each class is calculated as:

$$p_i = \frac{1}{x_j} \quad (5)$$

Provided that $max\ P_i=1$. The adjusted data $data_{adj}$ is found by the formula:

$$data_{adj} = x_i - \mu_m \quad (6)$$

At the moment, equation 6 adjusts the entire data with respect to origin. Note that at present the overall mean of adjusted data must be zero as: $\mu_m=0$.

2.1.2. Non-Singular AMFC Equation

MMC is a replica of conventional LDA which is also a supervised classification approach provided that it is more well-organized and robust learning algorithm as compared to PCA and LDA. MMC works like LDA but it is irrespective of singular matrix which increases the performance due to different objective functions [28]. Actually, MMC provides the maximum margin between to within class scatter matrix [7]. The maximized MMC equation is:

$$J(w)=w^T(S_b-S_w)w \quad (7)$$

In order to resolve 3S problem, AMFC equation is formulated from S_b obtained from LDA and $\left(S_w^{\alpha} + S_w^{\beta}\right)$ obtained from quick hull so, that the need to find the inverse of intra class scatter matrix $S_w^{-1}S_b$ is diminished as stated earlier and also it preserves the most discriminant information.

Before calculating the projection matrix, the scatter of the data is determined to examine the spread of data in the specified direction. Find the variance if the data is one dimensional 1D and covariance matrix for n dimensional $n*D$ [25]. To find the covariance matrix, the inter class scatter matrix S_b can be estimated by the formula [19]:

$$S_b = \sum_{i=1}^{n_C} P_i (\mu_i - \mu_m)(\mu_i - \mu_m)^T \quad (8)$$

Where P_i is the mix probability of class mean to overall mean. Similarly, the intra class scatter matrix is found by applying the quick hull by the equation 9 as shown below [31]:

$$Q_{conv}(S_w) = \left\{ \begin{array}{l} (S_w^{\alpha} + S_w^{\beta}) \mid , \\ \alpha = convex.point \\ \beta = between.convex.point \end{array} \right\} \tag{9}$$

Equation 9 finds the decision boundary within class features by calculating its most extreme points as shown in Figure 1. The main advantage of using quick hull is to preserve the discriminant information that is lost in LDA because it converts the data into low dimensional feature space whereas quick hull provides the decision boundary without any loss of data [31].

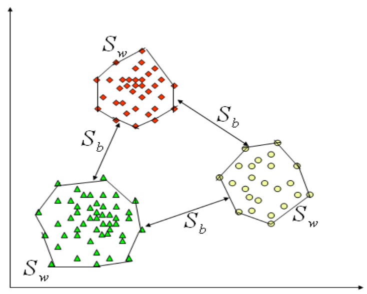

Figure 1. Feature classification between inter and intra class.

The exact value of the covariance matrix is not significant, hence their signs are important that demonstrate the relation between the data points [18, 25] as:

$$if \left\{ \begin{array}{lll} cov & = +ive & directrelation\ (+ivedirection) \\ cov & = -ive & indirectrelation\ (-ivedirection) \\ cov & = zero & uncorrelated \end{array} \right. \tag{10}$$

Equations 8 and 9 notify the stretch of intra and inter class of data that will be used to find AMFC. In fact, these equations actually present the modified fisher faces that are further used in the recognition process. In order to find Fisher's criterion following theorem must be proved:

Theorem: Let Ψ be the projection matrix of $Set_{org}\ (C_i)$ and the optimal projection matrix $\hat{\Psi}$ is the modified form of $J(w)$ in Equation 7 then:

$$argmax\left[\Psi^T(S_b - S_w)\Psi\right] \approx argmax\left[\Psi^T(S_b - (S_w^{\alpha} + S_w^{\beta})\Psi)\right] \tag{11}$$

Proof: The projection matrix with highest eigenvalues and eigenvectors is selected to transform into low dimensional space by equation 12:

$$F(\Psi) = argmax_{\Psi} \frac{\| \Psi^T S_b \Psi \|}{\| \Psi^T Q_{conv}(S_w) \Psi \|} \tag{12}$$

By the clarity of conventional LDA following criterion must be meeting:

$$criterion = Inv\left(cov_j\right) \times S_b \tag{13}$$

But, $\Psi = \left(S_w^{\alpha} + S_w^{\beta}\right)^{-1} S_b$, $\left(S_w^{\alpha} + S_w^{\beta}\right)^{-1}$ tends to formulate the SSS problem that has the generalized Eigenvector equation as:

$$S_b \Psi = \lambda \left(S_w^{\alpha} + S_w^{\beta}\right)^{-1} \Psi \tag{14}$$

Here, Ψ represents the Eigenvector and λ is the eigenvalue. Eigen decomposition formula segregates the eigenvalue from Eigenvector. In LDA, as equation 13 is unattainable due to singular matrix, that's why AMFC can be calculated by finding the optimal projection matrix $\hat{\Psi}$ as:

$$\hat{\Psi} = argmax \sum_{n=1}^{N} (\| \delta_n^b \|^2 - \| \delta_n^w \|^2) \tag{15}$$

Where δ_n^b and δ_n^w are between and within class differences which can be mathematically represented as:

$$\delta_n^b = \Psi^T \Delta^b \ where\ \Delta^b = \mu_i - \mu_m \tag{16}$$

$$\delta_n^w = \Psi^T \Delta^w \ where\ \Delta^w = S_w^{\alpha} + S_w^{\beta} \tag{17}$$

Now the modified optimal projection matrix is:

$$\begin{aligned} \hat{\Psi} &= \sum_{n=1}^{N} \left[\Psi^T (\mu_i - \mu_m)^T (\mu_i - \mu_m) \Psi \right] - \sum_{n=1}^{N} \left[\Psi^T (Q_{con}(S_w)) \Psi \right] \\ &= \sum_{n=1}^{N} \left[\Psi^T (\mu_i - \mu_m)(\mu_i - \mu_m)^T \Psi \right] - \sum_{n=1}^{N} \left[\Psi^T (Q_{con}(S_w)) \Psi \right] \\ &= \Psi^T \left[\sum_{n=1}^{N} (\mu_i - \mu_m)(\mu_i - \mu_m)^T \right] \Psi - \Psi^T \left[Q_{conv}(S_w) \right] \Psi \end{aligned}$$

Using equations 8 and 9 we have:

$$\begin{aligned} &= \Psi^T (S_b) \Psi - \Psi^T (S_w^{\alpha} + S_w^{\beta}) \Psi \\ &= \Psi^T \left\lfloor S_b - (S_w^{\alpha} + S_w^{\beta}) \right\rfloor \Psi \end{aligned} \tag{18}$$

Equation 18 is the modified form of equation 7 which is called AMFC equation. So, accordingly, the generalized Eigenvalues are shown in equation 14.

Algorithm 1: The AMFC Algorithm

Input: Unclassified data matrix $Set_{org}(Ci) \in R^D$
Output: Projection matrix $F(\Psi) \in R^d$
1. *Originate dataset and test set* $\in R^D$.
2. *Approximate class mean and overall mean using equations 3 and 4.*
3. *Calculate covariance matrix of equation 8 from LDA.*
4. *Calculate covariance matrix of equation 9 using Quick hull.*
5. *Find transformation matrix* $F(\Psi) \in R^d$ *from equation 18.*
6. *Select first* τ *columns of the projected matrix which have largest Eigenvalues.*

2.1.3. Intra Class Feature Selection with mRMR Algorithm

In order to reduce the computational cost of intra class, mRMR algorithm is applied to each class. mRMR algorithm works on correlated variables where small *m* is used for minimum redundancy that narrows down the processed area by eliminating those data points which were used multiple times. Correspondingly, large *M* is used to maximize the relevance across couple of data points where the mutual information shows the high correlation between points [1]. To formulate this hindrance, let within class C number of data points are represented as σ. The mutual information ξ between couple of variables is calculated by the formula [30]:

$$\xi(x,y)=\iint p(x,y)\log\frac{p(x,y)}{p(x)p(y)}dxdy \tag{19}$$

Where *p(x)* and *p(y)* are the probability distribution functions that show the chance of any variable to occur within the specified region and can be measured by taking its integral while *p(x,y)* are the combined probability of occurrence of two mutually correlated variables. By the definition of mRMR, maximum relevance and significance are achieved by estimating the largest mutual information that how two variables are strongly bounded and dependent to each other. The mathematical formulation of finding largest mutual information in such type of data is ruled by the equation:

$$\max D(\sigma,c),D=\frac{1}{|\sigma|}\sum_{x_i\in\sigma}\xi(x_i,c) \tag{20}$$

However, there are some algorithms that only focus to eliminate redundant data while some find common correlation between variables and some make attention towards both algorithms. mRMR is also a such type of method that locates the redundant features within class scatter matrix by the formula:

$$\min R(\sigma,c),R=\frac{1}{|\sigma|}\sum_{x_i,x_j\in\sigma}\xi(x_i,x_j) \tag{21}$$

By subtracting equation 20 from 21, we get the following mRMR equation:

$$\max \Phi(D,R),\Phi=D-R \tag{22}$$

By expanding equation 23, we have the following parameters:

$$\max_{x_i\in X-\sigma_{m-1}}\left[\xi(x_j,c)-\frac{1}{m-1}\sum_{x_i\in\sigma_{m-1}}\xi(x_j,x_i)\right] \tag{23}$$

Followed by the projection matrix, transformation is applied to the data set. The decision boundary is actually makeover to the single axis where n-dimensional data is transformed to convert into low dimensional feature space which can be obtained by the formula:

$$\phi=F'(\Psi)\times data_{adj} \tag{24}$$

Where ϕ shows the transformed low dimensional subspace. Finally, in order to provide the classification, pertinent distance function is used as:

$$\Delta=F(\Psi)'^{T}\times x-\mu_{ntrans} \tag{25}$$

Where Δ is the final distance function x is the test set while μ_{ntrans} is the mean of transformed data set.

Algorithm 2: Feature Selection using mRMR Algorithm

Input: Projected intra class matrix $S_w \in F(\Psi)$
Output: Classified dataset Δ
1. *Compute the largest mutual information using equation 20.*
2. *Compute overall mRMR parameters using equation 23.*
3. *Transformed dataset onto low dimensional subspace using equation 24.*
4. *Apply distance function to classify the dataset using equation 25.*

The overall flow chart of the system is depicted in Figure 2.

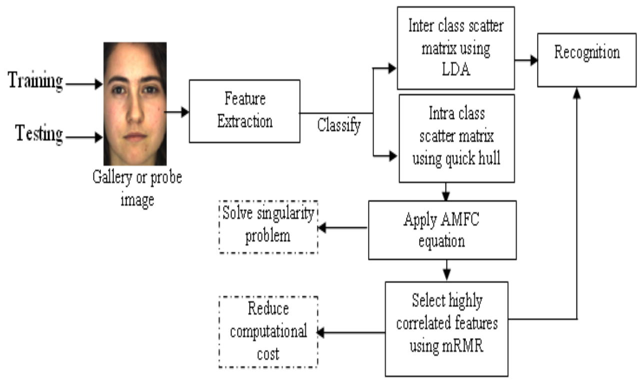

Figure 2. The overall flow of the system.

3. Experimental Results

In this section, the proposed technique is practically evaluated on different databases. The proposed algorithm is tested on the following trained databases i.e., AR, ORL extended Yale-B and CMU-PIE database. The images are taken under the controlled environment of different pose, lightning and expressions variation conditions. Primarily, images are tested on above mentioned four databases in order to check the recognition rate with different training datasets. Finally, a comparison is made against the proposed AMFC-LDA algorithm with Eigenfaces method PCA, Fisher's method LDA, Direct LDA method of combined PCA and LDA, Regularized Discriminant Analysis (RDA), Locality preserving projection LPP, simple maximum margin criterion MMC and Extended Linear Discriminant Analysis (ELDA) on Yale-B and CMU-PIE databases. These are the methods which are actually used to minimize the SSS problem in conventional LDA. The experimental outcome of proposed algorithm is tested on above mentioned databases which can be discussed as under.

3.1. Face Recognition Using AR Database

It contains 4,000 colored images with 126 peoples (70 males and 56 females) of two different sessions separated by two weeks. Each person has 26 different samples of 13 images per session. The dimension of each image is 120×165. The images are given in RAW and BMP file format.

In AR database, only facial expression images of neutral, smile, anger and scream are considered. The data set contains 10 images, 6 images are selected from first session in which 4 images contain expressions and 2 are occluded images while 4 expression images are selected from second session. So, total of 1260 (10×126) images are being used. From each subject, 10 samples are used in which a single image must be used as test image while remaining is used as trained subsets. Let Ω denotes the random trained subset that shows the number of training sets. The more the value of Ω, the more will be the recognition rate. For analyzing the proposed system, the results are taken with different training subsets (distinct value of Ω) and varying number of classes or subjects. The recognition rates using AR images are shown in Table 3.

In Table 3, the recognition rates are evaluated with different number of classes and training subsets. The entries provided in Table 3 show highest recognition rates for small number of classes and highest training subsets. It means that the recognition rate depends on inverse proportionality of number of classes and training sets. The right most entries show the average recognition rates of different number of subjects with number of training data. Figure 3 shows the graphical results using AR database. In Figure 3, the recognition rates are shown along y-axis and training subsets along x-axis. The recognition rate gets increased for large number of training data in which at least one image is used as test image.

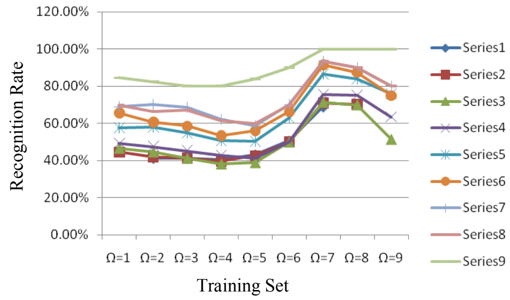

Figure 3. Recognition rates with different number of training data using AR database.

3.2. Face Recognition Using ORL Database

ORL database holds 400 images of 40 discrete subjects having 10 samples each. The images contain different lighting conditions, occlusions and facial expressions with dark setting environment. The file formats of the images are available in PGM with the dimensions of 92×112.

The dataset exploits all the images from ORL database. For each experiment, the number of subjects is minimized with the difference of five in order to analyze the recognition rate with different number of training images.

The same strategy is applied here as illustrated in the previous section. The main incentive is just to analyze the system with different number of subjects. The only difference is that a different type of database i.e., ORL database is used with different size of data. The variation is applied again to the number of subjects used and the size of training data. Table 4 illustrates different recognition rates with varying above mentioned parameters.

Table 3. Recognition rates using AR database with different training subsets.

No. of Classes	Training Set									Average
	Ω=1	Ω=2	Ω=3	Ω=4	Ω=5	Ω=6	Ω=7	Ω=8	Ω=9	
45	–	41.11%	40.95%	40.37%	43.11%	50.56%	68.89%	–	–	47.50%
40	44.44%	41.88%	41.07%	39.58%	42.50%	50.00%	70.83%	70.00%	–	50.04%
35	46.67%	44.64%	41.22%	38.10%	38.86%	50.00%	71.43%	70.00%	51.43%	50.26%
30	49.26%	47.50%	45.27%	42.78%	41.33%	50.00%	75.56%	75.00%	63.33%	54.45%
25	57.78%	58.00%	54.86%	50.67%	50.40%	63.00%	86.67%	84.00%	76.00%	64.60%
20	65.56%	60.63%	58.57%	53.33%	56.00%	66.25%	91.67%	87.50%	75.00%	68.28%
15	68.89%	70.00%	68.57%	62.22%	58.67%	70.00%	93.33%	90.00%	80.00%	73.52%
10	70.00%	66.25%	67.14%	61.67%	60.00%	70.00%	93.33%	90.00%	80.00%	73.15%
5	84.44%	82.50%	80.00%	80.00%	84.00%	90.00%	100.0%	100.0%	100.0%	88.99%

Table 4. Recognition rates using ORL database with different training subsets.

No of classes	Training Set									Average
	Ω=1	Ω=2	Ω=3	Ω=4	Ω=5	Ω=6	Ω=7	Ω=8	Ω=9	
40	62.22%	70.63%	72.86%	76.67%	78.00%	89.38%	91.67%	93.75%	95.00%	81.13%
35	65.40%	74.29%	76.33%	80.00%	80.57%	92.86%	94.29%	95.71%	97.14%	84.07%
30	66.30%	74.58%	79.05%	83.89%	85.33%	94.17%	95.56%	96.67%	96.67%	85.80%
25	68.89%	78.00%	83.43%	85.33%	86.40%	92.00%	93.33%	94.00%	96.00%	86.38%
20	65.56%	75.63%	85.00%	83.33%	85.00%	92.50%	93.33%	95.00%	95.00%	85.59%
15	76.30%	88.33%	88.57%	91.11%	90.67%	96.67%	97.78%	96.67%	93.33%	91.05%
10	85.56%	91.25%	95.71%	98.33%	98.00%	97.50%	100.0%	95.00%	90.00%	94.59%
5	91.11%	97.50%	97.14%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	98.42%

The same strategy is applied here as illustrated in the previous section. The main incentive is just to analyze the system with different number of subjects. The only difference is that a different type of database i.e., ORL database is used with different size of data. The variation is applied again to the number of subjects used and the size of training data. Table 4 illustrates different recognition rates with varying above mentioned parameters.

As shown from the above results, Table 4 also shows that recognition rate gets increased in right direction which means the more the value of *Ω* and less the number of classes; the more will be the recognition rate. Figure 4 shows that the system outperforms for larger training data.

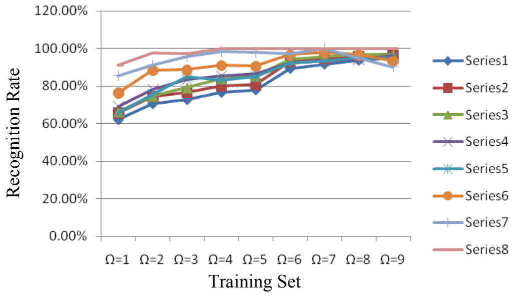

Figure 4. Recognition rates with distinct number of training data using ORL database.

3.3. Face Recognition Using Extended Yale-B Database

Extended Yale-B database holds 5760 images of 38 different persons. Each class contains 73 different samples of varying pose and lighting conditions (9 poses×64 illumination conditions). In our database twenty one samples of thirty eight subjects are used which approximately equals to 840 (21×40) images. The faces are cropped into 162×186 for recognition purpose. In each subject at least one image is used as test image while remaining is used for training out of 20 different samples. Let *Ω* denotes the random trained subset that shows the number of training sets. The more the value of *Ω*, the more will be the recognition rate. Table 5 depicts the comparison of recognition rate of different algorithms like PCA, LDA, D-LDA, RDA, LPP, MMC and ELDA with proposed AMFC-LDA.

Table 5. Comparasion on extended Yale-B database.

	Ω=5	Ω=10	Ω=15	Ω=20
PCA [9]	55.6±3.0%	61.1±1.9%	68.2±1.4%	73.5±1.2%
LDA [9]	71.2±1.9%	83.5±1.5%	86.3±1.2%	88.1±0.6%
D-LDA [9]	64.2±1.8%	73.6±0.9%	79.2±1.6%	80.0±1.2%
RDA [9]	64.8±2.1%	73.1±1.1%	78.7±1.3%	77.2±1.5%
LPP [9]	65.8±4.3%	78.4±3.9%	82.7±1.5%	83.2±1.9%
MMC [9]	65.4±2.5%	76.2±1.2%	79.1±1.8%	81.8±1.9%
ELDA [9]	72.1±2.3%	84.2±1.5%	87.5±1.7%	91.0±1.4%
AMFC-LDA	78.0±1.2%	89.38±1.5%	93.75±1.0%	95.0±1.5%

As depicted from Table 5, the error rate is added or subtracted from the recognition rate according to realistic results. The graphical results in Figure 5 demonstrate the clear cut performance of proposed system against different algorithms.

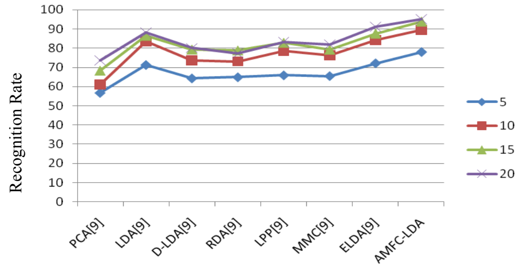

Figure 5. Graphical results of proposed system using Yale-B database.

3.4. Face Recognition Using CMU-PIE Database

In this section CMU-PIE database is chosen to examine the recognition rate of proposed system. CMU-PIE database contains a total of 41368 images with 68 subjects under varying lighting and pose conditions. Each subject of different samples is partitioned into 13 pose variations and 43 unusual lighting conditions. According to our study, we exploit 25 different subjects of 17 samples each which is the total of 425 (25×17) images.

The size of the face is cropped into 59×70. The same procedure is applied here that has been used for Yale-B database during recognition i.e., at least one image is used for testing while remaining images are used as trained images. Table 6 shows the results with different values of *Ω* as test images.

Table 6. Comparison on CMU-PIE database.

	Ω=4	Ω=8	Ω=12	Ω=16
PCA [9]	29.5±1.9%	37.2±0.9%	40.1±1.8%	45.3±1.9%
LDA [9]	39.0±1.4%	53.9±0.7%	60.1±1.2%	66.4±0.5%
D-LDA [9]	41.2±1.6%	56.1±0.8%	62.2±1.5%	64.0±1.2%
RDA [9]	42.7±1.5%	56.6±1.1%	62.5±1.4%	64.9±0.7%
LPP [9]	40.7±1.9%	52.1±2.4%	59.4±0.8%	65.1±1.9%
MMC [9]	34.6±3.9%	50.6±1.9%	56.9±1.5%	60.3±1.2%
ELDA [9]	51.8±1.8%	61.9±1.4%	68.3±1.1%	75.5±1.1%
AMFC-LDA	62.3±0.8%	65.9±1.2%	70.6±1.2%	80.9±1.3%

Graphical results of proposed system using CMU-PIE database are shown in Figure 6.

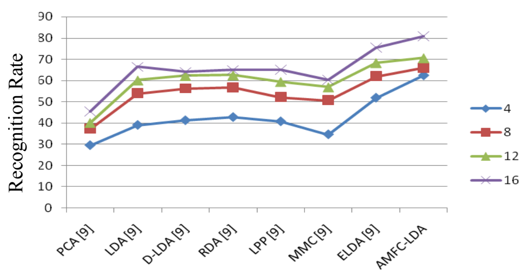

Figure 6. Graphical results of proposed system using CMU-PIE database.

4. Discussion and Performance Evaluation

The two basic linear subspace learning algorithms are PCA and LDA. As compared to PCA, LDA becomes more stable classification algorithm which provides discriminant information across the difference of classes [23]. PCA converts class feature data into low dimension by projecting the data onto a single space without classifying into classes unlike LDA. LDA offers simplified classification between classes to compute a Fisher face, that's why Fisher faces are more informative as compared to Eigenfaces [22].

Conventional LDA faces SSS problem within class matrix, so to avoid such dilemma, quick hull is used to determine the decision boundary across the data points that depicts the intra class scatter S_w. In fact, simple LDA loses some important discriminative information but Q_{conv} (S_w) finds the intra class without losing even a single point. Another main objective to use quick hull is to reduce the computational cost [31] that is very much high in MMC due to the calculation of neighboring data points. So, MMC is very time consuming during the training phase. In accordance with the face recognition performance, in this paper adaptive margin FC is applied in order to solve SSS problem appeared in conventional LDA. As already mentioned that although MMC works like LDA but MMC directly accessed the high dimensional input space which minimizes the loss of discriminative information unlike LDA due to which complexity increases as a consequence of large amount of processing data. However, the main participation of MMC in this paper is to achieve the modified MMC equation rather than using the whole MMC algorithm. The formulation of AMFC equation is depicted in the above theorem. Modified MMC handles the 3S problem very efficiently due to the change of objective function. As compared to PCA, it provides better separability between classes and gives accurate and effective classification results [8, 26]. Similarly, mRMR is used to reduce computational cost by assigning the highest priority to highly correlated data points and removing the redundancy [1, 30].

From the above experiments, it is observed that ORL provides better recognition rates as compared to the results obtained from AR images. Besides, the dimension size of ORL images is 92×112 which is less as compared to the dimension size of AR images i.e., 120×165. Additionally, AR images are colored images which are initially converted into grey level images in order to perform the experimental results. Moreover, AR database also, contains natural occluded images that increase the complexity and create obstruction during face recognition. From the results obtained from Yale-B and CMU-PIE databases, it is shown that AMFC-LDA performed the best as compared to other methods. It is also, observed that the proposed method chooses the best approximation projection which performs well over a range of different facial expressions and also, on unusual lighting conditions and pose variations. Hence, AMFC-LDA shows effective and accurate results under balanced computational cost which also reduces the 3S problem.

5. Conclusions

In this paper Adaptive Marin FC is proposed which is the modified form of fisher faces and MMC equation. The resultant AMFC is effective, efficient and has fast convergence rate that facilitates with robust face recognition. The false acceptance and false rejection rate is minimized because the proposed AMFC does not suffer from 3S problem in intra class and also, reduces the computational cost by most correlated feature selection and ignoring redundant variables on the basis of mRMR algorithm. mRMR also, prevents the important discriminative information being lost while linearly converting high dimensional input data into low dimensional feature subspace.

References

[1] Auffarth B., Lopez M., and Cerquides J., "Hopfield Networks in Relevance and Redundancy Feature Selection Applied to Classification of Biomedical High-Resolution Micro-CT Images," *in Proceedings of Spanish MEC Project*, Spain, pp. 16-31, 2008.

[2] Belhumeur P., Hespanha J., and Kriegman D., "Eigenfaces vs Fisher Faces: Recognition using Class Specific Linear Projection," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.

[3] Bland M., "Some Problems with Sample Size," *Presentation at the Joint Meeting of the Dutch Pathological Society and the Pathological Society of Great Britain & Ireland*, Leeds, pp. 1-6, 2008.

[4] Chen L-F., Mark H-Y., Ko M-T., Lin J-C., and Yu G-J., "A New LDA-Based Face Recognition System which Can Solve the Small Sample Size Problem," *Pattern Recognition*, vol. 33, no. 10, pp. 1713-1726, 2000.

[5] Chong L., Wanquan L., and Senjian A., "Face Recognition with Only One Training Sample," *in Proceedings of the 25th Chinese Control Conference*, Harbin, pp. 2215-2219, 2006.

[6] Dai D-Q. and Yuen P-C., "Face Recognition by Regularized Discriminant Analysis," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 37, no. 4, pp. 1080-1085, 2007.

[7] Gu Q. and Zhou J., "Local Relevance Weighted Maximum Margin Criterion for Text Classification," *in Proceedings of SIAM*

International Conference on Data Mining, vol. 21, pp. 1135-1146, 2009.

[8] Hariharan B., Manor L., Vishwanathan N., and Varma M., "Large Scale Max-Margin Multi-Label Classification with Priors," *in Proceedings of the 27th International Conference on Machine Learning*, Haifa, Palestine, pp. 423-430, 2010.

[9] Hu H., Zhang P., and Torre D., "Face Recognition using Enhanced Linear Discriminant Analysis," *IET Computer Vision*, vol. 4, no. 4, pp. 195-208, 2010.

[10] Hua Y. and Yang J., "A Direct LDA Algorithm for High-Dimensional Data with Application to Face Recognition," *Pattern Recognition*, vol. 34, no. 2001, pp. 2067-2070, 2001.

[11] Jiang Y. and Guo P., "Comparative Studies of Feature Extraction Methods with Application to Face Recognition," *in Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, Montreal, Canada, pp. 3627-3632, 2007.

[12] Juwei L., Plataniotis K., and Venetsanopoulos A., "Regularization Studies of Linear Discriminant Analysis in Small Sample Size Scenarios with Application to Face Recognition," *Elsevier Science Direct Pattern Recognition Letters*, vol. 26, no. 2, pp. 181-191, 2005.

[13] Kabir M., Jabid T., and Chae O., "Local Directional Pattern Variance (LDPv): A Robust Feature Descriptor for Facial Expression Recognition," *the International Arab Journal of Information Technology*, vol. 9, no. 4, pp. 382-391, 2012.

[14] Kalavdekar P., "Face Recognition using Extended Fisher Face with 3D Morph Able Model," *the International Journal of Computer Applications*, vol. 1, no. 16, pp. 344-523, 2010.

[15] Lajevardi S. and Hussain Z., "Novel Higher-Order Local Autocorrelation-Like Feature Extraction Methodology for Facial Expression Recognition," *IET Image Process*, vol. 4, no. 4 pp. 114-119, 2010.

[16] Li H., Jiang T., and Zhang K., "Efficient and Robust Feature Extraction by Maximum Margin Criterion," *IEEE Transaction on Neural Network*, vol. 17, no. 1, pp. 157-165, 2006.

[17] Li W., Ruan Q., and Wan J., "Two-Dimensional Uncorrelated Linear Discriminant Analysis for Facial Expression Recognition," *in Proceedings of IEEE 10th International Conference on Signal Processing*, Beijing, China, vol. 34, pp. 1362-1365, 2010.

[18] Liu C., "Learning the Uncorrelated, Independent, and Discriminating Color Spaces for Face Recognition," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 2, pp. 213-222, 2008.

[19] Lu H., Plataniotis K., Anastasios N., and Venetsanopoulos N., "Uncorrelated Multilinear Discriminant Analysis with Regularization and Aggregation for Tensor Object Recognition," *IEEE Transaction on Neural Networks*, vol. 20, no. 1, pp. 103-123, 2009.

[20] Pannagadatta K. and Jebara T., "Maximum Relative Margin and Data-Dependent Regularization," *Journal of Machine Learning Research*, vol. 11, pp. 747-788, 2010.

[21] Qiu X. and Wu L., "Nonparametric Maximum Margin Criterion for Face Recognition," *IEEE CiteSeer*, vol. 25, no. 7, pp. 1771-1782, 2005.

[22] Razzak M., Khan M., Alghathbar K., and Yousaf R., "Face Recognition using Layered Linear Discriminant Analysis and Small Subspace," *in Proceedings of the 10th IEEE International Conference on Computer and Information Technology*, Bradford, UK, pp. 1407-1412, 2010.

[23] Robinson P. and Clarke W., "Comparison of Principal Component Analysis and Linear Discriminant Analysis for Face Recognition," *in Proceedings of AFRICON*, Windhoek, Namibia, pp. 1-6, 2007.

[24] Sharif M., Mohsin S., Javed M., and Ali M., "Single Image Face Recognition using Laplacian of Gaussian and Discrete Cosine Transforms," *the International Arab Journal of Information Technology*, vol. 9, no. 6, pp. 562-570, 2012.

[25] Smith L., "A Tutorial on Principal Components Analysis," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 2-8, 2002.

[26] Tan X., Chen S., Zhou Z-H., and Liu J., "Face Recognition under Occlusions and Variant Expressions with Partial Similarity," *IEEE Transaction on Information Forensics and Security*, vol. 4, no. 2, pp. 421-426, 2009.

[27] Tian Q., Barbero M., Gu Z., and Lee S., "Image Classification by the Foley-Sammon Transform," *Optical Engineering*, vol. 25, no. 7, pp. 834-840, 1986.

[28] Yan J., Zhang B., Yan S., Yang Q., Li H., Chen Z., Xi W., Fan W., Ma W., and Cheng Q., "IMMC: Incremental Maximum Margin Criterion," *in Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, USA, vol. 1, pp. 725-730, 2004.

[29] Yang W., Wang J., Ren M., Yang J., Zhang L., and Liu G., "Feature Extraction Based on Laplacian Bidirectional Maximum Margin Criterion," *Pattern Recognition*, vol. 42, no. 11, pp. 2327-2334, 2009.

[30] Yang X., Tek B., Beddoe G., and Slabaugh G., "Feature Selection for Computer-Aided Polyp Detection using MRMR," *in Proceedings of Medical Imaging, Computer-Aided Diagnosis*, San Diego, USA, vol. 7624, pp. 1-8, 2010.

[31] Zhang X. and Tang Z., "A Fast Convex Hull Algorithm for Binary Image," *Informatica Journal*, *Guangxi*, vol. 34, pp. 369-376, 2010.
[32] Zhao H. and Yuen P-C., "Incremental Linear Discriminant Analysis for Face Recognition," *IEEE Transactions on Systems*, *Man*, *and Cybernetics*, *Part B: Cybernetics*, vol. 38, no. 1, pp. 210-221, 2008.
[33] Zhao W., Karishnaswamy A., Chellappa R., Swets D., and Weng J., "Discriminant Analysis of Principal Components for Face Recognition," *in Proceedings of the 3rd IEEE International Conference Automatic Face and Gesture Recognition*, Nara, Japan, vol. 83, pp. 336-341, 1998.
[34] Zhao W. and Chellappa R., "Image Based Face Recognition Issues and Methods," *in Proceedings of Image Recognition and Classification*, USA, pp. 375-402, 2002.
[35] Zheng W., "Heteroscedastic Feature Extraction for Texture Classification," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 766-769, 2009.

Marryam Murtaza is a lecturer at University of Wah, Pakistan. She has more than four years teaching and research experience. Her research interests include digital image processing and software engineering. She had completed her MS degree in computer science from CIIT Wah in 2011.

Muhammad Sharif is working as an assistant professor at the Department of Computer Science, COMSATS Institute of Information Technology, Pakistan. He is a PhD scholar at COMSATS Institute of Information Technology, Islamabad Campus. He has more than 16 years of experience of teaching undergraduate and graduate classes.

Mudassar Raza is a lecturer at COMSATS Institute of Information Technology, Pakistan. He has more than four years of experience of teaching undergraduate classes at CIIT Wah. He has also been supervising final year projects to undergraduate students. His areas of interest are digital image processing, and parallel and distributed computing.

Jamal Hussain Shah is a research associate in Computer Science Department at COMSATS Institute of Information Technology, Pakistan. His research areas are digital image processing and networking. He has more than 3 years experience in IT-related projects, he developed and designed ERP systems for different organizations of Pakistan.