Is Gender Classification Across Ethnicity Feasible using Discriminant Functions?

Tejas I. Dhamecha, Anush Sankaran, Richa Singh and Mayank Vatsa
IIIT Delhi, India
{tejasd, anushs, rsingh, mayank}@iiitd.ac.in

Abstract

Over the years, automatic gender recognition has been used in many applications. However, limited research has been done on analyzing gender recognition across ethnicity scenario. This research aims at studying the performance of discriminant functions including Principal Component Analysis, Linear Discriminant Analysis and Subclass Discriminant Analysis with the availability of limited training database and unseen ethnicity variations. The experiments are performed on a heterogeneous database of 8112 images that includes variations in illumination, expression, minor pose and ethnicity. Contrary to existing literature, the results show that PCA provides comparable but slightly better performance compared to PCA+LDA, PCA+SDA and PCA+SVM. The results also suggest that linear discriminant functions provide good generalization capability even with limited number of training samples, principal components and with cross-ethnicity variations.

1. Introduction

Gender recognition is an interesting problem that can be used to boost the performance of several important applications such as face recognition and video surveillance. There are several large scale applications such as US VISIT and India's UID project that store face images but do not use it for identification due to the limited performance of face recognition algorithms in large scale systems. However, gender classification can be utilized as an indexing technique to reduce the search space for automatic and manual face recognition. Further, other areas such as human computer interaction also have many interesting applications ranging from automatically identifying gender of individuals to image search over the internet.

Researchers have performed gender classification using several different cues such as face image, audio, and 3D face data. However, in this paper, we focus only on face image based gender classification. Different approaches, broadly subspace and point-based, have been proposed by the researchers. Matta *et al*. [9] combined temporal and spatial information such as head motion, mouth motion, and facial appearance to perform gender classification. The algorithm was evaluated on Italian TV speakers database that contains 208 images from 13 subjects. With 50% images used for training and the remaining 50% for testing, an accuracy of 99% was obtained. Moghaddam and Yang [10] proposed an algorithm using low resolution faces images, Support Vector Machine (SVM), and Gaussian kernel. The experiments were performed on around 1800 mugshot images of the FERET database [12] and reported 3.6% error with 80% training and 20% testing. Castrillon-Santana and Vuong [5] compared the performance of humans and automatic face recognition algorithms for gender classification. The automatic face recognition algorithm used Principal Component Analysis (PCA) [8] and SVM for feature extraction and classification. 7,000 images were used to build the PCA space and 50% of these were used for training and 50% for testing. The authors reported an accuracy of 70-75% for both humans and algorithms. Using Laplace and Gabor filters, Scalzo *et al.* [13] proposed an evolutionary genetic learning algorithm based framework to unify feature fusion and decision fusion. The performance of the algorithm was computed on a database of 400 frontal images and the results showed an error rate of 3.8%. Baluja and Rowley [2] proposed Adaboost with pixel comparison for gender classification. On frontal images from the FERET [12] database and 80%-20% non-overlapping train-test partitioning, the algorithm showed the maximum accuracy of 93%. Locally linear embedding algorithm based gender classification was proposed in [6] where low dimensional structure manifolds were learnt. Yishi *et al*. [18] studied the problem of gender classification across age variations using different variable selection and dimensionality reduction methods. Bekios-Calfa *et al.* [3] revisted linear discriminant techniques for gender classification. The paper compared the performance of Linear Discriminant Analysis (LDA), PCA+LDA, and Independent Component Analysis (ICA) + LDA along with SVM and Adaboost-based approaches. The results showed that for single database testing, SVM [17] and Adaboost yield best results followed by linear discriminant analysis approaches whereas for cross database tests, PCA+LDA provided the best results and SVM was the lowest.

1.1. Research Contribution

In this global era, a gender recognition system can have enrollees from all around the world, therefore it is important that the algorithm is robust to ethnicity variations. However, existing research has not analyzed the performance of gender recognition with ethnicity variations. Further, the experiments are generally performed with heavily trained classifiers that have seen the subject at least once. With such a setup, it is difficult to evaluate the generalization capability of the algorithm. *This research, therefore, aims at analyzing the performance of discriminant analysis techniques including PCA, PCA+LDA [8], PCA+Subclass Discriminant Analysis (SDA) [19], and PCA+SVM for gender classification on a database that comprises of images pertaining to different ethnicity*. The key contributions of this research are:

- A detailed analysis of three discriminant functions (PCA, LDA and SDA) with respect to the size of training data and number of principal components is presented. PCA and LDA have already been explored in literature but to the best of our knowledge, no work has been done with SDA for gender classification.
- Most of the existing research has focused on training samples from one database only. In this research, we have prepared a labeled database of 8112 images by combining images from the publicly available face databases. The database includes images of different ethnicity and nationalities such as Asian (Indian and Chinese), Caucasian, and African-American. A major portion of this database will be made available to the researchers.
- Experimental evaluation with cross database and cross ethnicity variations to evaluate the generalization capability of different discriminant functions and classifiers.

2. Discriminant Approaches for Gender Recognition

As mentioned previously, this research analyzes discriminant functions for gender recognition with ethnicity variations. The classification is formulated as a two class problem with the classes being *male* and *female*.

2.1. PCA-based Approach

As shown in Figure 1, PCA is used for feature extraction and Bayes' classification is used for classification. Let $\mathbf{x}$ be the $W \times H$ dimensional input image. As shown in *Algorithm 1*, PCA features are obtained by applying $\mathbf{y} = \mathbf{A}\mathbf{x}$ where the rows of $\mathbf{A}$ are the principal components. These features are further classified using Bayes' classification with Gaussian assumption (details are explained in Section 2.4).

2.2. PCA+LDA-based Approach

Bekios-Calfa *et al.* [3] have shown that if the selection of number of principal features is proper, then PCA+LDA can outperform LDA for gender recognition. PCA+LDA can be considered as applying LDA in the subspace spanned by the principal components. In this approach, PCA is used for dimensionality reduction followed by LDA for computing discriminant information. A $W \times H$ dimensional image is provided as input to compute the corresponding M dimensional representation. Top N eigenvectors are input to LDA which provides the d directions as the output. Not all principal components are always used to create the subspace because in this case the effect of applying LDA will be same as applying LDA in the original input feature subspace [8]. The mean and variance of both the classes (male and female) are computed from the training data and are considered as the estimates of Gaussian probability distribution for Bayes' classification. PCA+LDA approach is illustrated in Figure 1 and *Algorithm 2*.

2.3. PCA+SDA based Approach

Being a discriminant analysis technique, SDA follows the Fisher Rao's criterion which is defined as $\frac{|V^T AV|}{|V^T BV|}$, where V is the discriminant feature vector, A represents the between-class scatter matrix and B is the within-class scatter matrix. SDA, as the name suggests, captures the subclasses present within a class. B is defined as

$$\Sigma_B = \sum_{i=1}^{C-1} \sum_{j=1}^{H_i} \sum_{k=i+1}^{C} \sum_{l=1}^{H_k} p_{ij} p_{kl} (\mu_{ij} - \mu_{kl})(\mu_{ij} - \mu_{kl})^T \tag{1}$$

where $p_{ij} = \frac{n_{ij}}{n}$ is the prior probability of the j^{th} subclass of the i^{th} class, C is the number of classes, H_i is the number of subclasses in the i^{th} class, n is the total number of samples, and n_{ij} is the number of samples in the j^{th} subclass of the i^{th} class.

As explained in *Algorithm 3*, PCA+SDA performs subclass discriminant analysis on the PCA subspace. Top N eigenvectors are provided as input to SDA which then finds the subclasses and optimal projection directions. Similar to the other two approaches, Bayes' classifier with Gaussian distribution is used for classification.

2.4. Bayes' Classification

Let $D = \{d_1, d_2, ...d_n\}$ be the sample set in which the samples belong to either class C_1 (male) or C_2 (female) and $D_{train} \subset D$ be the labeled training set. Let

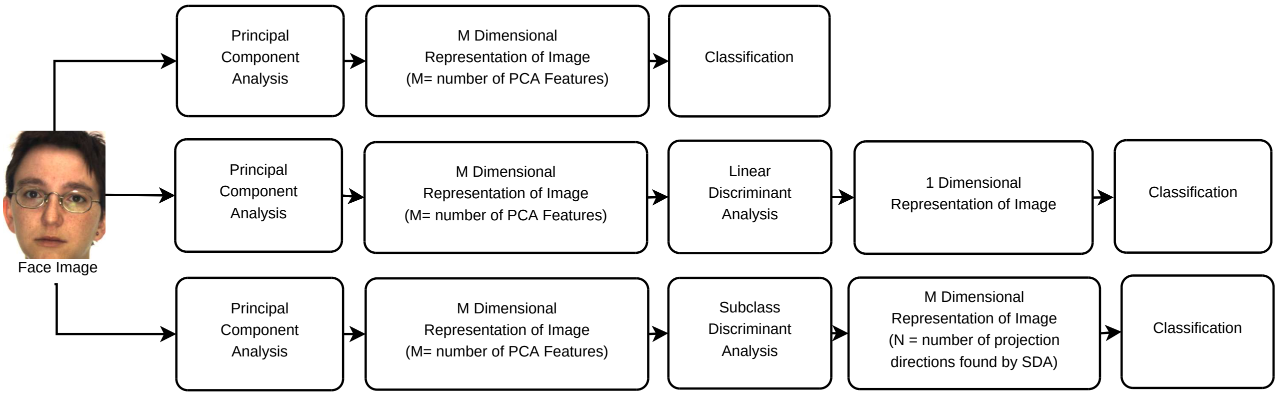


Figure 1. Illustrating the steps involved in the three algorithms.

$C_1 \sim N(\mu_1, \sigma_1^2)$ and $C_2 \sim N(\mu_2, \sigma_2^2)$, where the parameters μ_1, μ_2 and σ_1^2, σ_2^2 can be estimated from the training data. Therefore, $g_1(d) = N(d; \mu_1, \sigma_1^2)$ and $g_2(d) = N(d; \mu_2, \sigma_2^2)$ are the distribution functions of class C_1 and C_2 respectively. It is assumed that the prior $P(C_i) = 0.5, \forall i \in \{1, 2\}$. If we define $g(d) = g_1(d) - g_2(d)$, then the Bayesian decision function will be

$$f(d) = \begin{cases} C_1 & \text{if } g(d) > 0 \\ C_2 & \text{if } g(d) < 0 \end{cases}$$

This research assumes that the data follows a Gaussian distribution. However, it can be replaced by any other suitable distribution to which the data conforms.

Algorithm 1

Inputs: N_{PCA} (Number of PCA features), N_{tr} (Number of training images), $D = D_1 \cup D_2 \cup D_3 \ldots \cup D_N$, ($D$ = Database, N = Number of databases), L_D (Ground truths-labels)

1 Take random N_{tr} images from each gender of each database D_i, $1 \le i \le N$. These images are used as the train set D_{tr} and the remaining images comprise the test set D_{tt}.

2 Find principal components of D_{tr} and select the first N_{PCA} principal components to form the matrix $\mathbf{P}$ whose rows are the selected principal components.

3 Project the train and test samples in the PCA space, and obtain the transformed sets, D_{tr}^a and D_{tt}^a.

a) $D_{tr}^a = \{\mathbf{y} \big| \mathbf{y} = \mathbf{P}\mathbf{x}, \forall \mathbf{x} \in D_{tr}\}$

b) $D_{tt}^a = \{\mathbf{y} \big| \mathbf{y} = \mathbf{P}\mathbf{x}, \forall \mathbf{x} \in D_{tt}\}$

4 a) $Accuracy_{tr}^a = FindAccuracy(D_{tr}^a, L_{tr}, D_{tr}^a, L_{tr})$

b) $Accuracy_{tt}^a = FindAccuracy(D_{tt}^a, L_{tt}, D_{tr}^a, L_{tr})$

Algorithm 2

Inputs: N_{PCA} (Number of PCA features), N_{tr} (Number of training images), $D = D_1 \cup D_2 \cup D_3 \ldots \cup D_N$, ($D$ = Database, N = Number of databases), L_D (Ground truths-labels)

1 Take random N_{tr} images from each gender of each database D_i, $1 \le i \le N$. These images are used as the train set D_{tr} and the remaining images comprise the test set D_{tt}.

2 Find principal components of the D_{tr} and select the first N_{PCA} principal components to form the matrix $\mathbf{P}$ whose rows are the selected principal components.

3 Project the train and test samples in the PCA space, and obtain the transformed sets.

a) $D_{tr}^a = \{\mathbf{y} \big| \mathbf{y} = \mathbf{P}\mathbf{x}, \forall \mathbf{x} \in D_{tr}\}$

b) $D_{tt}^a = \{\mathbf{y} \big| \mathbf{y} = \mathbf{P}\mathbf{x}, \forall \mathbf{x} \in D_{tt}\}$

4 Compute the optimum projection direction of LDA ω and project the train and test datasets to compute the corresponding LDA projection, D_{tr}^b and D_{tt}^b.

a) $D_{tr}^b = \{\mathbf{y} \big| \mathbf{y} = \omega\mathbf{x}, \forall \mathbf{x} \in D_{tr}^a\}$

b) $D_{tt}^b = \{\mathbf{y} \big| \mathbf{y} = \omega\mathbf{x}, \forall \mathbf{x} \in D_{tt}^a\}$

5 a) $Accuracy_{tr}^b = FindAccuracy(D_{tr}^b, L_{tr}, D_{tr}^b, L_{tr})$

b) $Accuracy_{tt}^b = FindAccuracy(D_{tt}^b, L_{tt}, D_{tr}^b, L_{tr})$

3. Database

To evaluate the performance of gender recognition algorithms, we have prepared a database by combining images from several existing databases with different ethnicity and nationalities and is referred to as the *heterogeneous database*. The publicly available databases also contain images with other covariates including pose, illumination and expression. Since the focus of this research is gender recognition, we have selected frontal images with slight expression and illumination variations. Table 1 provides the

Algorithm 3

Inputs: N_{PCA} (Number of PCA features), N_{tr} (Number of training images), $D = D_1 \cup D_2 \cup D_3 \ldots \cup D_N$, ($D$ = Database, N = Number of databases), L_D (Ground truths-labels)

1 Take random N_{tr} images from each gender of each database D_i, $1 \leq i \leq N$. These images are used as the train set D_{tr} and the remaining images comprise the test set D_{tt}.

2 Find principal components of D_{tr} and select the first N_{PCA} principal components to form matrix $\mathbf{P}$ whose rows are the selected principal components.

3 Project the train and test samples in the PCA space, and obtain the transformed sets.

a) $D^a_{tr} = \{\mathbf{y} | \mathbf{y} = \mathbf{P}\mathbf{x}, \forall \mathbf{x} \in D_{tr}\}$

b) $D^a_{tt} = \{\mathbf{y} | \mathbf{y} = \mathbf{P}\mathbf{x}, \forall \mathbf{x} \in D_{tt}\}$

4 Compute the projection matrix $\mathbf{V}$ (whose rows are the projection directions) of SDA ω and find SDA projections of the train and test sets, D^c_{tr} and D^c_{tt}, respectively.

a) $D^c_{tr} = \{\mathbf{y} | \mathbf{y} = \mathbf{V}\mathbf{x}, \forall \mathbf{x} \in D^a_{tr}\}$

b) $D^c_{tt} = \{\mathbf{y} | \mathbf{y} = \mathbf{V}\mathbf{x}, \forall \mathbf{x} \in D^a_{tt}\}$

5 a) $Accuracy^c_{tr} = FindAccuracy(D^c_{tr}, L_{tr}, D^c_{tr}, L_{tr})$

b) $Accuracy^c_{tt} = FindAccuracy(D^c_{tt}, L_{tt}, D^c_{tr}, L_{tr})$

Algorithm 4 $FindAccuracy(D_x, Label_x, D_y, Label_y)$

Estimate the mean(μ_i) and variance(σ_i^2) of the classes in D_y and prepare the multivariate normal probability distribution function (MVNPDF) for them. For all the samples in D_x, predict the class label using the Bayesian decision boundary. These predicted labels are compared with the actual labels L_x to compute the accuracy value.

composition of the database. Out of these, the combined database is prepared along the lines of [15]. It contains images from the CMU PIE [14], Georgia Tech [1], GTAV [16] and FERET [12] face databases with neutral expression, minimum illumination variation, and no occlusion. The heterogeneous database contains total 0f 8,112 images out of which 4,246 are male faces and 3,866 are female face images. Haar-cascade detector in OpenCV is used for face detection and normalization. Figure 2 shows sample images from the heterogeneous database.

4. Experimental Results

To evaluate the performance of the three discriminant analysis functions for gender classification, three types of experiments are performed.

- The first experiment explores the effect of number of PCA features on the classification accuracy.
- The second experiment analyzes the effect of *amount of training* on the accuracy.
- The third experiment is a case study of the effect of ethnicity on the accuracy.

Table 1. Details of the heterogeneous database.

Database	No. of male face images	No. of female face images
AR [7]	527	453
Indian Face	250	250
FRGC [11]	600	407
Combined [15]	962	576
Notre Dame [4]	1712	580
Plastic Surgery [15]	195	1600
Total (8112)	**4246**	**3866**

Figure 2. Sample images from the database.

4.1. Gender Recognition and Number of PCA Components

In this experiment, the effect of the number of PCA components is explored by varying the number of PCA features from 10 to 100. The classification results are computed with three times cross validation.

- The results for PCA, PCA+LDA and PCA+SDA are shown in Figure 3 and Table 2. To accentuate the results, Table 2 shows the accuracy with 150 training images per gender per database. Therefore, in total, 1800 images are used for training and 6312 images are used for testing.
- On increasing the number of PCA features from 10-70, accuracy of PCA based approach increases and the

Table 2. Gender classification accuracy with varying number of PCA features.

Approach / PCA features	10	20	30	40	50	60	70	80	90	100
PCA	74.34	78.81	82.39	84.47	85.68	86.14	**86.47**	85.96	82.54	16.02
PCA+LDA	74.09	75.23	78.04	79.68	80.88	81.75	81.98	82.33	82.84	**83.29**
PCA+SDA	74.88	75.79	78.02	79.63	80.88	81.78	82.03	82.37	82.85	**83.29**
SVM Polynomial Kernel	· · ·	78.58	82.14	83.50	85.01	85.48	85.64	85.84	86.23	**86.33**

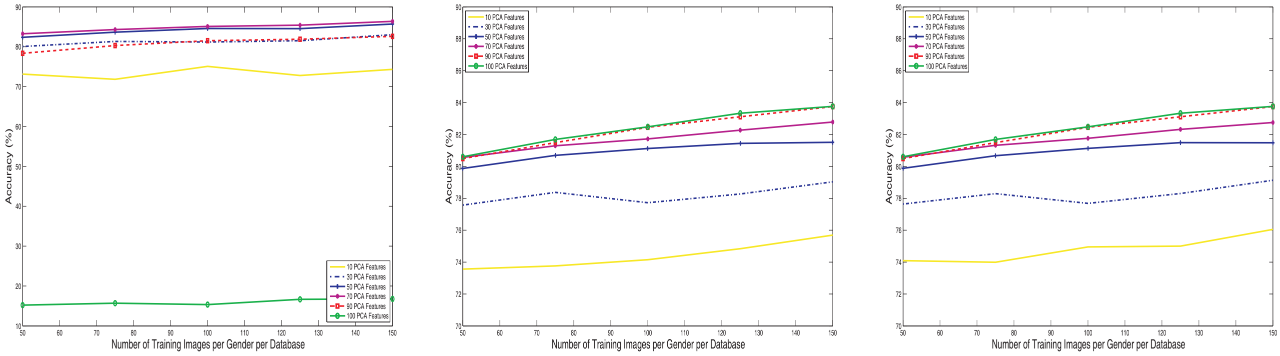


Figure 3. Classification performance of PCA, PCA+LDA, and PCA+SDA on varying the number of training images and principal components.

maximum performance achieved is 86.47% with 70 principal components. However, the algorithm yields around 82.39% accuracy with 30 principal components only and then there is small increment with increasing PCA features. This shows that with 30 principal components also, PCA is able to generalize to a good extent. On increasing the number of components beyond 70, the performance of PCA based approach reduces and for 100 components, it reduces to 16.02%. This sudden decrease can be attributed to the curse of dimensionality.

- PCA+LDA and PCA+SDA provide almost similar performances and unlike PCA, the performance increases with increasing the number of principal components. For this experiment, the performance of PCA+SVM is also evaluated with Polynomial kernel of degree 4. With 10 PCA features, PCA+SVM is not able to learn the classifier, however, for 20-100 PCA features, the performance is increasing and is in the range of 78-86%. It is interesting to note that the maximum performance of PCA is higher than the best results obtained by PCA+LDA, PCA+SDA, and PCA+SVM.

- Even though the input to LDA, SDA and SVM are dimensionality reduced principal components, these three algorithms are able to find good decision boundaries with large number of PCA components whereas PCA is providing very low accuracy for the same. It is because, greater the number of principal components, better is the approximation of the original image. Therefore, applying discriminant analysis or SVM over it, will lead to better/more generalized projection and hence, avoid the *curse of dimensionality*.

- Figure 4 shows sample images that are misclassified by all four approaches.

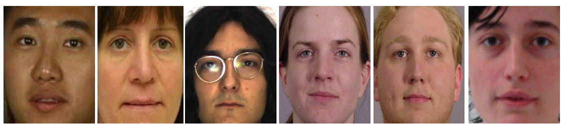


Figure 4. Face images misclassified by at least one of the approaches.

4.2. Gender Recognition and Number of Training Images

In large scale systems, it is difficult to initially provide a large number of training images. This experiment is designed to analyze the training requirement of discriminant functions. In this experiment, the number of training images is varied from 50 to 150 images per gender per database, with a step size of 25. To see the effect of the number of PCA features along with the number of training images, for each training size, we are also varying the number of PCA features. The results are presented in Figure 3.

- On varying the number of training images, minimum variation is observed in the performance of PCA. The accuracy does not change more than 2% for different number of training images. However, for PCA+LDA and PCA+SDA, the variation is comparatively higher.

- It is also observed that the classification performance remains unaltered or starts decreasing after the training dataset size is increased beyond 125 images per gender per database. either remains same or starts decreasing.

- PCA+LDA and PCA+SDA yield similar performance. We observed that in most of the cases no subclasses are formed and in such cases, SDA performs similar to LDA. It is consistent with the observations made by Zhu and Martinez [19] that in absence of classes, SDA and LDA provide similar performances.

4.3. Ethnicity Factor

This experiment is designed to understand "Is the gender classification a function of ethnicity?" Among the databases used for this research, one of the databases is the Indian face database in which all the subjects are of Indian ethnicity. The other databases contain face images pertaining to different ethnic origin and is challenging to separate them. Therefore, we have performed the experiments with Indian and non-Indian ethnicity[1]. This is a two fold experiment, in the first, the Indian face database is used for training and the remaining are used as test set. In the second experiment, the Indian face database is used for testing, while samples from the remaining databases are used for training. Since all the three approaches are providing high accuracies for different number of training images, we have performed extensive experiments with number of training images varying from 50 to 150 per gender per database. Figure 5 demonstrates the results of PCA, PCA+LDA, and PCA+SDA for this experiment.

In both the experiments, it is observed that PCA performs better than PCA+LDA and PCA+SDA for smaller values of PCA features and smaller training sizes. However, with increasing number of features, the performance of PCA approach reduces. This is consistent with the results obtained in Section 4.1. The performance of PCA+SDA and PCA+LDA remain nearly same for varying number of PCA features. However, the accuracy is considerably (15-20%) lower than that achieved in experiments shown in Sections 4.1 and 4.2. This experiment shows that ethnicity plays an important role in gender classification and PCA provides maximum generalization with respect to unseen ethnicity variations.

Bekios-Calfa *et al.* [3] have shown that with limited number of training data, discriminant functions provide better performance than non-linear functions. On the contrary, previous literature has shown that with sufficient training samples (on one database), non-linear functions are better than linear discriminant functions. However, the results in this research suggest that when higher number of training images are available, PCA+LDA, PCA+SDA, and PCA+SVM provide good performance. Moreover, when a range of training databases and principal components is available, PCA provides the best generalization performance compared to LDA, SDA, and SVM.

[1]The non-Indian database also contains some Indian face images but they are very few in number.

5. Conclusion

This research aims at studying two problems: (1) gender classification and (2) performance evaluation of discriminant functions with respect to their generalization capability. The performance of three discriminant functions is evaluated: PCA, PCA+LDA, PCA+SDA along with PCA+SVM as the non-linear classifier. The algorithms are studied in three scenarios: (1) varying number of training images, (2) varying number of principal components and training images, and (3) cross (unseen) ethnicity variations. To draw statistically meaningful inferences, a heterogeneous database of 8112 manually labeled images is prepared by combining several publicly available databases. The heterogeneous database comprises of face images pertaining to different ethnicity and minor variations in pose, illumination and expression. The results show that for gender classification, linear discriminant functions can provide good generalization performance even with limited training data and across ethnicity.

6. Acknowledgement

This research is supported through a grant from the Department of Information Technology, India.

References

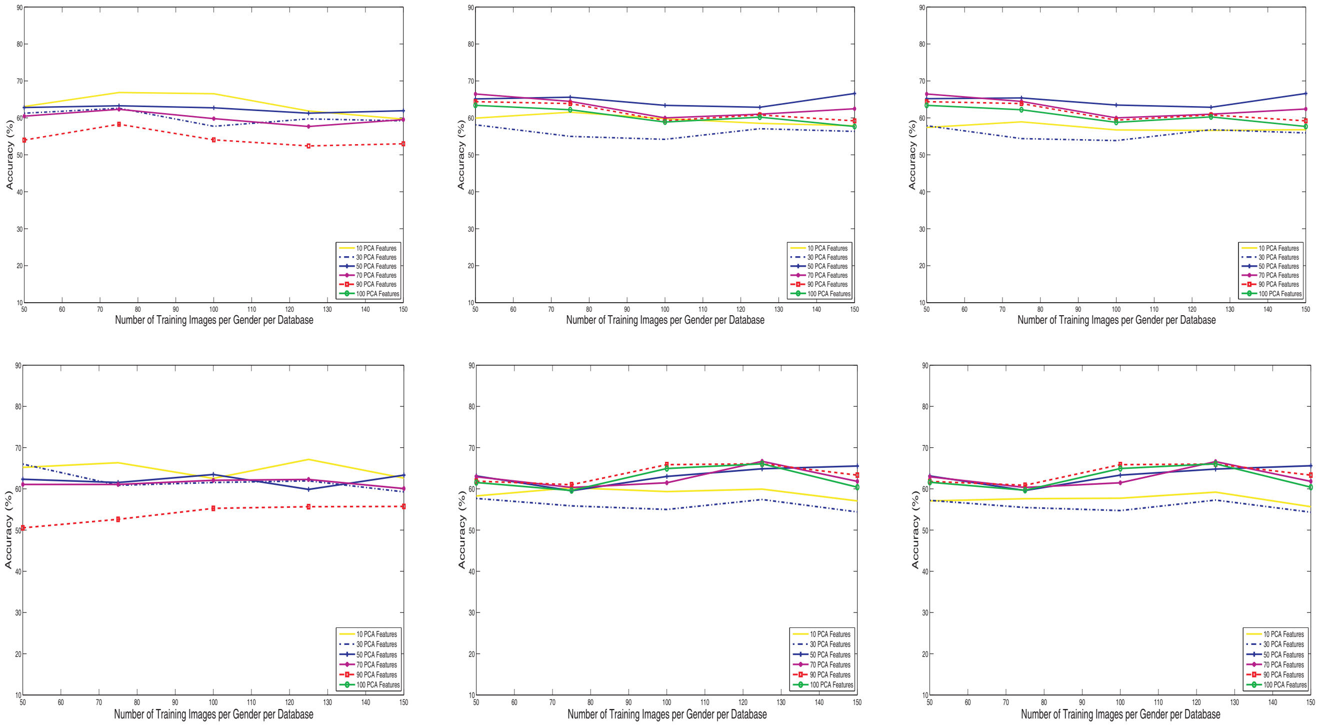
[1] Georgia tech face database, 2011. `ftp://ftp.ee.gatech.edu/pub/users/hayes/facedb/`.

[2] S. Baluja and H. A. Rowley. Boosting sex identification performance. *International Journal of Computer Vision*, 71(1):111–119, 2007.

[3] J. Bekios-Calfa, J. M. Buenaposada, and L. Baumela. Revisiting linear discriminant techniques in gender recognition. *IEEE Transactions on PAMI*, 33(4):858–864, 2011.

[4] K. W. Bowyer and P. J. Flynn. Notre dame biometrics database, 2011. `http://www.nd.edu/ cvrl/UNDBiometricsDatabase.html`.

[5] M. Castrillon-Santana and Q. C. Vuong. An analysis of automatic gender classification. In *Proceedings of Conference on Progress in Pattern Recognition, Image Analysis and Applications*, pages 271–280, 2007.



PCA PCA + LDA PCA + SDA

Figure 5. Classification performance of the three approaches for cross ethnicity experiments. The first row demonstrates the results when images with Indian ethnicity are used for training and the remaining are used for testing and the second row shows the results with Indian images for testing and a subset of the remaining images for training.

[6] A. Hadid and M. Pietikäinen. Manifold learning for gender classification from face sequences. In *Proceedings of International Conference on Biometrics*, pages 82–91, 2009.

[7] A. M. Martinez and R. Benavente. The AR face database. *CVC Technical Report #24*, 1998.

[8] A. M. Martínez and A. C. Kak. PCA versus LDA. *IEEE Transactions on PAMI*, 23(2):228–233, 2001.

[9] F. Matta, U. Saeed, C. Mallauran, and J.-L. Dugelay. Facial gender recognition using multiple sources of visual information. In *Proceedings of Workshop on Multimedia Signal Processing*, pages 785–790, 2008.

[10] B. Moghaddam and M.-H. Yang. Learning gender with support faces. *IEEE Transactions on PAMI*, 24(5):707–711, 2002.

[11] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. *In Proceedings of Conference on Computer Vision and Pattern Recognition*, 1:947–054, 2005.

[12] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi. The FERET evaluation methodology for face recognition algorithms. *IEEE Transactions on PAMI*, 22(10):1094–1104, 2000.

[13] F. Scalzo, G. Bebis, M. Nicolescu, and L. Loss. Feature fusion hierarchies for gender classification. In *Proceedings of International Conference on Pattern Recognition*, 2008.

[14] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression (PIE) database of human faces. Technical Report CMU-RI-TR-01-02, Robotics Institute, January 2001.

[15] R. Singh, M. Vatsa, H. Bhatt, S. Bharadwaj, A. Noore, and S. Nooreyezdan. Plastic surgery: A new dimension to face recognition. *IEEE Transactions on IFS*, 5(3):441 –448, 2010.

[16] F. Tarres and A. Rama. GTAV face database, 2011. `http://gps-tsc.upc.es/GTAV/ResearchAreas/UPCFaceDatabase/GTAVFaceDatabase.htm`.

[17] V. Vapnik. *Statistical learning theory*. Wiley, 1998.

[18] Y. Wang, K. Ricanek, C. Chen, and Y. Chang. Gender classification from infants to seniors. In *Proceedings of IEEE International Conference on Biometrics: Theory Applications and Systems*, pages 1–6, 2010.

[19] M. Zhu and A. Martinez. Subclass discriminant analysis. *IEEE Transactions on PAMI*, 28(8):1274 –1286, 2006.