International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:268
Vol.2, No.6, June 2013 DOI:10.15693/ijaist/2013.v2i6.85-92

Gender Classification based on Principal Component Analysis and Linear Discriminant Analysis Features with Sparse Representation

Annalakshmi.M
Assistant Professor/ Department of Communication Systems, Sethu Institute of Technology, Pulloor

Aishwarya.B
Department of Communication Systems, Sethu Institute of Technology, Pulloor.

Abstract— Computer vision and pattern recognition systems play an important role in our lives by means of automated face detection, face and gesture recognition, and estimation of gender and age. This paper addresses the problem of gender classification using frontal facial images. We have developed ¯gender classifier using dimensionality reduction techniques such as Principal Component Analysis (PCA) and Linear Discriminant techniques (LDA) along with Sparse Representation (SR). We experiment on 312 images (160 females and 152 males) randomly withdrawn from the LFW facial database. The input dataset is divided into training and testing dataset and experiments are performed by varying dataset size. The Bayesian Classifier has been used for the classification. This gender classification method classifies almost all the images with different image sizes. Overall classification rate of 87.50 % is achieved. As an additional contribution, a comparative study is made on Independent Component analysis (ICA), another dimensionality reduction technique. Our performance evaluation results show that (ICA+ LDA) have been outperformed by (PCA+LDA).

Keywords - Computer vision, Gender Classification, LDA, PCA, Sparse Representation, Bayesian Classifier, ICA

I. INTRODUCTION

In the last several years, various feature extraction and pattern classification methods have been developed for gender classification. Emerging applications of computer vision and pattern recognition in mobile devices and networked computing require the development of resource limited algorithms. Gender classification is a research topic with a high application potential in areas such as surveillance, face recognition, video indexing, and dynamic marketing surveys.

A. Fundamental Steps Involved In Gender Recognition

Generally Gender classification consists of the following steps. Figure 1 depicts these steps.

Preprocessing

Since, in real-life, it is unlikely that people will face directly and frontally towards the camera, face images often consist of some in-plane and out-of-plane rotations. Moreover, it is also unlikely that the light condition will be the same for all images. These variations greatly affect an accuracy of gender classifiers (e.g.: rotation, scaling and illumination change effects on classifier performance). The purpose of pre-processing step is thus to remove these variations as much as possible.

As with other computer vision applications, there is no unique solution to this problem. The common techniques involved in pre-processing step are face alignment, and light normalization. Face alignment tries to align faces such that they are closed to a common or specified pose of face as much as possible, whereas light normalization tries to get rid of the variation in illumination. One of the common employed normalization techniques in the gender classification field is histogram equalization.

Face Detection

In order to exploit uniqueness of faces in gender recognition, the first step is to detect and localize those faces in the images. This is the task achieved by face detection systems**.**

Most of the face detection systems are performed as a binary classification task. That is, given a part of image, the task is to decide whether it is a face or not. This is achieved by first transforming the given region into features and then using classifier trained on example images to decide if these features represent a human face.

As faces can appear in various locations and can also show themselves in various sizes, often, a window-sliding technique is also employed. The idea is to have the classifier classifying the portions of an image, at all location and scales, as face or non-face.Feature Extraction/ Dimensionality Reduction Working directly on raw pixel values can be very slow as one small face image can contain a thousand of pixels. Furthermore, not all the pixels will be useful. There can be an underlying structure that describes the differences between male and female faces better. Thus the feature detection module is employed here.

- **Approaches To Feature Extraction**

In general, gender classification in supervised learning setting requires extraction of features from face images, training classifiers using those features and finally performing classification of new faces. Generally gender recognition techniques can be divided into two groups based on the face representation they use:

Appearance-based, which uses holistic texture features and is applied to either whole-face or specific regions in a face image.

Feature-based, which uses geometric facial features (mouth, eyes, brows, cheeks etc.) and geometric relationships between them.

Among many approaches to the problem of face recognition, appearance-based subspace analysis, although one of the oldest, still gives the most promising results. Subspace analysis is done by projecting an image into a lower dimensional space (subspace) and after that recognition is performed by measuring the distances between known images and the image to be recognized. The most challenging part of such a system is finding an adequate subspace.

- **Techniques Of Dimensionality Reduction**

Since not all the detected features are useful, the feature selection (or dimensionality reduction) module is employed here to choose only a subset of representative features. Doing feature selection not only gives us the relevant features and thus the more accurate result but also give us an additional advantage of faster computation time as the dimensionality of data is reduced.

The popular feature selection techniques often employed in gender classification task are Principal Component Analysis (PCA), Independent Component Analysis (ICA)[, Adaboost and Genetic Algorithm.

Classification

With all necessary features have been extracted, the final task is to decide whether or not those features represent female or male face. As there are obviously two decisions to make this is essentially binary classification task, that is, the classifier is trained on the female and male example face images so that it learns the decision boundary between these two classes. After that it uses what it learn to make a decision on the given face images. For this purpose different types of classifiers are used. e.g. Bayesian, KNN,NN,and SVM.

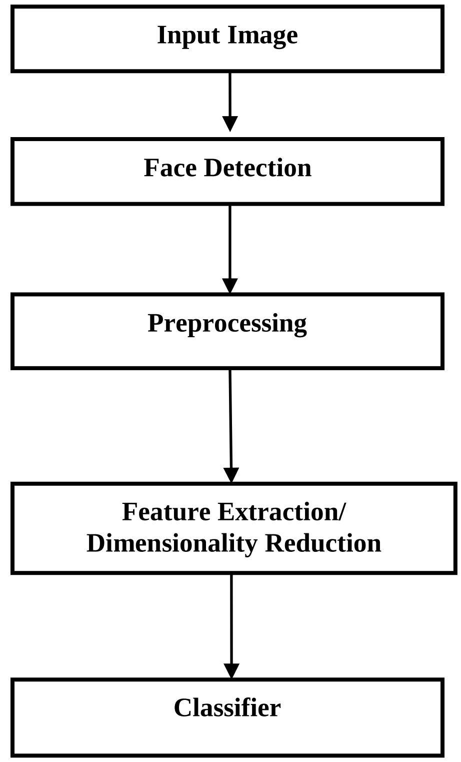

Figure 1. Fundamental Steps in Gender Classification

II. RELATED WORK

Almost all of the works in gender classification involves extracting features from faces and classifying those features using labeled data. They mostly differ in the way these two steps are performed. Therefore gender classification approaches can be categorized based on the feature extraction and classification methods. Feature extraction can be broadly categorized into a) Appearance-base methods, and b) Geometry-based methods. In appearance-based methods the whole image is considered rather than local features that are present in different parts of the face. On the other hand in geometry-based approaches, the geometric features (e.g. distance between eyes, face width, length, thickness of nose, etc.) of a face are considered. In this section, only the works related to appearance-based approaches are discussed. For the case of classification, most of the works use neural networks, discriminant analysis, nearest neighbors, and SVMs.

Early works in gender classification mostly used neural networks with face image as raw input. Some of these are Golomb et al.'s two-layer network called SEXNET [1], Tamura et al.'s multilayer neural network [2], etc. Gutta et al. in [3] takes a hybrid approach using neural network and decision trees.

Moghaddam et al. in [6] uses non-linear SVMs to classify faces from low-resolution "thumbnail" images of size 21-by-12. The authors also experimented with other types of classifiers including different types of RBFs, Fisher's linear discriminant, Nearest Neighbor, and Linear classifier. For SVM they looked at Gaussian RBF kernel and cubic polynomial kernels. They used a total of 1,755 thumbnails (1,044 males and 711 females) and reported the error rate of performing fivefold cross-validation. The best result was obtained for SVM with Gaussian RBF kernel which had an overall error rate of 3.38%, for males and females' error rates were 2.05% and 4.79% respectively.

Makinen and raisamo [13] performed a set of experiments using 411 images from the FERET database. They compared appearance based, feature-based, aligned, and unaligned approaches, among others. They got similar performance results for feature-based AdaBoost and appearance-based SVM+RBF classifiers.

Jain et al. in [7] presents an approach using ICA and SVM. They studied the performance of different classifiers namely- cosine classifier that finds the distance between two features lying on an hyper-sphere surface, linear discriminant classifier that finds the projection of the input image maximizing the ratio of the between-class scatter and within class scatter, and SVM which finds the maximal separating hyper-plane between the male and female features. A training set of 200 images out of a database of size 500 was used in their work. Using ICA 200 independent components were determined from the training set. They also experimented with different sizes of training set. In their work, SVM performed constantly well with respect to the other classifiers. The best performance they got was 95.67% using ICA and SVM for a training set of size 200.

With the notable exception of [2], existing approaches togender recognition focus mainly on high-performance computer systems. Emerging applications of video analysis in mobile devices and networked computing have recently attracted interest in the development of computer vision and pattern recognition algorithms for resource-limited devices. Linear classification techniques have an important role to play given their simplicity and low computational requirements at runtime.

III. OBJECTIVES & OVERVIEW OF THE PROPOSED MECHANISM

A. *Objectives*

In this paper, three most popular appearance-based subspace projection methods for gender recognition will be presented. Projection methods to be presented are: Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Linear Discriminant Analysis (LDA). PCA (Turk and Pentland, 1991) find a set of the most representative projection vectors such that the projected samples retain most information about original samples. ICA (Bartlett et al., 2002; Draper et al., 2003) captures both second and higher order statistics and projects the input data onto the basis vectors that are as statistically independent as possible. LDA (Belhumeur et al., 1996; Zhao et al., 1998) uses the class information and find a set of vectors that maximize the between-class scatter while minimizing the within-class scatter. .

Various algorithms were proposed and research groups across the world reported different and often contradictory results when comparing them. The aim of this paper is to present an independent, comparative study of three most popular appearance-based face recognition projection methods (PCA, ICA, and LDA) in completely equal working conditions regarding preprocessing and algorithm implementation. We are motivated by the lack of direct and detailed independent comparisons of all possible algorithm implementations (e.g., projection–metric combinations) in available literature. For consistency with other studies, LFW data set is used with its standard tests.

B. *Overview of the proposed Mechanism*

In our gender classification system, first we crop and resize images to 25 * 25 pixels using OpenCV's2 2.0.0 face detector, which is based on. Then, we apply an oval mask to prevent the background. Additionally if any noise occurred in our input image are also removed using median filter. We have extracted features using PCA, LDA methods and we also learnt that features to Bayesian classifier. Finally we test whether the input image is male or female, Bayesian classifier returns which gender the input image is belonging to the trained features.In our proposed method we have introduced sparse representation to arrange the feature values.

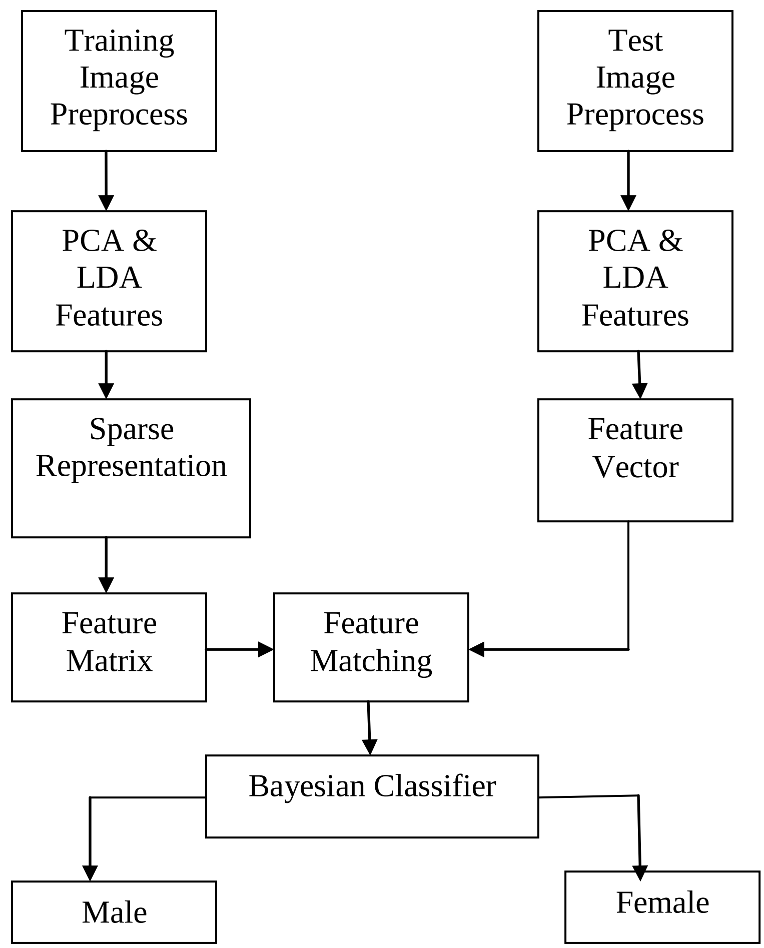

Figure 2. System Overview

IV. PROPOSED WORK

A. Modules In Proposed Work

The modules in our proposed work are,

- Load Image
- Face Detection
- Preprocessing
- Train Feature Extraction
- Test Feature Extraction
- Classification

B. Module Description

Load Image

In this module we load the test images as our option. Test images are the face that may or may not present in our database.

Face Detection

We crop and resize images to 25*25 pixels using OpenCV's2 2.0.0 face detector, which is based on Viola and Jones face detection method

Preprocessing

Oval mask are applied to the input image to prevent the background from influencing. And also if any noise occurred in our input image that are removed. Noise is nothing but any undesired information that contaminants the image. These noises are removal some types of filters. We are using Median Filter to removing the noises.

Train Feature Extraction

PCA & LDA features are extracted from our training images (dataset).In recognition test image PCA & LDA features are extracted and matched to the train image features then it returns the recognized face.

- **PCA:** Principal component analysis is a statistical tool used to analyze data sets. The mathematics behind principle component analysis is statistics and is hinged behind standard deviation, eigenvalues and eigenvectors.

  Feature Extraction Using PCA: Principal Components Analysis is a very well known approach for reducing the dimensionality of data. For applying PCA to images, the image is first represented as a column of vectors. A matrix is formed by concatenating the column of training set images. Let this matrix be **X,**

  X=[x1 x2…. xn], where xi is the ith column vector representing the ith training image.

  Then the mean is subtracted from each column and the covariance matrix is computed.

  Let the mean image be,

$$\bar{x}=\frac{1}{n}\sum_{i-1}^{n} x_i$$

  And **Y** = [$x_1 - \bar{x} \ldots x_n - \bar{x}$]

  The covariance matrix **Q** = cov(**Y**) = $\mathbf{YY^T}$

  Finally, eigenvalue decomposition is performed to find the highest ranking (based on eigenvalues) eigenvectors. These vectors, known as principal components span the low dimensional subspace. Out of these eigenvectors *m* most significant vectors are chosen, let these vectors be e1, e2…em. The value of *m* is chosen by considering the cumulative sum of the eigenvalues.

  The features of an image *x* is then computed by projecting it onto the space spanned by the eigenvectors as follows,

  g = $[e1\ e2\ \ldots em]^T$ (x $-\bar{x}$), where g is an *m* dimensional vector of features.

  This feature vector *g* is used during training and classification.

- **LDA:** The standard LDA can be seriously degraded if there are only a limited number of observations N compared to the dimension of the feature space n [5]. To prevent this from happen is it is recommended that the linear discriminant analysis be preceded by a principle component analysis. 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.

  Feature Extraction using LDA: Linear Discriminant Analysis (LDA) finds the vectors in the underlying space that best discriminate among classes. For all samples of all classes the between-class scatter matrix SB and the within-class scatter matrix SW are defined by:

$$S_B=\sum_{i=1}^{c} M_i\ (x_i-\mu)(x_i-\mu)^T$$

$$S_w=\sum_{i=1}^{c}\ \sum_{x_k\in X_i}\ (x_k-\mu_i)(x_k\ -\mu_i)^T$$

Where M_i is the number of training samples in class i, c is the number of distinct classes, μ_i is the mean vector of samples belonging to class i and Xi represents the set of samples belonging to class i with $x_k$being the k-th image of that class. S_W represents the scatter of features around the mean of each face class and S_B represents the scatter of features around the overall mean for all face classes. The goal is to maximize $S_B$while minimizing S_w, in other words, maximize the ratio det|S_B| / det|S_w |. This ratio is maximized when the column vectors of the projection matrix (W_{LDA}) are the eigenvectors of $S_w{}^{-1}S_B$. In order to prevent S_w to become singular, PCA is used as a preprocessing step and the final transformation is $W_{opt}^T=W_{LDA}^T W_{PCA}^T$.

- **ICA:** ICA captures both second and higher-order statistics and projects the input data onto the basis vectors that are as statistically independent as possible.

  Feature Extraction using ICA: The standard LDA can be seriously degraded if there are only a limited number of observations N compared to the dimension of the feature space n. To prevent this from happen is it is recommended Independent Component Analysis is another well known approach for blind signal separation where a signal is considered to be a linear combination of independent sources. If *s* is the vector representing the unknown sources, and **A** is the mixing matrix then the observed signal *x* is represented as ,

  x = **A**s

  ICA tries to find a separating matrix **W** such that

  u = **WA**s, where *u* is an estimation of *s*.

  For the case of images, **X** is the matrix whose columns are images in column vector form and **S** is the matrix whose columns are the independent components.

Train Feature Extraction

Like train images also, PCA & LDA features are extracted from our test image these test features are matched to our training dataset features then it returns which gender the input image is.

Classification

Finally trained image features are learnt by Bayesian classifier along with the true labels, then test features also passed through Bayesian classifier it matched to the trained features belonging the feature values and the true label Bayesian returns whether the test image is male or not..

V. Experimental Results

A. Dataset

The LFW dataset is used in this paper which contains 312 frontal images, representing different races, with different facial expressions, and under different lighting conditions. One- hundred and sixty of them are female, and the rest are male.

Figure 3. Some Of The Male Images From LFW Database

Figure 4. Some Of The Female Images From LFW Database

B. Results Of Gender Classification (PCA+LDA)

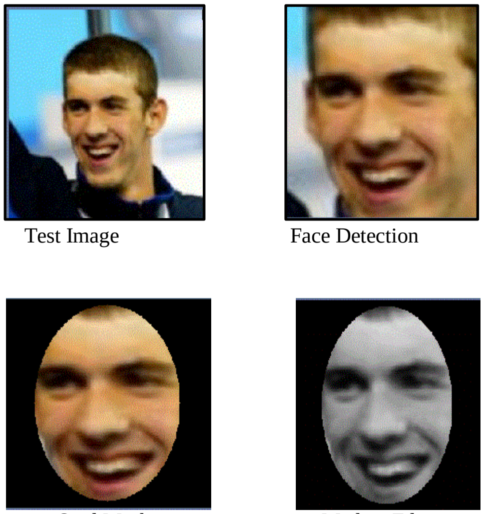

Figure 5. Gender Classification- Preprocessing (MALE)

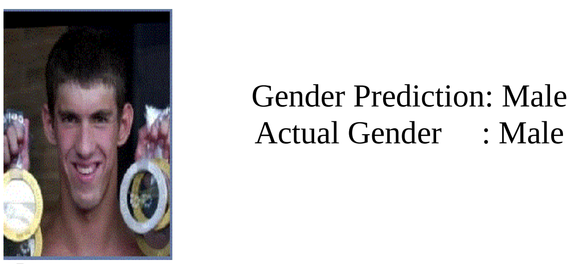

Figure 6. Gender Classification-- Recognition (MALE)

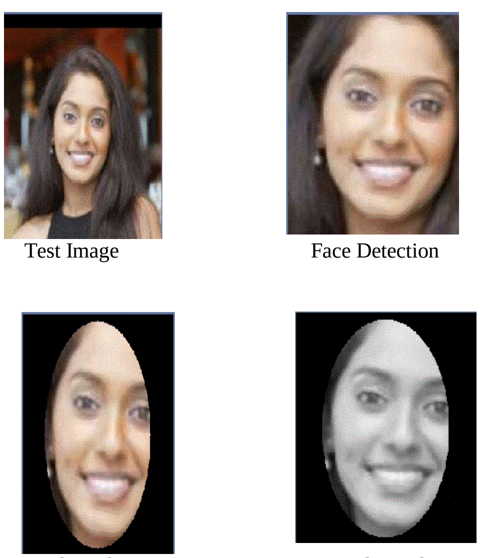

Figure 5. Gender Classification- Preprocessing (FEMALE)

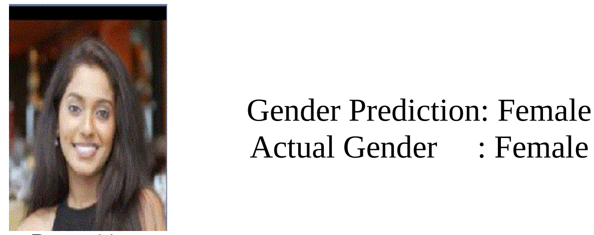

Figure 8. Gender Classification- Recognition (FEMALE)

VI. PERFORMANCE EVALUATION

A. Performance Metrics

We evaluate mainly the performance according to the following metrics.

Confusion Matrix: A confusion matrix (Kohavi and Provost, 1998) contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix. The following table shows the confusion matrix for a two class classifier.

The entries in the confusion matrix have the following meaning in the context of our study:

- *a* is the number of **correct** predictions that an instance is **Male**,
- *b* is the number of **incorrect** predictions that an instance is **Female**,
- *c* is the number of **incorrect** of predictions that an instance **Male**, and
- *d* is the number of **correct** predictions that an instance is **Female**.

		Predicted	
		Male	**Female**
Actual	**Male**	a	b
	Female	c	d

Several standard terms have been defined for the 2 class matrix:

- **The accuracy (AC)** is the proportion of the total number of predictions that were correct. It is determined using the equation:

$$AC = \frac{a+d}{a+b+c+d}$$

- The **recall or true positive rate (TP)** is the proportion of positive cases that were correctly identified, as calculated using the equation:

$$TP = \frac{d}{c+d}$$

- The **false positive rate (FP)** is the proportion of negatives cases that were incorrectly classified as positive, as calculated using the equation:

$$FP = \frac{b}{a+b}$$

- The **true negative rate (TN)** is defined as the proportion of negatives cases that were classified correctly, as calculated using the equation:

$$TN = \frac{a}{a+b}$$

- The ***false negative rate (FN)*** is the proportion of positives cases that were incorrectly classified as negative, as calculated using the equation:

$$FN = \frac{c}{c+d}$$

Sensitivity/ Recall:

Recall is a measure of the ability of a prediction model to select instances of a certain class from a data set. It is commonly also called sensitivity, and corresponds to the true positive rate. It is defined by the formula:

Recall = Sensitivity = tp/(tp+fn)

where tp and fn are the numbers of true positive and false negative predictions for the considered class. tp + fn is the total number of test examples of the considered class.

Specificity:

Recall/sensitivity is related to specificity, which is a measure that is commonly used in two class problems where one is more interested in a particular class. Specificity corresponds to the true-negative rate.

Specificity = tn/(tn+fp)

where tn and fp are the numbers of true negative and false positive predictions for the considered class. tp + fn is the total number of test examples of the considered class.

B. Performance Evaluation Result For (PCA+LDA):

Table 1. *Performance Evaluation*

Parameter Evaluation	
No. Of Observations	**312**
No. of Male Images	**152**
No. Of. Female Images	**160**
Accuracy (%)	**87.50**
Confusion Matrix	$\begin{bmatrix} 142 & 29 \\ 10 & 131 \end{bmatrix}$
Error Rate (%)	**12.50**
Sensitivity (%)	**93.42**
Specificity (%)	**81.87**

C. Performance Evaluation of (PCA+LDA) vs (ICA+LDA):

A comparative study is made on Independent Component analysis (ICA), another dimensionality reduction technique. Our performance evaluation results (recognition rate %) show that (ICA+ LDA) have been outperformed by (PCA+LDA).

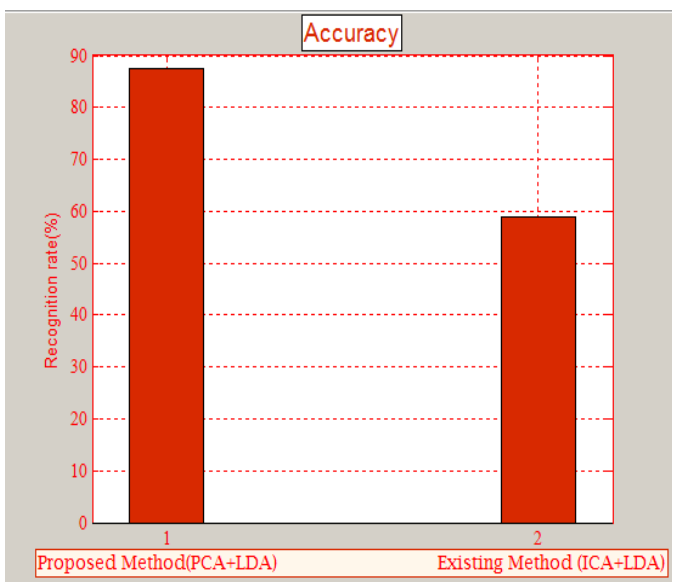

Figure 9. (PCA+LDA) Vs (ICA+LDA)

VII. CONCLUSION

In our proposed work, we performed gender classification using dimensionality reduction techniques namely PCA and LDA along with Sparse Representation is presented. Appearance-based approach is taken with the preprocessing step of oval masking and median filtering. Features are extracted after performing dimensionality reduction and classification is performed using Bayesian classifier. Results for varying dataset size and varying pose, illumination and expressions are presented. Both PCA and LDA can achieve very good accuracy with different combinations of the above parameters. In the current work, the database size was small. The performance of the approach can be better understood by using a larger database. It would also be interesting to see how the accuracy varies for people of different ethnicity. For this work all the highest ranking vectors found by PCA and ICA were used. For large training set it would be computationally efficient to choose a subset of these components. One possible research direction would be to determine the importance of the components for gender classification.

REFERENCES

[1]. B.A.Golomb, D.T.Lawrence, andT.J.Sejnowski. SEXNET:A neural network identifies sex from human faces. In Advances in Neural Information Processing Systems, pages 572--577, 1991.
[2]. [4] Shinichi Tamura, Hideo Kawai, and Hiroshi Mitsumoto. Male/female identification from 8×6 very low resolution face images by neural network. Pattern Recognition, vol. 29, no 2,pp 331-335, 1996.
[3]. [5] Gutta, S.; Wechsler, H.; Phillips, P.J., Gender and ethnic classification of face images, Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on , vol., no., pp.194-199, 14-16 Apr 1998
[4]. P. Phillips, H. Moon, P. Rauss, and S. Rizvi, “The Feret Evaluation Methodology for Face Recognition Algorithms,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 10, pp. 1090-1104, Oct. 2000.
[5]. A.M. Martinez and A.C. Kak, “PCA versus LDA,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 228-233, Feb 2001.
[6]. B. Moghaddam and M.-H. Yang, “Learning Gender with Support Faces,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 707-711, May 2002.
[7]. A. Jain and J. Huang, “Integrating Independent Components and Linear Discriminant Analysis for Gender Classification,” Proc. Int’l Conf. Automatic Face and Gesture Recognition, pp. 159-163, 2004.
[8]. M. Minear and D.C. Park, “A Lifespan Database of Adult Facial Stimuli,” Behavior Research Methods, Instruments and Computers, vol. 36, pp. 630-633, 2004.
[9]. P. Viola and M.J. Jones, “Robust Real-Time Face Detection,” Int’lJ. Computer Vision, vol. 57, no. 2, pp. 137-154, May 2004.
[10]. W. Zheng, L. Zhao, and C. Zou, “An Efficient Algorithm to Solve the Small Sample Size Problem for LDA,” Pattern Recognition, vol. 37, pp. 1077-1079, 2004.
[11]. J. Ye, R. Janardan, C.H. Park, and H. Park, “An Optimization Criterion for Generalized Discriminant Analysis on Undersample Problems,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 26, no. 8, pp. 982-994, Aug. 2004.
[12]. M. Zhu and A.M. Martı´nez, “Selecting Principal Components in a Two-Stage LDA Algorithm,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. I, pp. 132-137, 2006.
[13]. E. Ma¨kinen and R. Raisamo, “Evaluation of Gender Classification Methods with Automatically Detected and Aligned Faces,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 30, no. 3, pp. 541-547, Mar. 2008.
[14]. G. Guo, C.R. Dyer, Y. Fu, and T.S. Huang, “Is Gender Recognition Affected by Age?” Proc. IEEE Int’l Conf. Computer Vision Workshop Human-Computer Interaction, pp. 2032-2039, 2009.
[15]. Juan Bekios-Calfa, Jose M. Buenaposada, and Luis Baumela, “Revisiting Linear Discriminant Techniques in Gender Recognition”, IEEE Trans. Pattern Analysis And Machine Intelligence, Vol. 33, no. 4, April 2011.

Authors Profile

M. Annalakshmi received the **B.E.** degree in electronics and communication engineering from the Thiyagarajar College Of Engineering, Madurai, India, in 1998. She received **M.E.** degree in electronics and communication engineering (Optical Communication) in Alagappa Chettiar College Of Engineering and Technology, Karaikudi, India, in 2005. Her research interest includes Digital Image Processing, Linear Discriminant Analysis, Principal Component Analysis, Independent Component Analysis and Sparse Representation

B. Aishwarya received the **B.Tech** degree in electronics and communication engineering from the Kalasalingam University, Krishnankovil, Srivilliputtur, India, in 2011.Currently doing **M.E.** in electronics and communication engineering (Communication Systems) in Sethu Institute of Technology, Pulloor, Anna University, Chennai, Tamil Nadu, India. Her research interest includes Digital Image Processing, Linear Discriminant Analysis, Principal Component Analysis, Independent Component Analysis and Sparse Representation