Classification of Face Images Based on Gender using Dimensionality Reduction Techniques and SVM

Fahim Mannan
260 266 294
School of Computer Science
McGill University

Abstract
This report presents gender classification based on facial images using dimensionality reduction techniques such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA) along with Support Vector Machine (SVM). The input dataset is divided into training and testing dataset and experiments are performed by varying dataset size. The effect of performing image intensity normalization, histogram equalization, and input scaling are observed. The outcomes of the experiments are analogous to published works that apply similar techniques.

1. Introduction

Gender classification using facial images has been of interest for quite some time. Early works were mostly related to psychological research where the process by which humans determine gender from faces is studied. Humans are very good at determining gender from facial images. Even if the face is cropped to remove all gender cues, we can identify gender with very high accuracy ([1]).More recently automated gender classification from facial images has gained much interest in the computer vision and machine learning community. This is because of its extreme importance in Human Computer Interaction, demographic research, and security and surveillance applications. It can also augment other important areas like face recognition, age and ethnicity determination. Several approaches have been taken to classify facial images based on gender. This report addresses one particular approach using dimensionality reduction (ICA and PCA) and Support Vector Machine (SVM).

One of the challenges of automatic gender classification is to account for the effects of pose, illumination and background clutter. Practical systems have to be robust enough to take these issues into consideration. Most of the work in gender classification assumes that the frontal views of faces, which are pre-aligned and free of distracting background clutters, are available. Towes et al.[2] provides a framework that is free of these assumptions and can classify faces by first automatically detecting, then localizing and finally extracting features from arbitrary viewpoints. But in this project, only the facial images with full frontal views are considered.

The report is organized as follows - section 2 presents an overview of related research. Section 3, looks at the general approach and explains dimensionality reduction techniques as well as SVMs. Experiments are illustrated in section 4 and the results are discussed in section 5. Section 6 concludes with a discussion of possible future works.

2. 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 [3], Tamura et al.'s multilayer neural network [4], etc. Gutta et al. in [5] 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 five fold 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.

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 there 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.

3. Approach

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. This work uses appearance-based approach with dimensionality reduction techniques for feature extraction. The features extracted from the training set are used for training an SVM classifier. And finally images in the test set are classified using the classifier.

The general approach taken in gender classification is summarized in figure 1. First pre-processing operations on the input face image. These operations are face normalization, image intensity normalization and histogram equalization. In the face normalization step faces are cropped and aligned so that parts of the face (e.g. eyes, nose and mouth) fall into predefined locations. Then image intensity normalization and histogram equalization are performed to account for varying lighting conditions. Finally using PCA or ICA a set of basis vectors in a lower dimensional space is determined and the face images are projected onto the subspace spanned by those basis vectors. These basis vectors encode the discriminatory features of a human face. A training set is then formed by taking a set of labeled faces and extracting the features using the above approach. Then a classifier is trained with the labeled data and feature set pair. For a query image, the features are extracted in the same way. The classifier uses these features to determine the gender from a persons facial image.

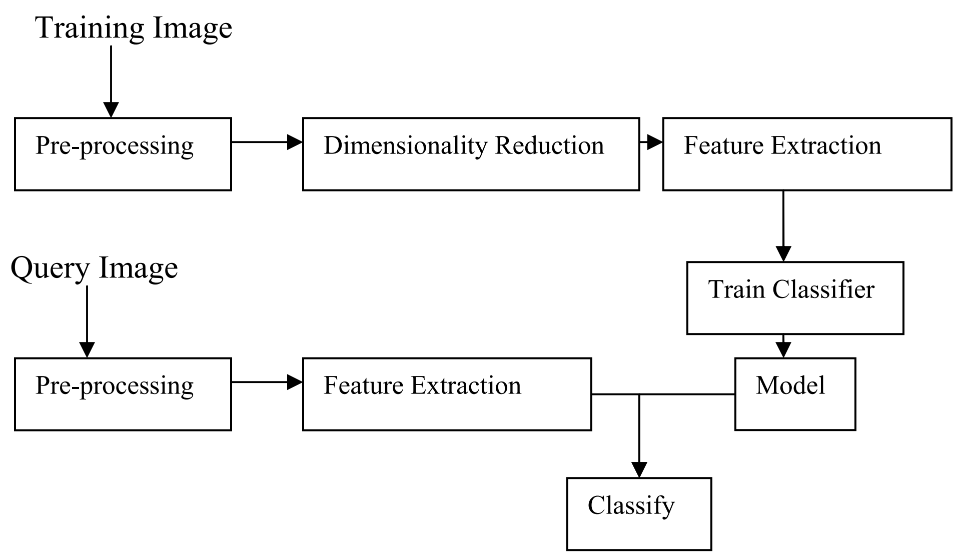


Figure 1. General Approach for Gender Based Face Classification using Dimensionality Reduction Techniques and SVM

3.1 Feature Extraction

As was described before feature extraction is done by projecting the face image onto a lower dimensional subspace. Dimensionality reduction techniques PCA and ICA are used for this approach. One of the objectives of this project was to see how the two techniques perform for gender classification. The motivation behind doing dimensionality reduction is to work with the most useful components of an image. This improves both the computation time and performance of the method that is used. In the following, the two dimensionality reduction methods are discussed.

3.1.1 Principal Component Analysis

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,**
$\mathbf{X} = [x_1\ x_2\ \dots\ x_n]$, where x_i is the i^{th} column vector representing the i^{th} 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 $\mathbf{Y} = [\, x_1 - \bar{x}\ \dots\ x_n - \bar{x}\,]$
The covariance matrix $\mathbf{Q} = \mathrm{cov}(\mathbf{Y}) = \mathbf{Y}\mathbf{Y}^T$

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 – $e_1, e_2, \dots, e_m$. 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 –

$g = [e_1\ e_2 \dots e_m]^T (x - \bar{x})$, where g is an *m* dimensional vector of features.

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

3.1.2 Independent Component Analysis

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 –

$x = \mathbf{A}s$

ICA tries to find a separating matrix **W** such that $u = \mathbf{WA}s$, where *u* is an estimation of *s*. In [8], an algorithm is given that finds an estimate of **W** by iteratively refining the columns of **W**.

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.

3.2 Classification using Support Vector Machine

Support vector machines are classifiers that construct a maximal separating hyperplane between two classes so that the classification error is minimized. For linearly non-separable data the input is mapped to high-dimensional feature space where they can be separated by a hyperplane. This projection into high-dimensional feature space is efficiently performed by using kernels. For instance-label pair (x_i, y_i) with $x_i \in \Re^n$, $y_i \in \{-1,1\}$ for $1 \le i \le n$ where n is the number of instances, the following optimization problem needs to be solved for SVMs –

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \xi_i$$

subject to $y_i (w^T \phi(x_i) + b) \ge 1 - \xi_i$,

$\xi_i \ge 0$

In the above equation, C is the penalty parameter for error term and ϕ maps a training instance x_i to higher dimensional space. The kernel K is defined as –

$K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$

For this project, a Radial Basis Function (RBF) kernel was used which is defined as –

$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ where $\gamma \ge 0$

The parameter γ controls the spread of a Gaussian cluster. Therefore in the above formulation there are two parameters, C and γ to control the performance of the classifier.

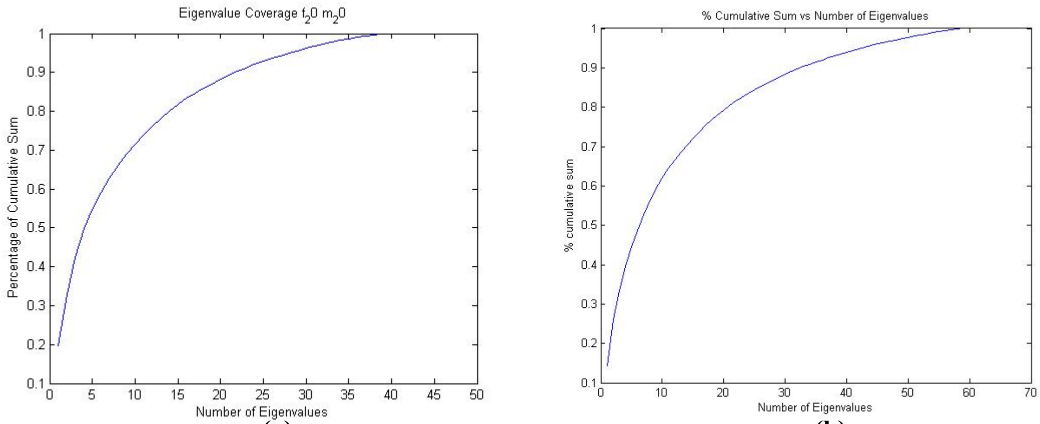
4. Experiments

Experiments were carried out with varying size of the training dataset, intensity normalization, histogram equalization, scaling training and testing dataset. Image database from [9] was used for this purpose. The database consists of full frontal face photos of 100 individuals (50 females and 50 males). The images were cropped by removing hair and background. The two training sets were of sizes 40 (20 females and 20 males) and 60 (30 females and 30 males). The rest of the images were used for testing. Same variations were made on both ICA and PCA approaches to observe the effects of different choices.

In the pre-processing step, image intensities are normalized (if applicable) and the images are reduced to 48x48 pixels. Histogram equalization is then performed (if applicable) on these images.

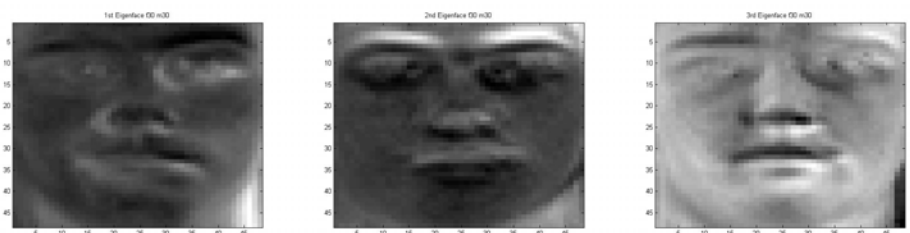
In PCA, the first 40 and 60 components were chosen for the training dataset of size 40 and 60 respectively. The plots in figure 2 show the percentage of the cumulative sum of the eigenvalues. The first three eigenfaces are shown in Figure 3 (a-c). These faces were obtained from the first three eigenvectors.

For the case of ICA, the FastICA [10] algorithm found 40 and 60 components respectively for the two sizes of training set. The top three ICA faces are shown in Figure 3(d-f).

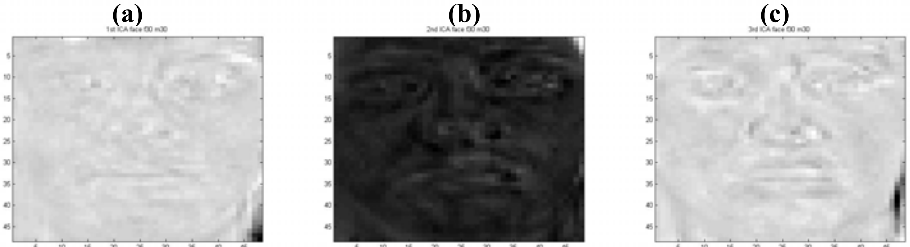


(a) **(b)**

Figure 2. Percentage of Cumulative Sum vs. Number of Eigenvalues for dataset of size a) 40 and b) 60



(a) **(b)** **(c)**



(d) **(e)** **(f)**

Figure 3. a) First three Eigenfaces for dataset of size 60 b) First three ICA faces for dataset of size 60

For SVM classification LIBSVM [11] was used. The choice of optimal C and γ was made by performing a grid search with values ranging from 2^{-n} to 2^{n} and doing five fold cross-validation for each choice. In the following table, the results for varying various parameters with the optimal values of C and γ are given. Some of the misclassified images are shown in figure 4.

Table 1:Results using ICA

Training Set Size	Intensity Norm.	Hist. Eq.	Scaling	Gaussian Kernel Parameters		CV Accuracy	Accuracy
				C	γ		
40	N	N	N	0.03125	0.0078125	50.00%	50.00%
40	N	N	Y	8192.0	0.0001221	92.50%	80.00%
40	N	Y	N	0.03125	0.0078125	50.00%	50.00%
40	N	Y	Y	32.0	0.000488	90.00%	81.67%
40	Y	N	N	0.5	0.000488	72.50%	58.33%
40	Y	N	Y	32.0	0.0078125	77.50%	75.00%
40	Y	Y	N	0.03125	0.000488	80.00%	70.00%
40	Y	Y	Y	32.0	0.000488	90.00%	81.67%
60	N	N	N	0.03125	0.0078125	50.00%	50.00%
60	N	N	Y	2048.0	0.0001221	88.33%	85.00%
60	N	Y	N	0.03125	0.0078125	50.00%	50.00%
60	N	Y	Y	32.0	0.000488	86.67%	**92.50%**
60	Y	N	N	0.03125	0.0019531	68.33%	57.50%
60	Y	N	Y	128.0	0.0078125	93.33%	**90.00%**
60	Y	Y	N	0.03125	0.000488	85.00%	85.00%
60	Y	Y	Y	3.0	0.0078125	86.67%	**92.50%**

Table 2:Results using PCA

Training Set Size	Intensity Norm.	Hist. Eq.	Scaling	Gaussian Kernel Parameters		CV Accuracy	Accuracy
				C	γ		
40	N	N	N	0.03125	0.000122	90.00%	68.33%
40	N	N	Y	32.0	0.0078125	75.00%	73.33%
40	N	Y	N	0.03125	0.0000305175	80.00%	70.00%
40	N	Y	Y	0.03125	0.0078125	90.00%	78.33%
40	Y	N	N	0.03125	0.0078125	90.00%	71.67%
40	Y	N	Y	1.0	0.00078125	77.50%	70.00%
40	Y	Y	N	0.5	0.0078125	90.00%	81.67%
40	Y	Y	Y	0.03125	0.0078125	90.00%	78.33%
60	N	N	N	0.03125	0.00003052	81.67%	65.00%
60	N	N	Y	2.0	0.03125	85.00%	**90.00%**
60	N	Y	N	0.03125	0.00003052	85.00%	85.00%
60	N	Y	Y	0.03125	0.5	68.33%	77.50%
60	Y	N	N	8.0	0.0078125	90.00%	**97.50%**
60	Y	N	Y	8.0	0.0078125	86.67%	**92.50%**
60	Y	Y	N	0.5	0.0078125	88.33%	**92.50%**
60	Y	Y	Y	32.0	0.0078125	63.33%	**95.00%**

Figure 4. A sample set of misclassified faces. The first three are of females and the last two males.

5. Discussion

The effect of using larger training set can be seen from Table 1 and 2. In general the results obtained for larger training set are better. In both PCA and ICA, for some cases accuracy was more than 90%. Similar trend was shown in [7] where using large training set gave an accuracy of 95.67%.

Image intensity normalization plays an important role in the performance of feature extraction and classification. Keeping all other configurations (dataset size, histogram equalization, and scaling) same, performing intensity normalization improved performance in most cases for both ICA and PCA. For example 1^{st} and 5^{th}, 3^{rd} and 7^{th},9^{th} and 13^{th} , 11^{th} and 15^{th} rows of both ICA and PCA show improvement for performing intensity normalization. On an average for ICA there is ~14% increase in performance and for PCA the improvement is ~13%.

The results in rows 1 and 3, 2 and 4, 5 and 7, 6 and 8, 9 and 11, 10 and 12, 13 and 15, 14 and 16 of both ICA and PCA shows the effects of histogram equalization while all other configurations are fixed. With the exception of two cases in PCA (10 and 12, 13 and 15) in all other cases there was improvement. The average improvement for ICA was ~7% and PCA ~6%.However, in some cases good accuracy (≥90%) was achieved even without using histogram equalization. This might be because of some loss of information due to down sampling and histogram equalization. Also the input images did not have enough illumination variation to affect the performance of classification. [7] also applies histogram equalization after face normalization. However, they used a slightly larger image (64x96).

Scaling seems to have more impact on the performance of ICA. For ICA, accuracy due to scaling improved by about 47% on average. Input scaling does not seem to improve the performance of PCA especially when intensity normalization or histogram equalization is performed.

6. Conclusion

In this report, gender classification using dimensionality reduction techniques namely ICA and PCA along with SVM is presented. Appearance-based approach is taken with the assumption that the input images are aligned and free of background clutter. Features are extracted after performing dimensionality reduction and classification is performed using SVM with Gaussian RBF kernel. Results for varying dataset size, image intensity normalization, histogram equalization, and input scaling are also presented. Both ICA and PCA can achieve very good accuracy with different combinations of the above parameters. The results that were obtained are similar to some of the works discussed in Section 2.

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. Currently a naïve grid search is performed for finding optimal values for C and γ, a more sophisticated search will definitely improve the performance of gender classification.

References

[1] Alessandro Cellerino, Davide Borghetti and Ferdinando Sartucci, Sex differences in face gender recognition in humans, Brain Research BulletinVolume 63, Issue 6, , 15 July 2004, Pages 443-449.

[2] M. Toews and T. Arbel, Detecting, localizing and classifying visual traits from arbitrary viewpoints using probabilistic local feature modeling, *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 154-167, Rio de Janeiro, Brazil, Oct. 2007.

[3] 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.

[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.

[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

[6] Moghaddam, B.; Ming-Hsuan Yang, Learning gender with support faces, *Transactions on Pattern Analysis and Machine Intelligence* , vol.24, no.5, pp.707-711, May 2002

[7] Amit Jain, Jeffrey Huang, Integrating Independent Components and Support Vector Machines for Gender Classification,ICPR , pp. 558-561, 2004.

[8] A. Hyvärinen. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Transactions on Neural Networks* 10(3):626-634, 1999.

[9] The Psychological Image Collection at Stirling (PICS) repository. http://pics.stir.ac.uk/cgi-bin/PICS/New/pics.cgi

[10] The FastICA Package for Matlab. http://www.cis.hut.fi/projects/ica/fastica/

[11] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm