Gender Identification Using Frontal Facial Images

Amit Jain, Jeffrey Huang and Shiaofen Fang
Department of Computer and Information Sciences,
Indiana University-Purdue University-Indianapolis, IN 46222
{ajain, huang, sfang}@cs.iupui.edu

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 classifiers with performance superior to existing gender classifiers. We experiment on 500 images (250 females and 250 males) randomly withdrawn from the FERET facial database. Independent Component Analysis (ICA) is used to represent each image as a feature vector in a low dimensional subspace. Different classifiers are studied in this lower dimensional space. Our experimental results show the superior performance of our approach to the existing gender classifiers. We get a 96% accuracy using Support Vector Machine (SVM) in ICA space.

1. INTRODUCTION

The face is an important biometric feature of human beings. Faces are accessible 'windows' into the mechanisms that govern our emotional and social lives [1]. A successful gender classification method has many potential applications such as human identification, smart human computer interface, computer vision approaches for monitoring people, passive demographic data collection, etc. This paper deals with gender classification based on frontal facial images.

Classification using images is in general difficult because of the inherent variability of image formation process. Some of the challenges and possible solutions in using facial images for classification are discussed in some detail by Chellapa et. al [2].

The earliest attempt to use computer vision techniques for gender classification was based on neural networks [4]. Golomb et. al [5] trained a fully connected two-layer network, called SEXNET, to identify gender from facial images. Tamura et al [6] used a multi layered neural network to identify gender from face images of different resolutions. Gutta et. al [7] proposed a hybrid approach that consists of a collection of neural networks and decision trees. A PCA based image representation was used along with radial basis functions and perceptron networks by Abdi et. al [8]. O' Toole et al [9] have also used PCA and neural networks and have reported good performance. Moghaddam et. al [10] investigated the use of SVMs for gender classification.

All above discussed techniques are appearance-based methods, that is, they learn the decision boundary between male and female classes from training images, without extracting any geometrical features such as distances, face width, face length, etc. V. Bruce et al [11] identified 73 points on a face image and discriminant analysis was used to classify gender using point to point distances. Brunelli and Poggio [12] compute 16 geometric features (like pupil to eye brow separation, eye brow thickness, etc.) from the frontal images of a face. These features are used for identifying the gender.

Bebis et. al [13] used genetic feature subset selection from frontal images. Experiments were carried on four different classifiers after features are evolved (selected) using genetic algorithm. A multimodal gender classification approach using images and voice is proposed in [14].

The variety of methods published in the literature show that there is not a unique or generic solution to the gender classification problem. In this paper, we propose two different methods for identifying the gender using facial images based on Independent Component Analysis (ICA) for projection of data into lower dimensional space. Different classifiers are studied in this low dimensional space for gender classification. The outline of the paper is as follows: Section 2 provides framework for image representation using Independent Components (ICs). Section 3 introduces the classifiers we have experimented. Experimental results are discussed in Section 4 and conclusions are presented in Section 5.

2. IMAGE REPRESENTATION

In general the first step in any recognition process is to choose good discriminatory features. A well-known problem in pattern recognition is the "curse of dimensionality"- more features do not necessarily imply a better classification success rate. Feature extraction in pattern recognition involves the derivation of salient features from the raw input data in order to reduce the

0-7803-9332-5/05/$20.00 ©2005 IEEE

amount of data used for classification and simultaneously provide enhanced discriminatory power.

A fundamental problem in digital signal processing is to find a suitable representation for a signal. Usually different signal representations are based on linear transformations of the signals onto different bases. The transformations help representing the signal in a lower dimensional space. We use Independent Component Analysis for facial images representation [15].

Using ICA, one tries to model the underlying data so that in the linear expansion of the data vectors the coefficients are as independent as possible. ICA bases of the expansion must be mutually independent. Let **s** be the vector of unknown source signals and **x** be the vector of observed mixtures. If **A** is the unknown mixing matrix, then the mixing model is written as **x=As**. It is assumed that the source signals are independent of each other and the mixing matrix is invertible. Based on these assumptions and the observed samples, ICA tries to find the mixing matrix **A** or the separating matrix **W** such that

$$\mathbf{u} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s} \tag{1}$$

is an estimation of the independent sources [16].

We use a fast and computationally simple fixed-point rule developed by Hyvarinen and Oja [17]. Let $\mathbf{x}(t) = [\mathbf{x}_1(t),\ldots, \mathbf{x}_n(t)]^T$ be the n-dimensional t^{th} data vector. In ICA, the data vector $\mathbf{x}(t)$ can be expanded as

$$\mathrm{x}(t) = \mathrm{As}(t) + \mathrm{n}(t) = \sum_{i=1}^{m} \mathrm{s}_i(t)\mathrm{a}_i + \mathrm{n}(t) \tag{2}$$

where $\mathbf{s}(t) = [\mathrm{s}_1(t), \ldots, \mathrm{s}_m(t)]^T$ are the m independent source signals, $\mathbf{A} = [\mathrm{a}_1, \ldots, \mathrm{a}_m]^T$ is a n-by-m matrix, called the mixing matrix and $\mathbf{n}(t)$ is additive noise. The number of independent components, m, is usually $m \leq n$. The expansion in Eqn (1) is determined by enforcing that the coefficients $\mathrm{s}_i(t)$, where $i=1..\ m$, are mutually independent. Hence the basis vectors a_i are generally not mutually orthogonal.

The independent components are computed by finding the separating matrix **W** so that

$$\mathbf{y}(t)=\mathbf{W}\mathbf{x}(t) \tag{3}$$

becomes an estimate of $\mathbf{s}(t)$. As a preprocessing step, the observed samples are whitened, and now the separating equation is

$$\mathbf{y}(t)=\mathbf{W}\mathbf{v}(t) \tag{4}$$

where $\mathrm{v}(t)=\mathbf{V}\mathbf{x}(t)$ and **V** is the whitening matrix. In the fixed-point algorithm described in [18], some initial values for the columns w_i $(i=1..\ m)$ are randomly chosen. The fixed-point algorithm iteratively computes a new $(k+1)^{th}$ estimate for w_i

$$\mathrm{w}_i^*(k+1) = E\{\mathrm{v}g(\mathrm{w}_i(k)^T\mathrm{v}) - g'(\mathrm{w}_i(k)^T\mathrm{v}) \cdot \mathrm{w}_i(k) \tag{5}$$

$$\mathrm{w}_i(k+1) = \frac{\mathrm{w}_i^*(k+1)}{\left\|\mathrm{w}_i^*(k+1)\right\|} \tag{6}$$

Here, $E\{\cdot\}$ denotes the mathematical expectation and $g(\cdot)$ can be any odd, sufficiently regular nonlinear function and $g'(\cdot)$ is its first derivative. Most common choice of $g(\cdot)$ is $g(u)=u^3$ or $\tanh(u)$. By choosing $g(u)=u^3$, the kurtosis criterion [16] is maximized. The choice of $g(u)=\tanh(u)$ is motivated as it is a robust non-linearity that grows less than linearly. It has been shown in [18] that $\mathrm{w}_i(k)$ converges (up to the sign) to one of the columns of **W**. For preventing w_i $(i=1,\ldots, m)$ from converging to the same directions, they are orthogonalized against each other. The basis vectors $\mathbf{a}_\mathbf{i}$ of ICA are obtained from w_i by

$$\mathbf{a}_i=\boldsymbol{\Xi}\boldsymbol{\Lambda}^{1/2}\,\mathbf{w}_i \tag{7}$$

where Ξ contains the eigenvectors and the diagonal matrix Λ has the corresponding eigenvalues as its elements.

Hence, by use of ICA, a sparse representation of the data set is generated. This representation of the images in lower dimensional space is used for the task of gender classification.

3. CLASSIFICATION

We have experimented on three different classifiers in the lower dimensional space defined by the ICA vectors. The classifiers we have experimented on are the cosine classifier, linear discriminant classifier (LDA) and the support vector machine (SVM).

Since the ICA projects the images into a space whose bases vectors are not orthogonal, hence Euclidean distance is not a good metric of similarity. Imagine the image vectors are normalized and located over an hyper-sphere surface, the angle between them represents the distance above its surface. This angle is calculated by

$$\cos(x, y) = \frac{x^T y}{\| x \| \| Y \|} \tag{8}$$

The objective of the LDA is to find projection $y=W^T x$ (where x is the input image and W is the projection matrix), that maximizes the ratio of the between-class scatter and the within class scatter [19]. Let $\omega_1, \omega_2, \ldots, \omega_L$ and $N_1, N_2,\ldots N_L$ be the classes and number of images in each class, respectively. Let $M_1, M_2,\ldots M_L$ and M be the means of the classes and the grand mean respectively. Then, the within and between class scatter matrices S_w and S_b are defined as

$$\begin{aligned} S_w &= \sum_{i=1}^{L} P(\omega_i)\Sigma_i \\ &= \sum_{i=1}^{L} P(\omega_i)E\left\{[X - M_i][X - M_i]^t \mid \omega\right\} \end{aligned} \tag{9}$$

and

$$S_b = \sum_{i=1}^{L} P(\omega_i)\left(M_i - M\right)\left(M_i - M\right)^t \qquad (10)$$

where $P(\omega_i)$ and Σ_i are a priori probability and the covariance matrix of class ω_i respectively. LDA derives a projection matrix W that maximizes the ratio $|W^t S_b W| / |W^t S_w W|$. This method defines *L-1* basis vectors called Fisherfaces, also known as Most Discriminating Features (MDF). Since gender classification is a two class classification, hence the LDA space is one-dimensional.

Support Vector Machines (SVM) are primarily two class classifiers that are known to have robust performance for learning linear or non-linear decision boundaries. Given a set of points that belong to either of the two classes, SVM finds the hyper-plane leaving the largest possible fraction of points of the same class on the same side and maximizing the distance of either class from the hyper-plane. We use the SVM toolbox developed by Ahalt et al [20].

4. EXPERIMENTS

Our methodology for gender classification is shown is Figure 1. The use of facial images for gender classification starts by normalizing the face image to account for geometry and illumination changes using the eye location [21]. The images are normalized using the location of eye and maintaining constant distance between the eyes. The normalized images are cut into elliptical region that contains only the face. Histogram equalization is done to each normalized image to account for the different lighting conditions. These normalized face images are used for the gender identification using appropriate image representation and classification algorithms. We define the Independent Components derived from the preprocessed training images as an optimal basis along which faces are projected, leading to a compact and efficient coding. Let $X \in R^n$ be a vector representation of the image obtained by row stacking. $T = [\theta_1, \theta_2, \ldots \theta_m]$, $T \in R^{nxm}$ be the optimal basis derived using Independent Components. Then the new feature set $U \in R^m$ is defined as $U_i = T^t X_i$ where X*i* is the *i*-th training image and U_i is the corresponding feature vector.

The data set used is the FERET facial data set consisting of frontal shots of persons at a resolution of 256x384 with 256 gray levels. The normalized images are at a resolution of 64x96. The dataset consists of 500 images, 250 of male subjects and 250 of female subjects. Out of these images, 200 images (100 male and 100 female subjects) are used for training. The remaining 300 images are used for testing.

200 independent components were found to describe this lower dimensional space. Fig 2 shows the reconstructed images of 4 independent components. Different classifiers are used in this sparse space. The cos-distance classifier is a nearest neighbor classifier. Since the goal is to identify the gender based on facial images, it is a 2-class classification problem and hence, LDA method defines one basis vector, which is sufficient for efficient classification. The class of the test case is defined as the class whose mean is nearest to the test sample. Fig. 3 shows the reconstructed image of MDF that is used for the classification. Using the experiments described, the average performance for the ICA + COS classifier was found to be 85.33%. For the ICA + LDA, the accuracy was 93.33%. For ICA + SVM, the accuracy observed is 95.67%.

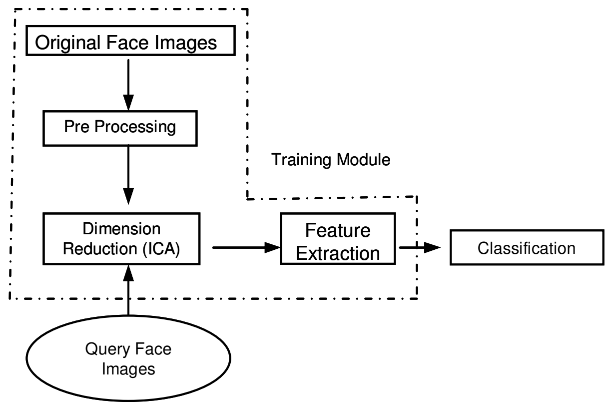

Figure 1. System overview

Figure 2. The first four IC vectors

Figure 3. Reconstructed Image of MDF

We also experiment on the effect of the size of the training set. Table 1 shows the performance accuracy when the training sample size is varied from 50 to 200 in discrete increments of 50.

Comparing our methodology with existing research, we get better performance than already reported techniques. ICA projects the images into a lower dimensional subspace and the best performance is obtained by using a SVM in this subspace. LDA also gives good classification accuracy since it extracts the most discriminating features (MDF) from the dataset. The motivation for using MDF for classification is that it discounts those features that are irrelevant to classification (like the lighting direction).

Table 1. Effect of training sample size on classification accuracy.

Training Set Size	Classifiers ICA+COS	ICA+LDA	ICA+SVM
50	60.67%	64.67%	68.30%
100	71.67%	73.67%	76.00%
150	80.33%	83.00%	86.67%
200	85.33%	93.33%	95.67%

5. CONCLUSIONS

In this paper, we introduce ICA as an image representation technique for gender classification. ICA is used to represent face images as a linear superposition of basis functions. It learns basis images that are adapted to the data ensemble. ICA is used to reduce the dimensions of the data. Different classifiers are used for the task of classification by finding discriminating features for the data. This paper deals with use of two-dimensional facial images for gender classification. Efficient use of 3-D face data might achieve better results at a lower computational expense. Our future work concentrates on exploring 3-D face data for applications like gender classification and face recognition.

6. REFERENCES

[1] P. Ekman, T.S. Huang, T.J. Sejnowski, and J. C. Hager, "Final Report to NSF of the Planning Workshop on Facial Expression Understanding", 1993.

[3] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces: A Survey", *Proc. of IEEE*, vol. 83, no. 5, pp. 705-740, 1995.

[4] G. Cottrell, J. Metcalfe, and E. Face, "Gender and Emotion Recognition using Holons", 1991.

[5] B. Golomb, D. Lawrence and T. Sejnowski, "Sexnet: a neural network identifies sex from human faces", *Advances in Neural Information Processing Systems*, pp. 572-577, 1991.

[6] S. H. Tamura, Kawai, and H. Mitsumoto, "Male/Female Identification from 8 x 6 Very Low Resolution Face Images by Neural Network", *Pattern Recognition*, vol. 29, no. 2, pp. 331-335, 1996.

[7] S. Gutta, H. Weschler, and P. J. Phillips, "Gender and Ethnic Classification of Human Faces using Hybrid Classifiers", *Proc. of IEEE International Conf. on Automatic Face and Gesture Recognition*, pp. 194-199, 1998.

[8] H. Abdi, D. Valentin, B. Edelman, and A. O'Toole, "More about the difference between men and women: evidence from linear neural networks and the principal component approach", *Perception*, vol. 24, pp. 539-562, 1995.

[9] A. O'Toole, H. Abdi, K. Deffenbacher and D. Valentin, "A low-dimensional representation of faces in higher dimensions of space", *Journal of the Optical Society of America,* vol. 10, pp. 405-411, 1993.

[10] B. Moghaddam and M. Yang, "Gender classification with Support Vector Machines", *4th IEEE Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 306-311, 2000.

[11] A. Burton, V. Bruce and N. Dench, "What's the Difference Between Men and Women? Evidence from Facial Measurements", *Perception*, vol. 22, pp. 153-176, 1993.

[12] R. Brunelli and T. Poggio, "Hyberbf Networks for Gender Classification", *DARPA Image Understanding Workshop*, pp. 311-314, 1992.

[13] Z. Sun, G. Bebis, X. Yuan, and S. J. Louis, "Genetic Feature Subset Selection for Gender Classification: A Comparison Study", *IEEE Intl. Conf. on Image Processing*, Aug. 2002.

[14] L. Walavalkar, M. Yeasin, A. Narasinmhmurthy, and R. Sharma, "Support Vector Learning for Gender Classification Using Audio Visual Cues: A Comparison", International Journal of Pattern Recognition and Artificial Intelligence, vol. 17, no. 3, pp. 417-439, 2003.

[15] P. Comon , "Independent component analysis, a new concept?, " *Signal Processing*, vol. 36, pp. 287-314, 1994

[16] A. Hyvarinen and E. Oja, "Independent component analysis: A tutorial," [Online] Available: http://www.cis.hut.fi/projects/ica

[17] A. Hyvarinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," Neural Computation, vol. 9, pp. 1483-1492, 1997.

[18] A. Hyvarinen, "A family of fixed-point algorithms for independent component analysis," *Proc. IEEE Int'l. Conf. on Acoustics, Speech and Signal Processing (ICASSP'97)*, pp. 3917-3920, Germany, 1997.

[19] R. Duda, P. Hart and D. Stork, *Pattern Classification*. Jon-Wiley, 2001.

[20] J. Ma, Y. Zhao and S. Ahalt, *OSU SVM Classifier Matlab Toolbox*, [Online] Available: www.eleceng.ohio-state.edu/~maj/osu_svm/

[21] J. Huang and Wechsler, H., "Eye Detection Using Optimal Wavelet Packets and RBFs", *Int. Journal of Pattern Recognition and Artificial Intelligence* (IJPRAI), 13 (6), pp.1009-1026.