SAPIENZA
UNIVERSITÀ DI ROMA

Gender Classification Of Human Faces

Nadine Abu Rumman
Dipartimento di Informatica e Sistemistica Antonio Ruberti
Via Ariosto 25, 00185, Rome, Italy
aburumman@dis.uniroma1.it

September 03, 2013

Contents

1 INTRODUCTION

Face is one of the most important part in human. Arguably, images of human face provide a lot of information such as age, gender, ethnicity, identity, expression...etc. Therefore gender recognition by face is consider one of the active research area of computer vision; there are huge numbers of applications where gender recognition can be useful, like biometric, authentication, high-tech surveillance, security systems, criminology and augmented reality...etc. Not surprisingly, thus, that a very large number of studies has investigated gender recognition from face perception in humans [1].

Gender recognition is consider as one of the interesting visual task for an extremely social animal like us humans, many social interactions critically depend on the correct gender perceptions. In contrast, as the field of artificial life emerged, researchers began applying principles such as gender recognition to achieve socially interactive robot behavior [2]. Toward to implement a social competence in a humanoid robot, first task has to be detecting human faces, then recognize and identify more information about detected faces such as gender.

Gender recognition can be regarded as classification problem of detected faces into two classes (male and female). In the literature there are three popular approaches to detection of features for gender classification: Principal Components Analysis (*Eigenfaces*), Linear Discriminant Analysis(*Fisherfaces*) and Local Binary Pattern Histogram (*LBP*), where previous work in machine learning focused on different types of classifiers such as Nearest-Neighbor Classifiers (*NN*) versus Radial Basis Function networks (*RBF*), Adaboost-Based Classifiers or Support Vector Machine (SVM).

This project addresses the problem of gender classification using frontal facial images as training images, where the goal of this project is to automatically detect faces on images or video, quickly and reliably classify the gender of the detected faces. Our gender classifier has been trained using two different datasets; more details are in the evaluation section.

Fisherfaces method [3] is used in our implementation to represent each image as a feature vector and project each feature vector (linearly projecting) from the high-dimensional image space to a significantly lower dimensional feature space which insensitive both to variation in lighting direction and facial expression, the projection directions are nearly orthogonal to the within-class scatter, projecting away variations in lighting and facial expression while maintaining discriminability. Fisherfaces, a derivative of Fishers Linear Discriminant (FLD), which maximizes the ratio of between-class scatter to that of within-class scatter. On above of Fisherfaces extraction features method, nearest neighbour based on Euclidean distance is used as a classifier, in order to classify a new image.

In literature, Eigenface method considers as quite popular technique among pattern classification techniques for solving the face recognition problem. Also Eigenface is based on linearly projecting the image space to a low dimensional feature space [3]. However, the Eigenface method, which uses Principal Components Analysis (*PCA*) for dimensionality reduction, yields to projection directions that maximize the total scatter across all classes, i.e., across all images of all faces. In choosing a projection, which maximizes total scatter, *PCA* retains unwanted variations due to lighting.

Here a brief explanation of the main pipeline of solving this classification problem [4], gender classification system consists features extraction and classifier module, In the training phase features extraction module reduces the data by measuring certain features of the training face images that are useful for classification. After this features are stored in the database. While in the testing phase, feature of the test face image are extracted and these extracted features are used by the classifier to classify the image with the help of the database, which is created during the training phase, and makes the final decision.

This report is organized as follows: Section 2 defines the gender classification problem, with formally specifying the inputs and outputs; Section 3 discusses the overview of gender classification system as learning system, gives more details about the feature extraction method that we defined, show the target function employed in this project, and describe the representation of the input images; in Section 4 the database, evaluation and experiments is described in detail; finally section 5 concludes the work and its future scope.

2 PROBLEM DEFINITION

Gender classification consider as binary classification problem, where there is a single target variable $t \in \{0,\ 1\}$ such that t = 1 represents class *Female* and t = 0 represents class *Male*, this problem can simply stated: Given a set of face images labeled with the persons gender identify (*learning set*) and an unlabeled set of face images (*test set*), seeks to identify each person's gender in the test images.

In this project we approach this problem within the appearance-based pattern classification, considering each of the pixel values in a sample image as a coordinate in a high dimensional space (*image space*). More formally let us consider a set of N samples images each of size mxn, training images as I_{train} $=X_i\{x_1,x_2,...,x_m\}$ and assume that each image belong to one of classes $\{X_{male}$ $,X_{female}\}$, and the set of testing images $I_{test}= E_i$; $\{e_1,e_2,...,e_n\}$. At this point, we have not chosen a mathematical representation of the image, and therefore the notation used is abstract.

Given those N individuals and M training face images, the algorithm to learn the model is described as follows: Initialize N sample sets, for each training image I_i, extract the feature points, compute feature descriptors for each feature point and construct the appearance model for each individual.

Fisherfaces is consider as supervised learning, takes the class labels of the training images into account and maps them to a subspace that maximizes between class variance, the intuition here is that Fisherfaces preserves information that differentiates class, which makes sense in gender classification problem [3]. When an unknown face image is presented to the model, it extracts a feature vector from the image, and then searches for the closest feature vector of a known face, which allow to obtain the class label; Figure1: shows the algorithm of face gender classification workflow.

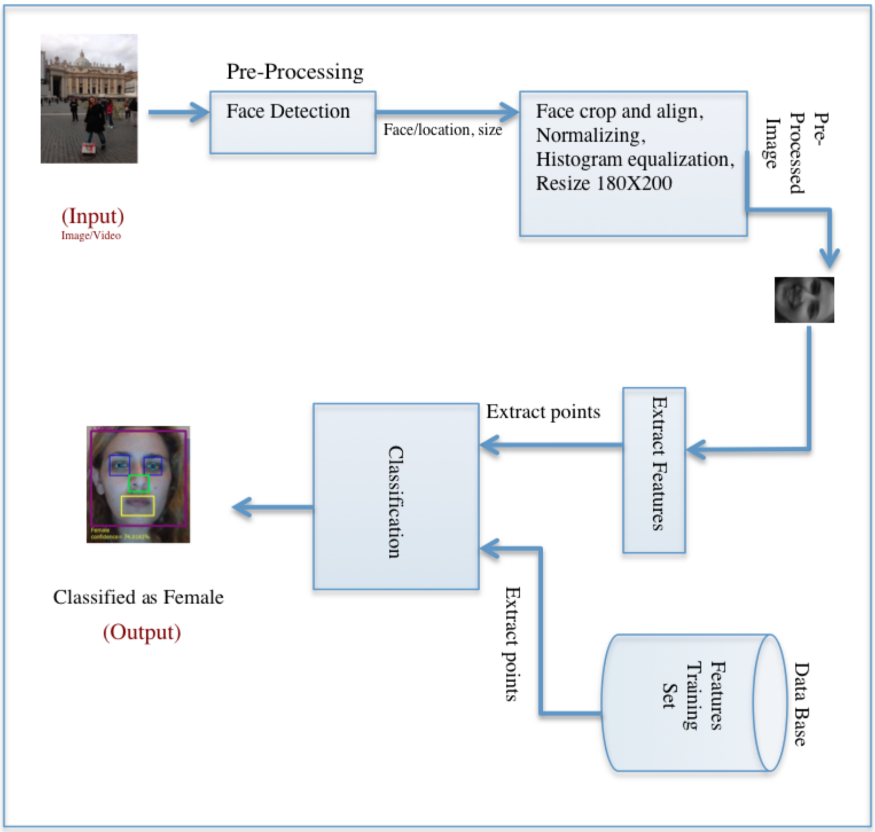


Figure 1: Gender classification workflow.

3 GENDER CLASSIFICATION SYSTEM

Gender classification system in this project consists of three main modules: automatically detect faces on images or video, feature extraction/selection, and classification. Extract facial images from image/video using face detector; LBP Cascade detector (Local Binary Pattern; which is extended set of Haar-Cascade) used here for rapid face detection [5]. After the face is detected in the image, some pre-processing may be applied. It helps to reduce the sensitivity of the classifier to variations such as illumination, pose and detection inaccuracies.

Pre-processing that may be applied to the face image include:

- Normalize for contrast and brightness using histogram equalization [6]; histogram equalization is a very simple method of automatically standardizing the brightness and contrast of facial image.

- Geometric alignment (like eye- alignment either manually or using automatic method).
- Crop to remove the background and resizing images to a standard size, which is *180X200* pixels in our case, why? Because of the natural face proportion, where height suppose to be bigger than width, and we try to keep the aspect ratio unchanged (as much we can).

The purpose of normalize the input image is to make the condition of the processing face as close as possible with the ones stored in the database see the figure below.

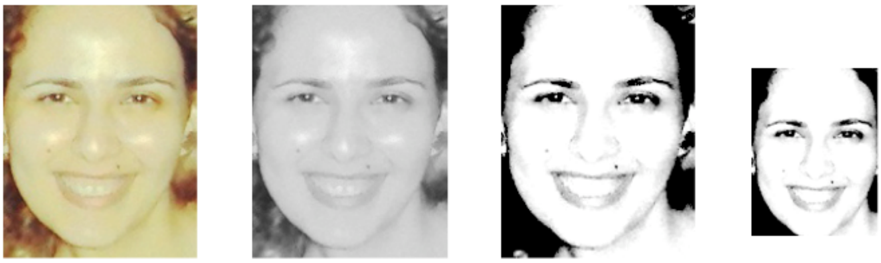


Figure 2: Illustrates the Pre-Processed stage.

3.1 IMAGE REPRESENTATION

We worked on the vector representation of the image, for example; for each image of size $i \times j$ pixels is represented as a vector (concatenation of the image columns) in an $n = i \cdot j$ dimensional space, so each 2D image data (image space) represented as vector (subspace of the image space which called face space), consider an image with *180X200* pixels, it represented as *36,000X1* vector image.

Face as 2D image data has very high dimensionality and owing to the curse of dimensionality. Dimensionality reduction techniques should be applied before using any machine learning method for classification [3]. In order to reduce the data size, while keeping all information necessary for classification of novel data. This means project the training data onto a much smaller subspace to enable an efficient image handling.

The purpose of this dimensionality reduction is also to find meaningful low-dimensional structures hidden in high-dimensional observations, in other words finds an optimal transformation that suitably represent the data [7]. Therefore, the goal of subspace method is to find a projection that transforms the training data (images), so a new unknown image can be efficiently classified.

Let M be the number of images of two different classes in the training set, those images represented as aligned columns of the matrix X=$\{x_1,x_2,...,x_m\}$$\in$R^{nxm}. When performing the dimensionality reduction in the dataset X in the high-dimensional space, X is mapped to a lower-dimensional subspace giving the matrix Y. The main point is that by this projection T (consider as the optimal transformation), where the redundancy in the data is reduced while as much information as possible is preserved. Thus, linear feature extraction is defined as a linear projection from a n-dimensional image space onto a r-dimensional subspace (feature space):

$$T\colon R_n \rightarrow R_r$$
$$Y = T^T X \tag{1}$$

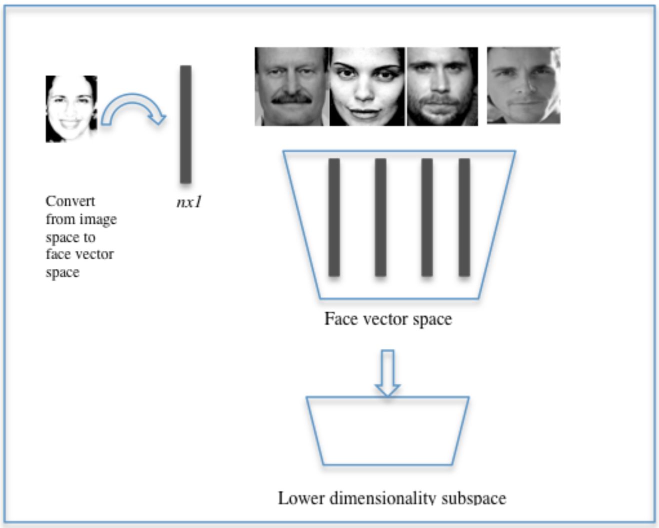

Convert from image space to face vector space
nx1
Face vector space
Lower dimensionality subspace

Figure 3: Visualize the dimensionality reduction.

3.2 FEATURES EXTRACTION

This stage is to extract salient features from the face; the output of this stage is a vector storing information about the salient feature points of each face images.

Previous studies showed two categories of feature extraction approach:

- Appearance feature based (global/ local features, such as *LDA*) [8], which is working based on some operation or transformation performed on the pixels of an image. This can be done at the global (holistic) or local level. At the local level, the face may be divided into defined regions such as eyes, nose and mouth.
- Geometrical feature based (local features, such as *LBP*), is based on measurements of facial landmarks like eyebrow, noise and mouth, geometric relationships between these points are maintained and then the process of extracting the point locations is preformed.

The method used in this project is implemented based on appearance feature based feature approach, where the geometric relationships are naturally maintained [3], this is consider as advantageous when the gender discriminative features are not exactly known.

As mentioned before dimensionality reduction is optimal transformation to find meaningful low-dimensional structures hidden in high-dimensional observations, in other words its a process of mapping the original data into more effective features (seeks to extract those salient features from the images), the method which used here is based on Linear Discriminant Analysis technique called Fisher Linear Discriminant Analysis (*Fisherfaces*).

3.3 SYSTEM STRUCTURE

In this subsection we will explain the procedure followed by Fisherfaces method [3], where Fisherface recognize the variation within each class lies in a linear subspace of the image space. Hence, the classes are convex, and, therefore, its linearly separable, then perform dimensionality reduction using this linear projection, and still preserve linear separability.

This method selects linear projection/transformation W to map images from n-dimensional (image space) onto a r-dimensional subspace (feature space), in such a way that the ratio of the between-class scatter and the within class scatter is maximized.

Given training images as I_{train}=$X_i\{x_1,x_2,...,x_m\}$ with orthonormal columns, and assume that each image belong to one of classes $\{X_{male},\ X_{female}\}$:

Where the between-class scatter matrix defined as:

$$\mathrm{S}_B=\sum_{i=1}^{c}\mathrm{N}_i(\mu_i\text{-}\mu)\ (\mu_i\text{-}\mu)^T \tag{2}$$

And the within-class scatter matrix be defined as:

$$\mathrm{S}_W=\sum_{i=1}^{c}\ \sum_{x_k\in X_i}\ (\mathrm{x}_k\text{-}\mu_i)\ (\mathrm{x}_k\text{-}\mu_i)^T \tag{3}$$

Where μ is the mean of all training images, μ_i is the mean image of class X_i, and N_i is the number of samples in class X_i. If S_W is nonsingular, the optimal projection W_{opt} is chosen as the matrix with orthonormal columns which maximizes the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within class scatter matrix of the projected samples, in other word the target function is to maximize the scatter between classes and minimize the scatter within classes:

$$W_{opt}=arg\ max\ \frac{|W^T S_B W|}{|W^T S_W W|} \tag{4}$$

$$=[w_1 w_2 w_3 .... w_m]$$

Where $\{\mathbf{w_i}|\mathbf{i=1,2,,m}\}$ is the set of generalized eigenvectors of S_B and S_W corresponding to the m largest generalized eigenvalues $\{\lambda_i|i=1,2,,m\}$

$$S_B\mathbf{w_i}=\lambda_i S_w \mathbf{w_i},\ \mathrm{i=1,2,..,m} \tag{5}$$

Note that there are at most *c-1* nonzero generalized eigenvalues, and so an upper bound on m is *c-1* , where c is the number of classes, because we have just two classes in our case male or female, thus the upper bound of m is equal to one [m=1].

In order to enforce nonsingular in S$_W$ matrix, we have to project the image set to a lower dimensional space in such way that the resulting within-class scatter matrix S$_W$ is nonsingular. This is achieved by using PCA to reduce the dimension of the feature space to N - c, and then applying the standard FLD defined by (4) to reduce the dimension to c - 1. More formally, $Wopt$ is given by:

$$(W_{opt})^T = (W_{fld})^T (W_{pca})^T \quad (6)$$

where

$$W_{pca} = argmax_W |W^T S_T W| \text{ where } S_T = \sum_{k=1}^{n} (x_k - \mu)\ (x_k - \mu)^T \quad (7)$$

$$W_{fld} = argmax_W \frac{|W^T (W_{pca})^T S_B W_{pca} W|}{|W^T (W_{pca})^T S_W W_{pca} W|} \quad (8)$$

Note:

- Fisher linear discriminant, actually it is not a linear discriminant, but allows efficient linear model for classification.
- The optimization for W_{pca} is performed over n x$(N$ - $c)$ matrices with orthonormal columns, while the optimization for W_{fld} is performed over $(N$ - $c)$ x m matrices with orthonormal columns. In computing W_{pca} , we have thrown away only the smallest c - 1 principal components.
- All the images were normalized to have zero mean and unit variance, as this improved the performance of this method.

3.4 THE ALGORITHM

Given a training database of human facial images labeled with gender, train an automated system to identify a person gender in a new image (*test set*), how?! Fisherface classifies a face image by looking for the training image that's closest to it in the PCA subspace. Finding the closest training example in a learned subspace is a very common AI technique. It's called Nearest Neighbor matching which is computes distance from the projected test image to each projected training example. The distance basis here is "Squared Euclidean Distance" used to find the closest image, and then finds the class associated with it.

Given: Suppose we have a set of N sample images I_{train} =$\{X_i\}$; $\{x_1, x_2, ..., x_m\}$ each of size $m \times n$, these training images is converted to column vectors. The training set in vector form is defined as $\Gamma = [\Gamma_1, \Gamma_2, \Gamma_3, ..., \Gamma_n]$.

Normalize the training set:

- Calculate the average face of the training set (Γ_i) as μ:
  $\mu = \frac{1}{N} \sum_{i=1}^{N} \Gamma_i$

- Normalize the training set by subtracting mean from each sample, A=$[\Phi_1, \Phi_2, \Phi_3, ..., \Phi_n]$,where Φ_is are defined as Φ_i= $\Gamma_i - \mu$

Projection and Feature Extraction:

Compute PCA

- Apply PCA on A to get a set of N eigenvectors u_k and their associated eigenvalues λ_k which best describes the distribution of the data.

Compute FLD

- Compute the mean of female class and male class:

  $\mu_{male} = \frac{1}{N_{male}} \sum_{n \in C_{male}} \Phi_n$

  $\mu_{female} = \frac{1}{N_{female}} \sum_{n \in C_{female}} \Phi_n$

- Compute between-class scatter matrix

  $S_B = (\mu_{female} - \mu_{male})(\mu_{female} - \mu_{male})^T$

- Compute within-class scatter matrix

  $S_W = \sum_{n \in C_{male}} (\Phi_n - \mu_{male})(\Phi_n - \mu_{male})^T + \sum_{n \in C_{female}} (\Phi_n - \mu_{female})(\Phi_n - \mu_{female})^T$

- Project into *(N-c)* by *(N-c)* subspace using PCA

- Differentiating (4) with respect to w, we need to find w that maximized W$_{opt}$ is when : (w^{T}S$_B$w)S$_W$w = (w^{T}S$_W$w)S$_B$w

- Generalized eigenvalue decomposition to get u_k

- Those values can be obtained from calculation of eigenvectors and eigenvalues of the covariance matrix $A^T A$ and then multiplied by A, to get Av_i as the eigenvectors, where covariance matrix is :
  C=$\frac{1}{N} \sum_{i=1}^{N} \Phi_i^T \Phi_i$

- The Eigenfaces can be calculated as,
  $u_i = \sum_{k=1}^{N} v_{lk} \Phi_k$, where *l=1,2,3 ..... N*

- Project the training images to features space, where
  $\Psi_i = \sum_{j=1}^{M} w_j u_j$, where $w_j = u_j^T \Phi_i$ (features subspace)

Nearest-Neighbor Classification:

- Obtain a new projected face image {in test set} (Γ_{new}) and transform it into its features space as $w_k = u_k^T(\Gamma_{new} - \mu)$

- Find class label that minimizes the Euclidean distance with training images as $\varepsilon = min\|w_k - w_{ki}\|$

- Classify the test image as class male or female, if $\varepsilon \leq \Theta$, where Θ is a threshold.

4 EXPERIMENTAL EVALUATION

Here we describe the experiments that we did with our learning system in detail:

4.1 DATA SETS

To train our classification algorithm for this application [which is gender recognizer], we used two different databases. Georgia Tech Database which is publicly available online hosted at the website of Ara Nefuan [9]. Where in this database for each person, there are 15 images with various angles; we used subset of this database to perform our gender classification algorithm (with 135 males and 105 females). Secondly, we have constructed Custom Database using frontal face images of celebrities and friends, which contains 141 images (with 71 males and 70 females). In order to evaluate our algorithm, collected images from web have been used as test set, this test set contains 15 images of frontal face (we use this data set to examine the generalization ability of the gender discriminating model), in addition to real-time evaluation from video done for numbers

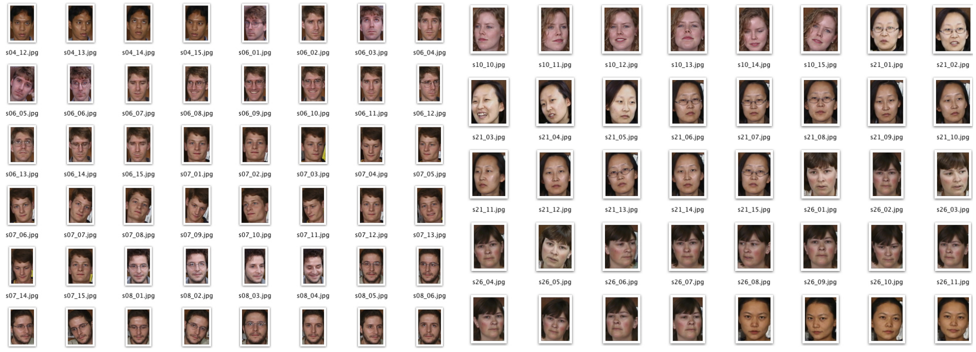


Figure 4: Snapshot of Georgia Tech Database.

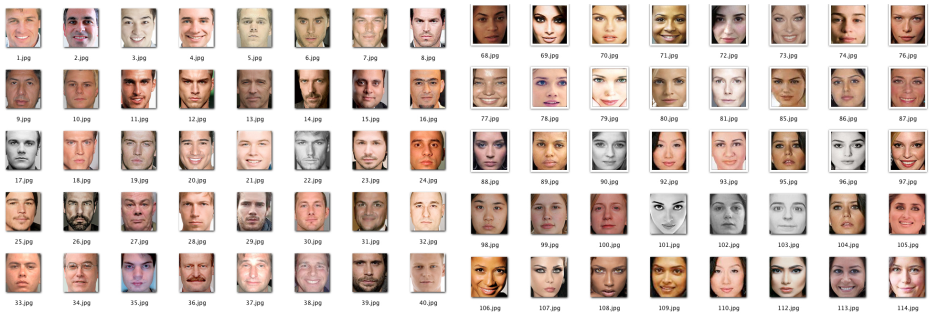


Figure 5: Snapshot of Custom Database.

of friends, and colleagues, see the figure below.

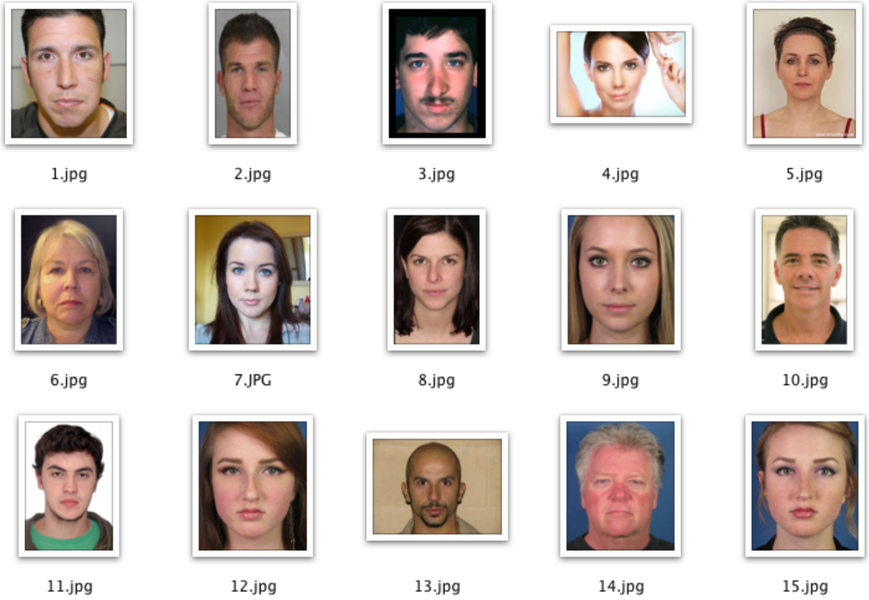


Figure 6: Snapshot of the test set.

All the images in all datasets undergo the pre-process stage that we mentioned before (in section 2), with eyes aligned horizontally, see the sample below.

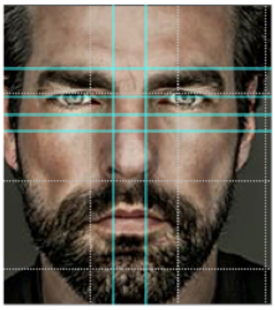


Figure 7: Eyes-aligned horizontally.

4.2 LEARNING

Parameters:

- The training images: The faces we want to learn.
- Labels: The labels corresponding to the images, it has to be given as 0 for male and 1 for female.

- Number of principal components: The performance of the algorithm could vary with the number of principal components, but because of applying Fisherface based on PCA. Thus, we have to maintain our parameter to get best data fitting. Fisherface works based on number of components, which should be less or equal c-1, where c is the number of classes. In our case you have just two classes male and female, this left us with just one principal component.

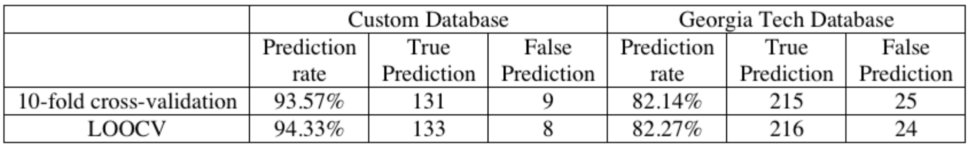
- Threshold: The threshold applied in the prediction. If the distance to the nearest neighbor is larger than the threshold, this method returns -1, which means that the gender is unrecognized, but we didnt specify any threshold value in our algorithm, instead the algorithm predict the gender of detected face with specific confidence.

- Number of iteration: we used four iteration and we got good result, more iteration would improve the algorithm performance, but at the same time it's computationally expensive.

Because the database is consider as one of most important factor that affects any learning algorithm performance, we have been validate the two training databases using two different validate criteria, which helps to determine if overfitting is occurred:

- k-Fold Cross-Validation, where we used k=10: a single subsample is retained as the validation data for testing the model, and the remaining 9 subsamples are used as training data then the process repeated 10 times (the folds), with each of the 10 subsamples used exactly once as the validation data. The 10 results from the folds then can be averaged to produce a single estimation [10].

- Leave-One-Out Cross-Validation ($LOOCV$): which uses a single observation from the original sample as the validation data, and the remaining observations as training data [10].



	Custom Database			Georgia Tech Database		
	Prediction rate	True Prediction	False Prediction	Prediction rate	True Prediction	False Prediction
10-fold cross-validation	93.57%	131	9	82.14%	215	25
LOOCV	94.33%	133	8	82.27%	216	24

Figure 8: DataBase validation.

4.3 RESULTS

The algorithm successfully classify 11 faces out of 15 faces in the test set, which means 73.33% recognize rate, the figure below shows some of the them.

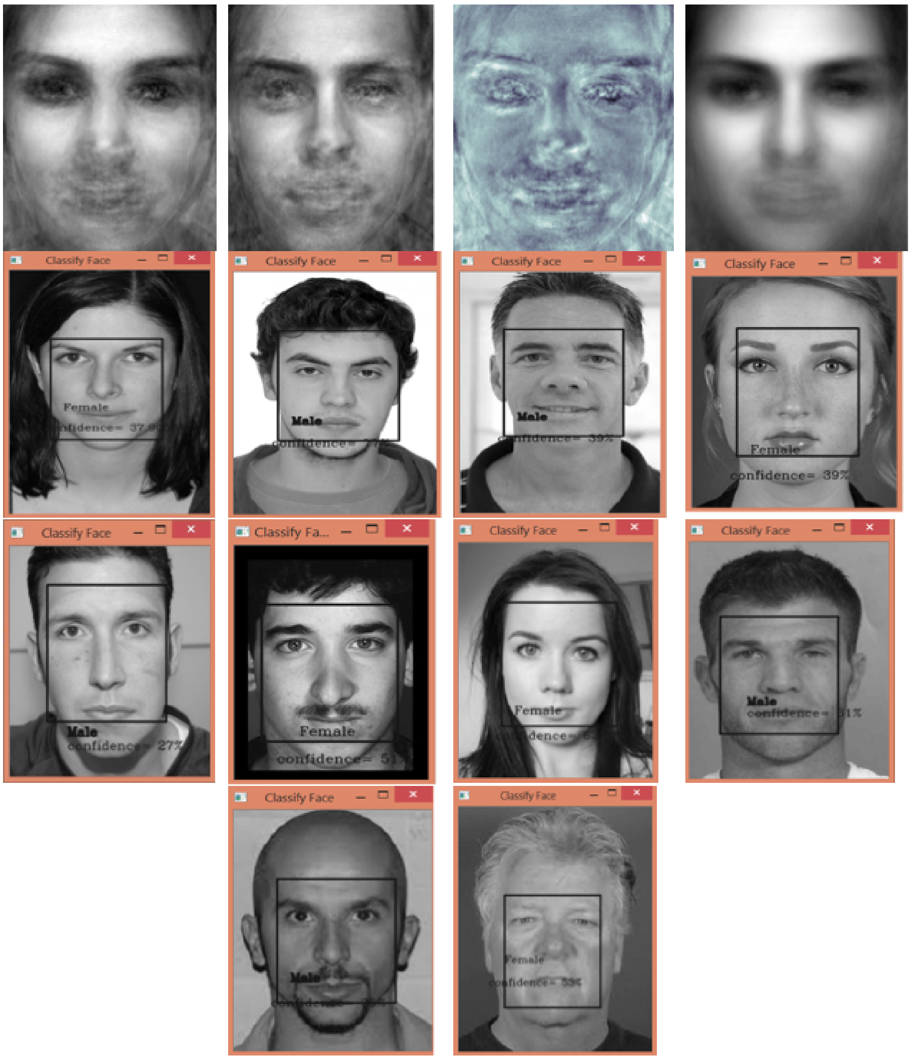


Figure 9: : First four images respectively show the mean image in each class, the mean for all images in Fisherfaces and the mean for all images in Eigenfaces, the rest are some results for classifying the test set.

In order to make this application more immersive and fun especially in real time recognition. We implement role-based reaction using virtual face (see the figure below), to react differently when seeing face of male than when seeing face of female, some examples of the virtual face reaction are shown in the appendix part.

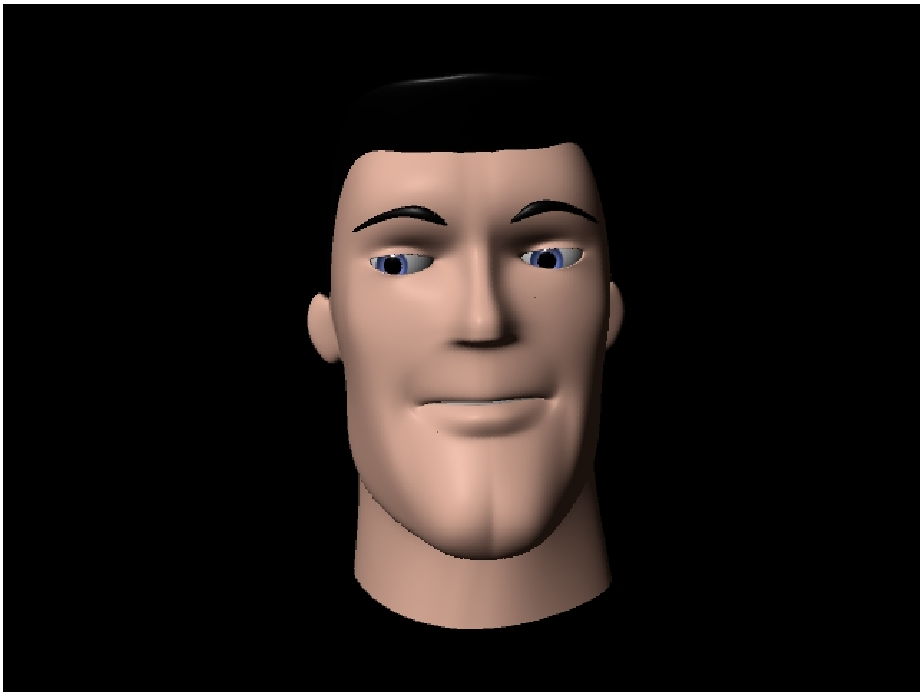


Figure 10: Virtual face used in role-based implemention.

5 CONCLUDING REMARKS AND FUTURE WORK

We noticed that pre-processing stage is highly effected the algorithm performance as initially we faced problem with generalization, where the algorithm misclassify 12 faces out of 15 faces in test set. Crop and align the eyes horizontally in the training database images improved a lot the performance of the gender recognition.

The algorithm limitation:

- The algorithm might misclassify when a person wearing glasses, especially sunglasses.
- The algorithm might misclassify black man and woman.
- The algorithm might misclassify old man and woman.
- Beautiful men with beautiful eyes its really consider as big limitation here, they always classify as female.
- This algorithm is sensitive to the head position and angles variation, so it might misclassify some times when it preforms in real-time (video).

The gender classification work could be enhance to retch more than 96% classification accuracy using SIFT as feature extraction technique which is consider as scale invariant technique, which helps in case of distance detected faces. Followed by SVM as machine learning classifier, where in literature shown that the robustness of SVM around 94.08% [11].

6 REFERENCES

[1] W. Zhao, R. Chellappa, P.Phillips, and A. Rosenfeld, Face Recognition: A Literature Survey, ACM 4th ser.35, 399-458, November 2003.

[2] L.Aryananda, A few days of a robot's life in the human's world: toward incremental individual recognition, Ph.D. Thesis, MIT CSAIL, January 2007.

[3] P. Belhumeur, J. Hespanha and D. Kriegman, Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, IEEE Trans. PAMI, Special Issue on Face Recognition, 19(7), 711-20, July 1997

[4] M.Feld and C.Müller, Speaker Classification for Mobile Devices, Proceedings of the 2nd IEEE International Interdisciplinary Conference on Portable Information Devices (Portable 2008), Germany, August 2008.

[5] P.Viola and M.Jones, Rapid Object Detection using a Boosted Cascade of Simple Features, CVPR (1): 511-518, July 2001.

[6] R. Gonzalez and R.Woods, Digital Image Processing, Third Edition, 2008.

[7] M.Uray, P.Roth and H.Bischof, Efficient Classification for Large-scale Problems by Multiple LDA Subspaces, VISAPP (1) 299-306, September 2009.

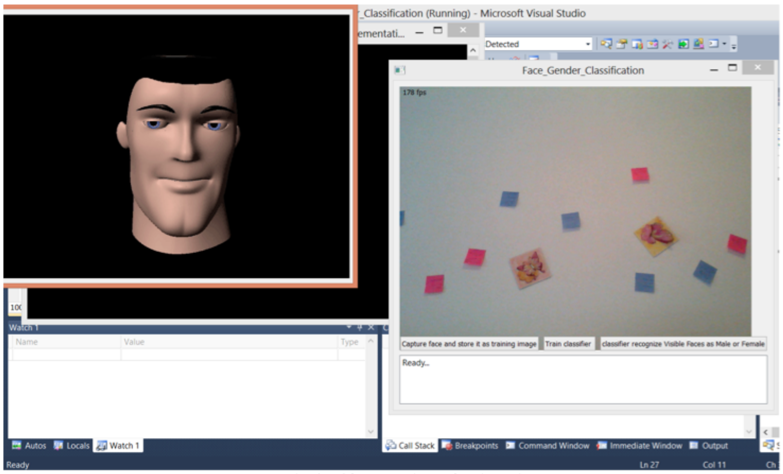
[8] X.Zhixiang, K.Weinberger and O.Chapelle, Distance Metric Learning for Kernel Machines, CoRR abs/1208.3422, 2012.

[9] $www.anefian.com/research/GTdb\_crop.zip$

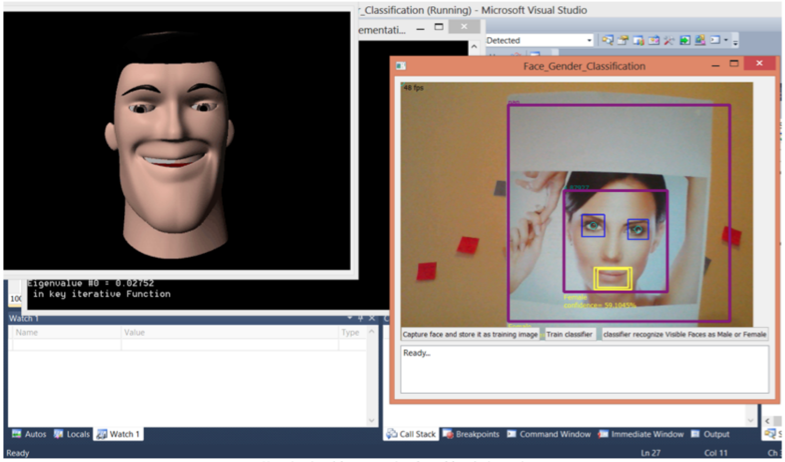
[10] P.Devijver and J.Kittler, Pattern Recognition: A Statistical Approach, Prentice-Hall, London, GB, 1982

[11] P.Nguyen, L.Trung, D.Tran, X.Huang, and D.Sharma, Fuzzy support vector machines for age and gender classification, INTERSPEECH2806-2809, 2010.

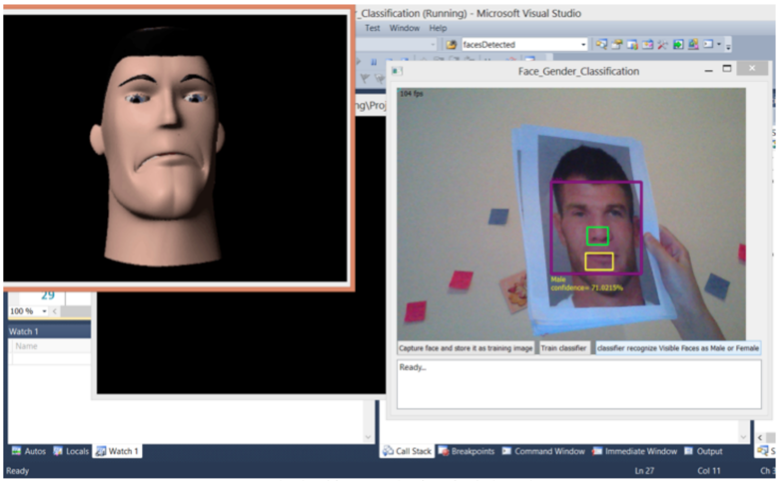
7 APPENDIX


_Classification (Running) - Microsoft Visual Studio
Face_Gender_Classification
178 fps
Capture face and store it as training image
Train classifier
classifier recognize Visible Faces as Male or Female
Ready..
Watch 1
Name
Value
Type
Autos
Locals
Watch 1
Call Stack
Breakpoints
Command Window
Immediate Window
Output
Ready
Ln 27
Col 11

How the virtual face react when no face detected


```
Eigenvalue #0 = 0.02752
 in key iterative Function
```
_Classification (Running) - Microsoft Visual Studio
Face_Gender_Classification
48 fps
Female
confidence= 59.1045%
Capture face and store it as training image
Train classifier
classifier recognize Visible Faces as Male or Female
Ready..
Watch 1
Name
Value
Type
Autos
Locals
Watch 1
Call Stack
Breakpoints
Command Window
Immediate Window
Output
Ready
Ln 27
Col 11

How the virtual face react when face of female detected



How the virtual face react when face of male detected