SMAI-Project
Gender Detection From Facial Features

Guided By:

Bapi Raju Surampudi Sir

Mentored By:

Subba Reddy

Submitted by:

Siddharth Saklecha (201505570)

Athresh G (201505565)

Viplav Sanghvi (201505575)

Abstract

The objective of this project is to identify the gender of a person by looking at his/her photograph. This is a case of supervised learning where the algorithm is first trained on a set of female and male faces, and then used to classify new data.We have not taken genders other than Male and Female into account

Introduction

We tried to identify gender from facial features, we are often curious about what features of the face are most important in determining gender.Are localized features such as eyes, nose and ears more important or overall features such as head shape, hairline and face contour more important?

In this project, the following methods were used for classification:

- Eigenface Method
- K-means
- SVM that performs supervised learning on reduces space of PCA

Data Set and Preprocessing

The data we have is a set of high resolution colour images of 396 female faces and 389 male faces obtained from the MUCT database. All images are frontal views of the face. The database provides diversity of lighting,age and ethnicity.

The images also have variations in :

- Subject's head rotation and tilt
- Subject's facial expression
- Subject's face/Hair accessories

- Position of the face in the image

We run a python script to center all the images in our database by centering the images the faces are aligned at the axis of symmetry of the face.Hence, we have a set of centered and uncentered images.We also use coloured (RGB) and B/W versions of the given images.Colour images have been compressed to 140x140 pixels and B/W to 64x48 pixels.

Implementation

1. EigenFace Method

A popular method in face recognition is the Eigenface algorithm. It takes a holistic approach by processing the entire face and coming up with an eigenface basis.In this method, a Principle Component Analysis (PCA) is performed to reduce the dimensionality.

Table1:Eigenfaces method on Dataset1

Gender	Error
Male	0.154
Female	0.885

Table2:Eigenfaces method on Dataset2

Gender	Error
Male	0.852
Female	0.17

On Dataset 1, the algorithm shows very good recognition for males but a very poor one for females. We conclude here that the algorithm is basically identifying almost every new face to be male, hence contributing to the large error for females.
One disadvantage of PCA is that it cannot give you an intuitive sense of why the algorithm is favouring males. But upon looking at the data where the algorithm misclassified the person, we conclude that female subjects who have short hair, hair tied back or in a scarf were almost always labeled male. Having insufficient examples for them to train on might have resulted in this outcome.

Running the same algorithm on Dataset 2 reduced the excessive bias towards males, as now the female faces were equally well-centered.

In all cases, the number of principal components was chosen to be 200.

Table 3:For Precision,Recall,F1-Score,Support

Class Labels	**Precision**	**Recall**	**F1-Score**
Female	0.76	0.67	0.71
Male	0.70	0.59	0.68
Avg/Total	**0.73**	**0.63**	**0.69**

Confusion Matrix:

Table 4:Confusion Matrix

Class Labels	**Male**	**Female**
Male	50	32
Female	18	51

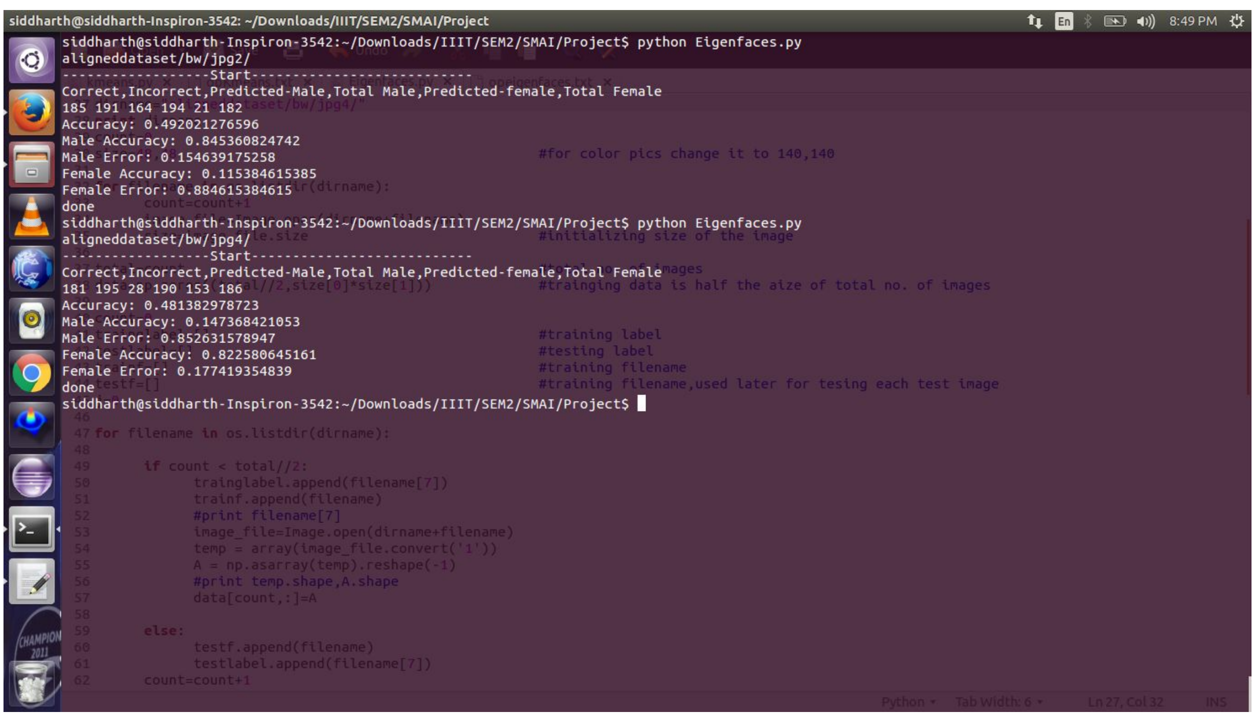

Figure 1 :Output Snapshot of EigenFaces Method

2. **K-Means**

We apply K-means directly on the pixel data that we get from images to obtain 10 clusters for female faces and 10 for male faces. We would like to call these the 10 most representative female and male faces. We then run the K Nearest Neighbours algorithm to classify our test images. K was chosen to be 5.

Table5:K-means method on dataset1

Gender	Error
Male	0.437
Female	0.478

Table6:K-means method on dataset2

Gender	Error
Male	0.289
Female	0.703

Table 7:For Precision,Recall,F1-Score,Support

Class Labels	Precision	Recall	F1-Score
Female	0.63	0.68	0.77
Male	0.61	0.66	0.75
Avg/Total	**0.64**	**0.67**	**0.76**

Confusion Matrix:

Table 8:Confusion Matrix

Class Labels	Male	Female
Male	60	22
Female	25	45

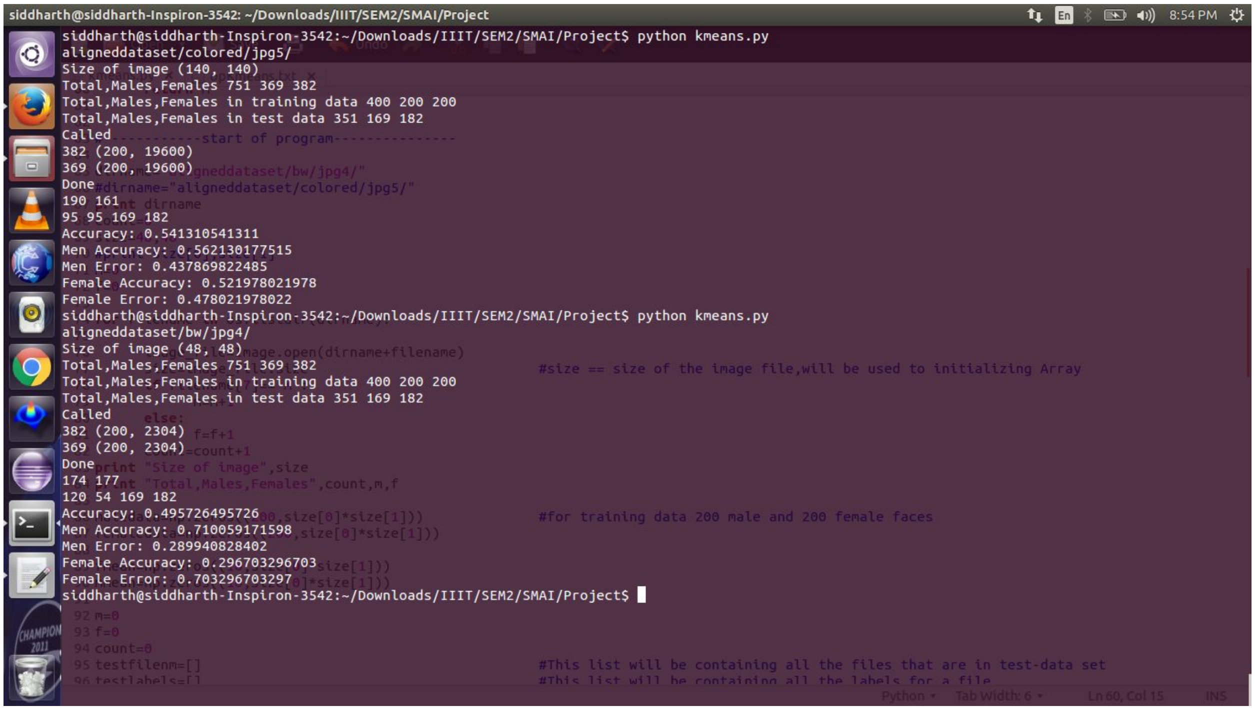

Figure 2 :Output Snapshot of K-means Method

3. PCA with SVM

SVM is yet another way of performing supervised learning over the reduced space. The PCA was applied to reduce dimensionality of the vectors that serve as inputs to the SVM. The SVM then does supervised learning. Sometimes this method is called the fisher discriminant analysis.

The number of principal components was chosen to be 150.

Table3:SVM method on dataset1

Gender	Error
Male	0.540
Female	0.115

Table4:SVM method on dataset2

Gender	Error
Male	0.3315
Female	0.25

Table 5:For Precision,Recall,F1-Score,Support

Class Labels	**Precision**	**Recall**	**F1-Score**	**Support**
Female	0.86	0.76	0.81	82
Male	0.75	0.86	0.80	69
Avg/Total	**0.81**	**0.80**	**0.80**	**151**

Confusion Matrix:

Table 5:Confusion Matrix

Class Labels	**Male**	**Female**
Male	66	16
Female	15	54

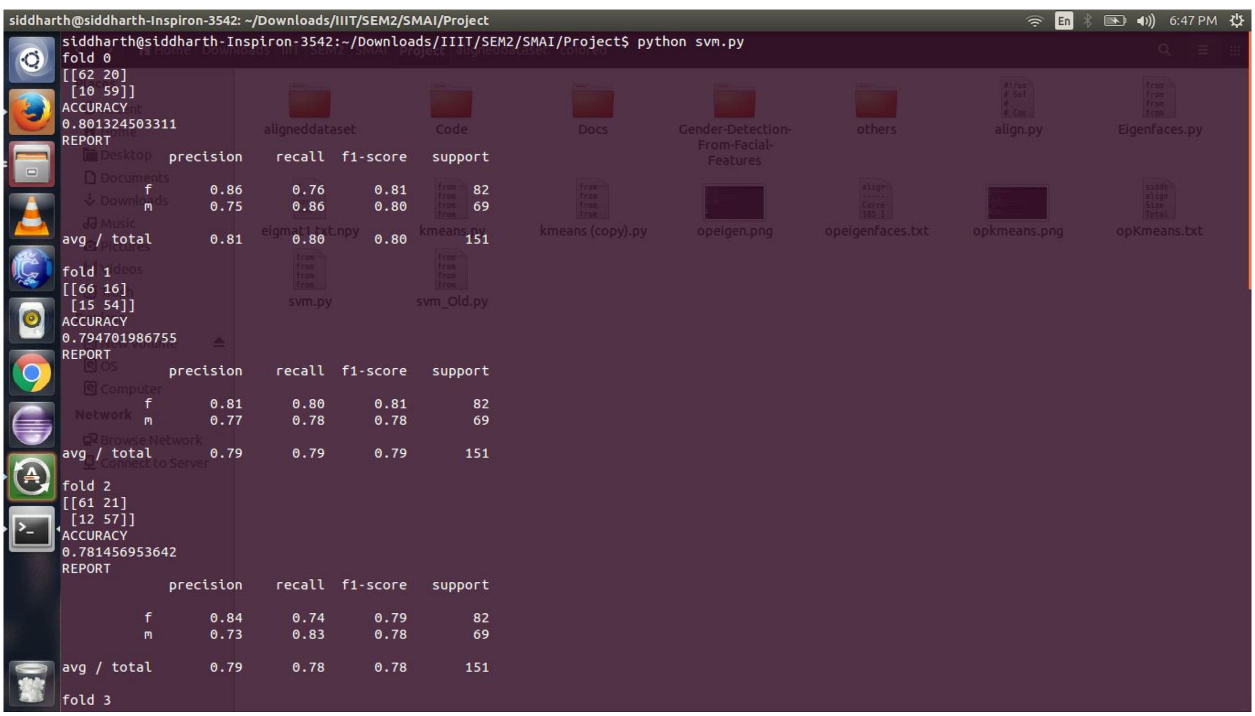

Figure 3 :Output Snapshot of SVM Method