WROCŁAW UNIVERSITY OF SCIENCE AND TECHNOLOGY
FACULTY OF ELECTRONICS

FIELD: Automatics and Robotics (AIR)
SPECIALIZATION: Embedded Robotics (AER)

INTERMEDIATE PROJECT

Gender recognition based on face proportions

AUTHOR:
Piotr Dziębowski

SUPERVISOR:

Witold Paluszyński Ph.D.

Abstract:
The main goal of the project was to develop gender recognition
program. It can be divided on few smaller tasks: Database collection,
database preparation, choosing characteristic facial features,
implementing machine learning, final results.
Project ended with success. Database of 1300 pictures goes through
features extraction algorithm, then machine learning comes in action.
Naive-bayes algorithm obtained positive guesses at level around 60%.

WROCŁAW 1 February 2017

1   Introduction

The main goal of the project was to develop face recognizing program. To obtain this task the big amount of face pictures was needed, so the program could teach itself, by comparing the most significant features. Facebook.com is the good source of face pictures. All faces used in this project was obtained with GraphApi, which is interface to get Facebook users data. Unfortunately in version 2.8 most of features are locked for beginners, for example it is easy to get user profile picture or name, but gender is unavailable. After downloading over 3000 pictures, some processing was needed. For face detection two open-source libraries was used opencv and dlib, but the second occur to be better (it found more faces with less mistakes). Finally opencv was used just for resizing and saving separate faces in gray-scale. Then all faces were manually segregate by gender, here is the point where human factor could influence results. Next inspired by Nastasias Peters drawing tutorial, interesting facial features was chosen. To get all needed features modified face_landmark_detection.py example script from dlib was used. All dlib examples are in the public domain license. All facial features are represented by angles between characteristic vectors. Then each face is represented by vector in CSV file, where last attribute is gender class. Having this representation allowed to run Naive-Bayes algorithm, which learns to recognize gender on learning data(80% of database), then checks results on test data(20% of database).

2   Database collection and preparation

To obtain the set of face pictures GraphApi and guide from article [1] was very useful. Unfortunately only data of people who accepted GraphApi is available this way. Explanation: from 150 facebook groups each around 20000 members this algorithm found 3325 unique profile pictures, when if it would be done before release of version 2.0 in April 30, 2014 the algorithm would acquire around 3 million.
Still a lot of gathered pictures were bad for gender recognition purpose. To obtain only faces default front face classifiers from opencv and dlib libraries was tested. Results of Second one was much better.
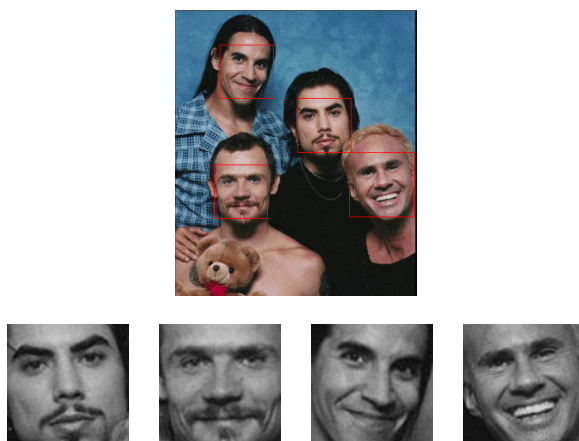

In the picture above, there is an example of multiple faces detection and extraction to separate pictures.

3 Facial features

The most significant values of each picture where chosen accordingly to instructions from drawing tutorial [6]. Each value represent angle between vectors in radians. On the picture below, there are landmarks that can be obtained using dlib and chosen vectors for features.
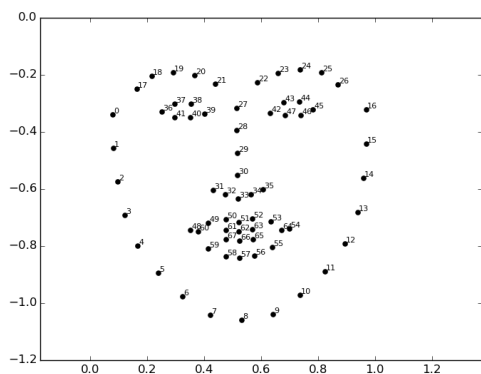
(f) dlib_landmarks
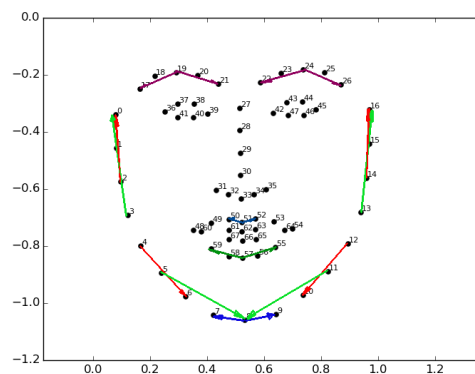
(g) chosen vectors

(h) female

(i) male

After computing the angles, they are write into CSV file with class parameter at the end 1 for male, 0 for female.

4 Machine learning

Finally for gender classification Naive-Bayes was implemented with this task article [7] was very helpful.

The Naive Bayes algorithm is an intuitive method that uses the probabilities of each attribute belonging to each class to make a prediction.

Algorithm works in following steps:

- Handle Data – load the data from CSV file and split it into training and test datasets,

- Summarize Data – summarize the properties in the training dataset for calculating probabilities and making predictions,

- Make a Prediction – use the summaries of the dataset to generate a single prediction,

- Make Predictions – generate predictions given a test dataset and a summarized training dataset,

- Evaluate Accuracy – Evaluate the accuracy of predictions made for a test dataset as the percentage correct out of all predictions made.

5   Summary

The algorithm is around 60% accurate in predicting gender.
The gender recognition is quite complex task and it may be solved in many different ways.
After some research, I can say that 60% is not best score but it is still better than random.
A lot of things influence the final score. In the future this approach can be modified to
check influence of different factors of picture. It can be modified for other class predictions.

6   References

[1] https://sahildua2305.wordpress.com/2014/08/31/this-is-how-i-downloaded-profile-pics-of-my-facebook-friends-using-graph-api/
[2] https://developers.facebook.com/docs/graph-api
[3] https://docs.python.org
[4] http://opencv.org/
OpenCV is released under a BSD license
[5] http://dlib.net/
Boost Software License - Version 1.0 - August 17th, 2003
[6] https://design.tutsplus.com/articles/the-differences-between-male-and-female-portraits–vector-14954
[7] http://machinelearningmastery.com/naive-bayes-classifier-scratch-python/