Image Recognition Capstone Project
Andrey Semenov

I. Image recognition capstone project

Project Overview

The idea of the project has been taken from the description of one of project examples: [Computer Vision Capstone Project](#). The goal of the project was to create a learning

agent in the form of bot for Telegram messenger which is able to recognize faces of famous people.

Problem Statement

The task formulation was: create the bot for Telegram messenger allowing to identify face of famous person.

Python programming language has been chosen for the implementation of user application. Telegram messenger platform has been chosen for user interface development due to the open API. telepot framework was used for the bot development.

Machine learning component of application (database of faces, functions) was taken from sklearn face_recognition.

Metrics

Different models build on training dataset were evaluated by the following criteria classification report:

- F1 score
- Recall score
- Precision score

II. Analysis

Data Exploration

Publically available faces dataset was used. Dataset of labeled faces in wild contains more than 13,000 images of faces collected from the web. Each face has been labeled with the name of the person pictured. 1680 of the people pictured have two or more distinct photos in the data set.

Exploratory Visualization

The dataset is a named series of famous faces people:

Database by name, all

[A)[Alf)[Ang)[B)[Bin)[C)[Che)[Col)[D)[Daw)[Don)[E)[Eri)[F)[G)[Goe)[H) [I)[J)[Jav)[Jes)[Joh)
[Jos)[K)[Kim)[L)[Lil)[M)[Mark)[Mel)[Mik)[N)[O)[P) [Per)[Q)[R)[Ric)[Rog)[S)[Sha)[Ste)[T)
[Tim)[U)[V)[W)[X)[Y)[Z)

AJ Cook (1)   AJ Lamas (1)   Aaron Eckhart (1)   Aaron Guiel (1)   Aaron Patterson (1)

Aaron Peirsol (4)   Aaron Pena (1)   Aaron Sorkin (2)   Aaron Tippin (1)   Abba Eban (1)

Abbas Kiarostami (1)   Abdel Aziz Al-Hakim (1)   Abdel Madi Shabneh (1)   Abdel Nasser Assidi (2)   Abdoulaye Wade (4)
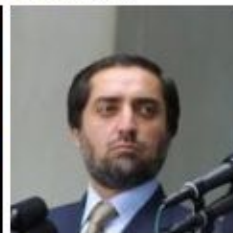
Abdul Majeed Shobokshi (1)   Abdul Rahman (1)   Abdulaziz Kamilov (1)   Abdullah (4)   Abdullah Ahmad Badawi (1)

Many persons have more than 1 image (in brackets).

Algorithms and Techniques

There are three main techniques to be used for the solving this image recognition task:

1) Principal Components Analysis [PCA](#). In order to decrease the complexity of the problems (comparison of faces) and extract only important features with maximum information the PCA filtering was used. Different number of features (5,10,50,100,150,200,250) was compared in order to achieve optimal number of principal components.

2) Machine learning problems in such case is a supervised learning problems. The model is supposed to be trained trained on the part of the dataset (training dataset), tested on testing data set and used on images received from the user. For supervised learning part the Support Vector Classifier was used [SVM](#). In the supervised learning project, the SVM classifier got one of the highest scores.

3) In order to make the learning agent able to learn new images in addition to existing dataset, if the image is not recognized, the image will be stored in the dataset. And in the next request, the classifier will take into account the stored image received from user.

III. Methodology

Data Preprocessing

A few data preprocessing steps was used for the dataset:

- The time of the bot response is crucial. So in order to manage the part of dataset on which the model is trained, the `min_faces_per_person` variable is using in order to take the faces with the >threshold number of pictures. All images from dataset are resizing with `resize=(0,1]` ratio parameter.
- The next step is splitting the data between training and testing data set. For the existing data set the ratio of 3:1 was used, for the bot image recognition all images are using as testing data set and only 1 received from user as testing data set.

- After receiving the image from user, the image is processing:

```
pil_im=Image.open(image)
image_resized=pil_im.resize((w,h))
```

- Also image is converting to 1D numpy array:

```
face = array(image_resized.convert("L"), "f")
face_1D= np.transpose(face.ravel())
```

Implementation

1. *PCA was implemented by the following way.*

```
# Compute a PCA (eigenfaces) on the face dataset (treated as unlabeled
# dataset): unsupervised feature extraction / dimensionality reduction
n_components = 150
print("Extracting the top %d eigenfaces from %d faces"
    % (n_components, X_train.shape[0]))
t0 = time()
pca = PCA(n_components=n_components, svd_solver='randomized',
    whiten=True).fit(X_train)
print("done in %0.3fs" % (time() - t0))
eigenfaces = pca.components_.reshape((n_components, h, w))
print("Projecting the input data on the eigenfaces orthonormal basis")
t0 = time()
X_train_pca = pca.transform(X_train)
X_test_pca = pca.transform(X_test)
print("done in %0.3fs" % (time() - t0))
```

In the beginning the number of principal components was chosen (150 in this example). After that the training and testing datasets were converted into the space of principal components.

2. *SVM classification implementation.*

```
# Train a SVM classification model
print("Fitting the classifier to the training set")
t0 = time()
param_grid = {'C': [1e3, 5e3, 1e4, 5e4, 1e5],
        'gamma': [0.0001, 0.0005, 0.001, 0.005, 0.01, 0.1], }
clf = GridSearchCV(SVC(kernel='rbf', class_weight='balanced'), param_grid)
clf = clf.fit(X_train_pca, y_train)
print("done in %0.3fs" % (time() - t0))
print("Best estimator found by grid search:")
print(clf.best_estimator_)
```

SVC classifier from sklearn package was used with variable parameters 'C' and 'gamma'.

3. Recognizer bot implementation in telegram.

Telepot framework was used and class Bot was created. Bot Recognizer was created by BotFather in Telegram messenger. Received Token was used for authorization. Bot is handling messages by method: on_chat_message. In case of receiving an image, the image is storing and used for recognition.

Refinement

In order to found the best solution using PCA, the following number of components was used: (5,10,50,100,150,200,250). And results log for 12 persons with the highest number of pictures are below. The performance of the model depends upon number of PC could be compared by the average precision/recall/F1 score.

1. n_components=5

```
Total dataset size:
n_samples: 1288
n_features: 1850
n_classes: 7
Extracting the top 5 eigenfaces from 966 faces
done in 0.273s
Projecting the input data on the eigenfaces orthonormal basis
done in 0.021s
Fitting the classifier to the training set
done in 210.672s
Best estimator found by grid search:
SVC(C=50000.0, cache_size=200, class_weight='balanced', coef0=0.0,
  decision_function_shape=None, degree=3, gamma=0.1, kernel='rbf',
  max_iter=-1, probability=False, random_state=None, shrinking=True,
  tol=0.001, verbose=False)
Predicting people's names on the test set
done in 0.010s
             precision   recall  f1-score   support

   Ariel Sharon     0.04     0.08     0.05      13
   Colin Powell     0.37     0.42     0.39      60
Donald Rumsfeld     0.16     0.30     0.21      27
```

```
     George W Bush       0.62     0.42     0.50        146
 Gerhard Schroeder       0.10     0.12     0.11         25
      Hugo Chavez        0.11     0.13     0.12         15
       Tony Blair        0.25     0.22     0.24         36

      avg / total        0.41     0.34     0.36        322
```

2. n_components=10

```
Total dataset size:
n_samples: 1288
n_features: 1850
n_classes: 7
Extracting the top 10 eigenfaces from 966 faces
done in 0.249s
Projecting the input data on the eigenfaces orthonormal basis
done in 0.020s
Fitting the classifier to the training set
done in 76.697s
Best estimator found by grid search:
SVC(C=1000.0, cache_size=200, class_weight='balanced', coef0=0.0,
  decision_function_shape=None, degree=3, gamma=0.1, kernel='rbf',
  max_iter=-1, probability=False, random_state=None, shrinking=True,
  tol=0.001, verbose=False)
Predicting people's names on the test set
done in 0.016s
                 precision   recall  f1-score   support

    Ariel Sharon       0.09     0.15     0.11         13
    Colin Powell       0.46     0.53     0.49         60
 Donald Rumsfeld       0.24     0.30     0.27         27
   George W Bush       0.68     0.60     0.64        146
Gerhard Schroeder      0.19     0.20     0.20         25
     Hugo Chavez       0.14     0.13     0.14         15
      Tony Blair       0.59     0.47     0.52         36

      avg / total      0.50     0.48     0.49        322
```

3. n_components=50

```
Total dataset size:
n_samples: 1288
n_features: 1850
n_classes: 7
Extracting the top 50 eigenfaces from 966 faces
done in 0.370s
```

Projecting the input data on the eigenfaces orthonormal basis
done in 0.031s
Fitting the classifier to the training set
done in 11.607s
Best estimator found by grid search:
SVC(C=1000.0, cache_size=200, class_weight='balanced', coef0=0.0,
  decision_function_shape=None, degree=3, gamma=0.01, kernel='rbf',
  max_iter=-1, probability=False, random_state=None, shrinking=True,
  tol=0.001, verbose=False)
Predicting people's names on the test set
done in 0.028s

	precision	recall	f1-score	support
Ariel Sharon	0.67	0.77	0.71	13
Colin Powell	0.82	0.90	0.86	60
Donald Rumsfeld	0.75	0.56	0.64	27
George W Bush	0.85	0.92	0.88	146
Gerhard Schroeder	0.75	0.72	0.73	25
Hugo Chavez	0.77	0.67	0.71	15
Tony Blair	0.89	0.67	0.76	36
avg / total	0.82	0.82	0.82	322

4. n_components=100

Total dataset size:
n_samples: 1288
n_features: 1850
n_classes: 7
Extracting the top 100 eigenfaces from 966 faces
done in 0.333s
Projecting the input data on the eigenfaces orthonormal basis
done in 0.034s
Fitting the classifier to the training set
done in 20.996s
Best estimator found by grid search:
SVC(C=1000.0, cache_size=200, class_weight='balanced', coef0=0.0,
  decision_function_shape=None, degree=3, gamma=0.005, kernel='rbf',
  max_iter=-1, probability=False, random_state=None, shrinking=True,
  tol=0.001, verbose=False)
Predicting people's names on the test set
done in 0.044s

	precision	recall	f1-score	support
Ariel Sharon	0.77	0.77	0.77	13
Colin Powell	0.86	0.92	0.89	60

Donald Rumsfeld	0.76	0.59	0.67	27
George W Bush	0.88	0.97	0.92	146
Gerhard Schroeder	0.83	0.76	0.79	25
Hugo Chavez	1.00	0.60	0.75	15
Tony Blair	0.90	0.75	0.82	36
avg / total	0.86	0.86	0.86	322

Example of visual analysis below: prediction vs true name. And display of first 11 eighteen vectors:

5. n_components=150

Total dataset size:
n_samples: 1288
n_features: 1850
n_classes: 7
Extracting the top 150 eigenfaces from 966 faces
done in 0.453s
Projecting the input data on the eigenfaces orthonormal basis
done in 0.049s
Fitting the classifier to the training set
done in 32.176s
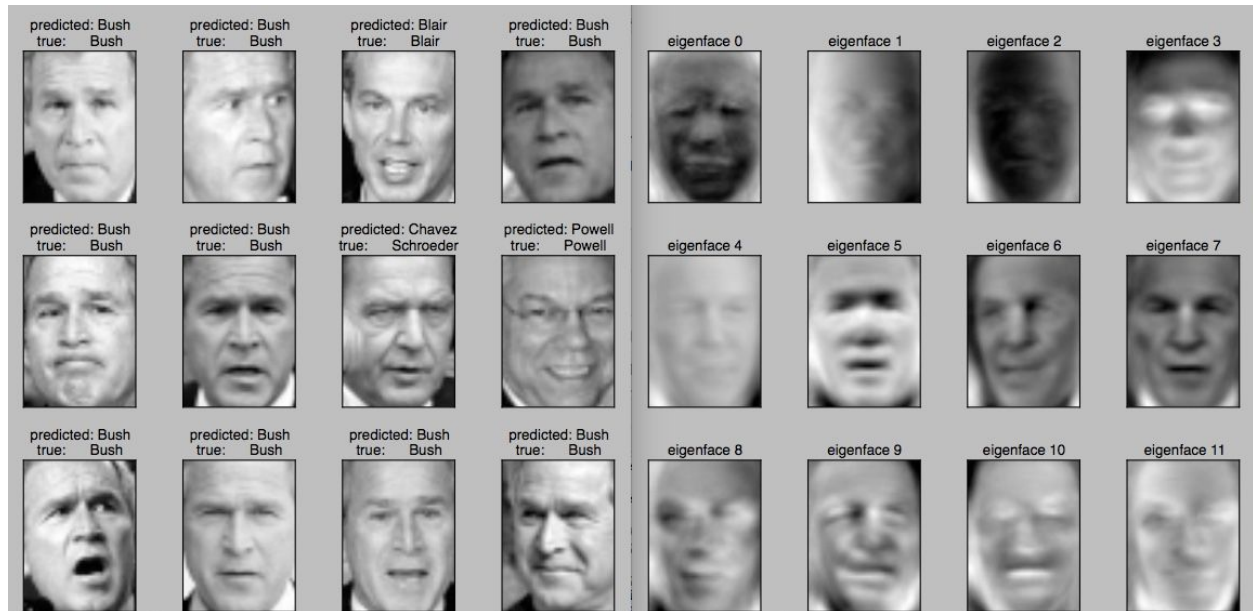Best estimator found by grid search:
SVC(C=1000.0, cache_size=200, class_weight='balanced', coef0=0.0,
  decision_function_shape=None, degree=3, gamma=0.005, kernel='rbf',
  max_iter=-1, probability=False, random_state=None, shrinking=True,
  tol=0.001, verbose=False)
Predicting people's names on the test set

done in 0.060s

```
               precision    recall  f1-score   support

   Ariel Sharon       0.73      0.62      0.67        13
   Colin Powell       0.81      0.87      0.84        60
Donald Rumsfeld       0.85      0.63      0.72        27
  George W Bush       0.84      0.97      0.90       146
Gerhard Schroeder     0.90      0.76      0.83        25
    Hugo Chavez       1.00      0.53      0.70        15
     Tony Blair       0.97      0.83      0.90        36

    avg / total       0.86      0.85      0.85       322
```

6. n_components=200

Total dataset size:
n_samples: 1288
n_features: 1850
n_classes: 7
Extracting the top 200 eigenfaces from 966 faces
done in 0.663s
Projecting the input data on the eigenfaces orthonormal basis
done in 0.061s
Fitting the classifier to the training set

done in 41.205s
Best estimator found by grid search:
SVC(C=1000.0, cache_size=200, class_weight='balanced', coef0=0.0,
  decision_function_shape=None, degree=3, gamma=0.001, kernel='rbf',
  max_iter=-1, probability=False, random_state=None, shrinking=True,
  tol=0.001, verbose=False)
Predicting people's names on the test set
done in 0.071s

```
               precision    recall  f1-score   support

   Ariel Sharon       0.77      0.77      0.77        13
   Colin Powell       0.79      0.88      0.83        60
Donald Rumsfeld       0.76      0.70      0.73        27
  George W Bush       0.90      0.91      0.90       146
Gerhard Schroeder     0.84      0.84      0.84        25
    Hugo Chavez       0.80      0.53      0.64        15
     Tony Blair       0.88      0.83      0.86        36

    avg / total       0.85      0.85      0.85       322
```

As it could be seen from the beginning, in case of low number of PC (5,10) the dimension is not enough to recognize faces and scores are poor. But starting from 50, the average F1/Precision/Recall scores are increasing, which means that model is improving the predictive ability. The local maximum is found for n=100 components (all average scores are equal 0,86) and later increase of number of components to 150 and 200 only a little bit decrease average scores. It is mean, that 100 components are allowing to extract all information from images needed for this task.

IV. Results

Model Evaluation and Validation

The final model was constructed for famous people with >40 images in the database and 100 principal components and have the following parameters:

```
Total dataset size:
n_samples: 1867
n_features: 1850
n_classes: 19
Extracting the top 100 eigenfaces from 1400 faces
done in 0.810s
Projecting the input data on the eigenfaces orthonormal basis
done in 0.057s
Fitting the classifier to the training set
2016-10-30 00:14:32,994 Starting new HTTPS connection (1): api.telegram.org
2016-10-30 00:14:55,005 Starting new HTTPS connection (1): api.telegram.org
done in 54.465s
Best estimator found by grid search:
SVC(C=1000.0, cache_size=200, class_weight='balanced', coef0=0.0,
  decision_function_shape=None, degree=3, gamma=0.005, kernel='rbf',
  max_iter=-1, probability=False, random_state=None, shrinking=True,
  tol=0.001, verbose=False)
Predicting people's names on the test set
done in 0.114s
```

	precision	recall	f1-score	support
Ariel Sharon	0.71	0.62	0.67	16
Arnold Schwarzenegger	1.00	0.33	0.50	15
Colin Powell	0.73	0.87	0.79	61
Donald Rumsfeld	0.64	0.68	0.66	34
George W Bush	0.77	0.93	0.84	135
Gerhard Schroeder	0.76	0.70	0.73	23
Gloria Macapagal Arroyo	1.00	0.82	0.90	11

	precision	recall	f1-score	support
Hugo Chavez	0.76	0.72	0.74	18
Jacques Chirac	0.80	0.36	0.50	11
Jean Chretien	1.00	0.71	0.83	14
Jennifer Capriati	0.86	0.67	0.75	9
John Ashcroft	0.67	0.62	0.64	13
Junichiro Koizumi	1.00	0.94	0.97	16
Laura Bush	1.00	0.82	0.90	11
Lleyton Hewitt	0.88	0.70	0.78	10
Luiz Inacio Lula da Silva	1.00	0.80	0.89	15
Serena Williams	0.80	0.62	0.70	13
Tony Blair	0.71	0.71	0.71	34
Vladimir Putin	0.62	0.62	0.62	8
avg / total	0.79	0.78	0.77	467

Justification

Based on the avg/total scores for precision, recall, F1>0.77 the model is able to recognize the face for 19 famous people. As it is shown below, the faces predicted are correct for the majority of predicitons.
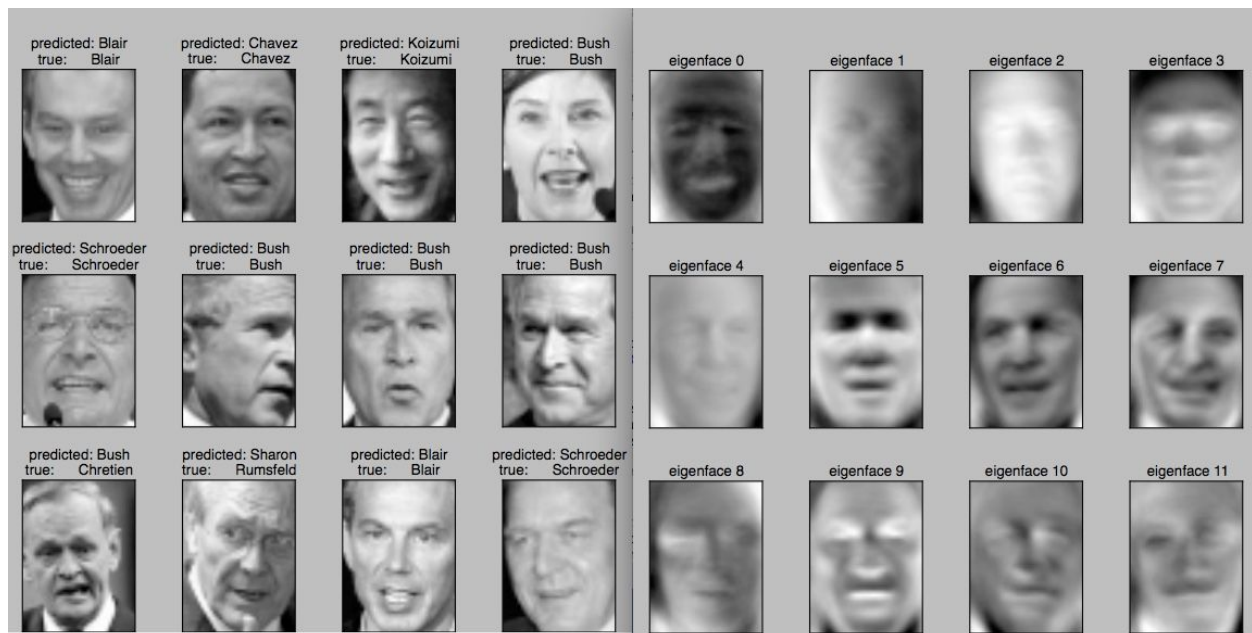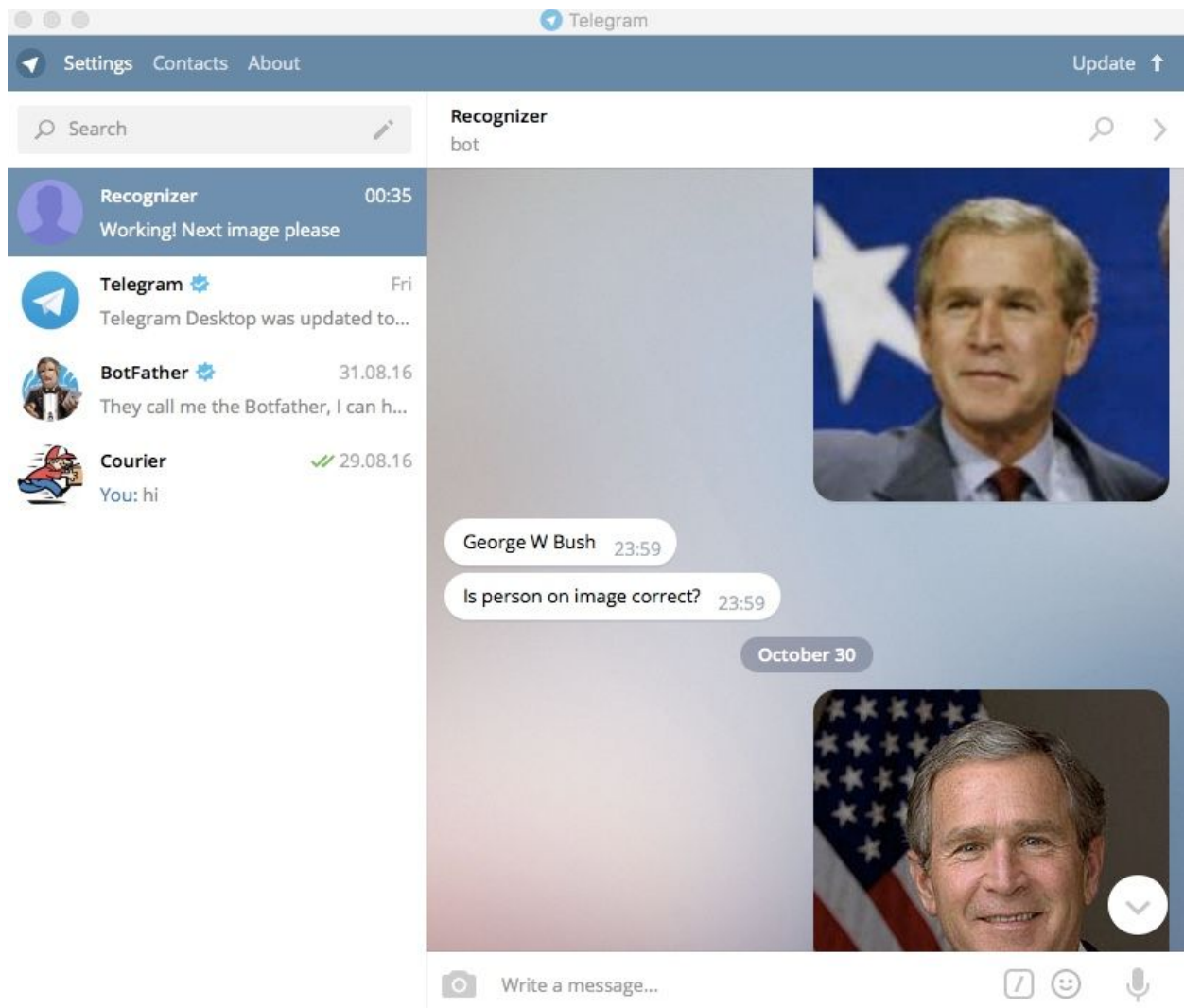

V. Conclusion

Interesting and complex problem of face recognition was used in the scope of the project and implemented in the form of Telegram bot.

Free-Form Visualization

The user interface of application is looks like this:

The image could be send to bot and bot will advice to which famous person the face is similar.

Reflection

The were two particularly interesting and difficult at the same time parts of the project:

1. Image recognition techniques. The whole set of instruments for image recognition: filtering, image representation in the form of arrays, PCA transform, etc... were quite new for me. Great help for me was coverage of PCA on this example in the education materials.

2. Bot user interface. Thanks to open API it was possible to build user interface for the application available on different platforms: mobile, desctop, etc... Interesting part was in building the bot and handling the meassages.

Improvement

Derived model is robust enough in case of the big number (>50 images) per person. In case of decreasing the threshold of images per person to 10 for example, the model has the following performance:

```
Total dataset size:
n_samples: 4324
n_features: 1850
n_classes: 158
Extracting the top 100 eigenfaces from 3243 faces
done in 1.293s
Projecting the input data on the eigenfaces orthonormal basis
done in 0.136s
Fitting the classifier to the training set
done in 620.780s
Best estimator found by grid search:
SVC(C=1000.0, cache_size=200, class_weight='balanced', coef0=0.0,
  decision_function_shape=None, degree=3, gamma=0.005, kernel='rbf',
  max_iter=-1, probability=False, random_state=None, shrinking=True,
  tol=0.001, verbose=False)
Predicting people's names on the test set
done in 2.009s
precision   recall  f1-score   support
```

	precision	recall	f1-score	support
Abdullah Gul	1.00	0.60	0.75	5
Adrien Brody	0.00	0.00	0.00	1
Alejandro Toledo	0.22	0.50	0.31	4
Alvaro Uribe	0.55	0.60	0.57	10
Amelie Mauresmo	0.50	0.33	0.40	6
Andre Agassi	0.06	0.33	0.11	3
Andy Roddick	0.00	0.00	0.00	5
Angelina Jolie	0.00	0.00	0.00	2
Ann Veneman	0.00	0.00	0.00	1
Anna Kournikova	0.00	0.00	0.00	2
Ari Fleischer	0.00	0.00	0.00	4
Ariel Sharon	0.36	0.71	0.48	17
Arnold Schwarzenegger	0.35	0.35	0.35	17
Atal Bihari Vajpayee	0.33	0.50	0.40	2
Bill Clinton	0.50	0.50	0.50	8
Bill Gates	1.00	0.60	0.75	5
Bill McBride	0.50	0.33	0.40	3

Bill Simon	0.25	1.00	0.40	1
Britney Spears	1.00	0.25	0.40	4
Carlos Menem	0.00	0.00	0.00	4
Carlos Moya	0.25	0.33	0.29	3
Catherine Zeta-Jones	1.00	1.00	1.00	1
Charles Moose	0.00	0.00	0.00	4
Colin Powell	0.44	0.89	0.58	53
Condoleezza Rice	0.00	0.00	0.00	3
David Beckham	1.00	0.40	0.57	5
David Nalbandian	0.00	0.00	0.00	2
Dick Cheney	1.00	0.40	0.57	5
Dominique de Villepin	0.50	0.67	0.57	3
Donald Rumsfeld	0.51	0.78	0.62	27
Edmund Stoiber	0.00	0.00	0.00	6
Eduardo Duhalde	0.00	0.00	0.00	1
Fidel Castro	1.00	0.29	0.44	7
George HW Bush	0.00	0.00	0.00	2
George Robertson	0.75	0.50	0.60	6
George W Bush	0.48	0.93	0.63	140
Gerhard Schroeder	0.45	0.76	0.57	25
Gloria Macapagal Arroyo	0.71	0.62	0.67	8
Gonzalo Sanchez de Lozada	0.00	0.00	0.00	6
Gordon Brown	1.00	0.25	0.40	4
Gray Davis	0.40	0.33	0.36	6
Guillermo Coria	0.71	0.62	0.67	8
Halle Berry	1.00	0.33	0.50	3
Hamid Karzai	0.50	0.14	0.22	7
Hans Blix	0.83	0.62	0.71	8
Harrison Ford	1.00	0.50	0.67	2
Hillary Clinton	0.50	0.33	0.40	3
Howard Dean	0.00	0.00	0.00	4
Hu Jintao	1.00	1.00	1.00	4
Hugo Chavez	0.71	0.71	0.71	21
Ian Thorpe	1.00	0.20	0.33	5
Igor Ivanov	0.00	0.00	0.00	5
Jack Straw	0.00	0.00	0.00	6
Jackie Chan	0.00	0.00	0.00	2
Jacques Chirac	0.38	0.55	0.44	11
Jacques Rogge	0.00	0.00	0.00	3
James Blake	1.00	0.17	0.29	6
James Kelly	0.00	0.00	0.00	2
Jason Kidd	0.00	0.00	0.00	4
Javier Solana	0.00	0.00	0.00	2
Jean Charest	1.00	0.12	0.22	8
Jean Chretien	0.79	0.69	0.73	16
Jean-David Levitte	0.00	0.00	0.00	2
Jeb Bush	0.00	0.00	0.00	2
Jennifer Aniston	1.00	0.29	0.44	7
Jennifer Capriati	0.38	0.83	0.53	6
Jennifer Garner	1.00	0.40	0.57	5
Jennifer Lopez	0.75	0.38	0.50	8
Jeremy Greenstock	0.80	0.50	0.62	8
Jiang Zemin	1.00	0.75	0.86	4
Jiri Novak	0.00	0.00	0.00	1
Joe Lieberman	0.00	0.00	0.00	6
John Allen Muhammad	1.00	0.75	0.86	4
John Ashcroft	0.56	0.90	0.69	10
John Bolton	0.67	0.67	0.67	3
John Howard	0.00	0.00	0.00	11
John Kerry	1.00	0.17	0.29	6
John Negroponte	0.25	0.29	0.27	7
John Paul II	0.00	0.00	0.00	5
John Snow	0.50	0.25	0.33	4
Joschka Fischer	1.00	0.60	0.75	5
Jose Maria Aznar	1.00	0.60	0.75	5
Juan Carlos Ferrero	0.50	0.67	0.57	6

Name				
Julianne Moore	0.75	0.50	0.60	6
Julie Gerberding	1.00	0.67	0.80	3
Junichiro Koizumi	0.75	0.88	0.81	17
Keanu Reeves	0.00	0.00	0.00	4
Kim Clijsters	0.00	0.00	0.00	1
Kim Ryong-sung	0.75	0.75	0.75	4
Kofi Annan	0.57	0.67	0.62	6
Lance Armstrong	0.00	0.00	0.00	2
Laura Bush	0.86	1.00	0.92	6
Lindsay Davenport	0.50	0.17	0.25	6
Lleyton Hewitt	0.67	0.55	0.60	11
Lucio Gutierrez	1.00	0.25	0.40	4
Luiz Inacio Lula da Silva	0.82	0.75	0.78	12
Mahathir Mohamad	0.00	0.00	0.00	4
Mahmoud Abbas	1.00	0.62	0.77	8
Mark Philippoussis	0.00	0.00	0.00	2
Megawati Sukarnoputri	0.88	0.78	0.82	9
Meryl Streep	0.50	0.25	0.33	4
Michael Bloomberg	0.00	0.00	0.00	4
Michael Jackson	1.00	0.33	0.50	3
Michael Schumacher	0.50	0.67	0.57	3
Mike Weir	0.00	0.00	0.00	4
Mohammad Khatami	1.00	0.50	0.67	2
Mohammed Al-Douri	0.00	0.00	0.00	5
Muhammad Ali	1.00	0.33	0.50	3
Nancy Pelosi	0.00	0.00	0.00	5
Naomi Watts	0.25	0.67	0.36	3
Nestor Kirchner	0.33	0.20	0.25	10
Nicanor Duarte Frutos	0.00	0.00	0.00	3
Nicole Kidman	0.00	0.00	0.00	6
Norah Jones	0.00	0.00	0.00	7
Paradorn Srichaphan	0.00	0.00	0.00	2
Paul Bremer	1.00	0.29	0.44	7
Paul Burrell	1.00	0.50	0.67	2
Paul Wolfowitz	0.00	0.00	0.00	1
Pervez Musharraf	1.00	0.25	0.40	4
Pete Sampras	1.00	0.50	0.67	4
Pierce Brosnan	0.00	0.00	0.00	1
Queen Elizabeth II	0.00	0.00	0.00	2
Recep Tayyip Erdogan	0.75	0.86	0.80	7
Renee Zellweger	1.00	0.75	0.86	4
Ricardo Lagos	0.43	0.50	0.46	6
Richard Gephardt	1.00	0.25	0.40	4
Richard Gere	0.00	0.00	0.00	3
Richard Myers	0.00	0.00	0.00	9
Roger Federer	1.00	0.14	0.25	7
Roh Moo-hyun	0.67	0.55	0.60	11
Rubens Barrichello	0.33	0.33	0.33	3
Rudolph Giuliani	1.00	0.25	0.40	4
Saddam Hussein	0.67	0.25	0.36	8
Salma Hayek	0.00	0.00	0.00	4
Serena Williams	0.53	0.60	0.56	15
Sergey Lavrov	0.50	1.00	0.67	1
Sergio Vieira De Mello	0.00	0.00	0.00	5
Silvio Berlusconi	0.11	0.17	0.13	6
Spencer Abraham	1.00	0.40	0.57	5
Taha Yassin Ramadan	0.00	0.00	0.00	3
Tang Jiaxuan	1.00	1.00	1.00	3
Tiger Woods	0.50	0.38	0.43	8
Tim Henman	0.50	0.25	0.33	4
Tom Cruise	1.00	0.67	0.80	3
Tom Daschle	0.60	0.60	0.60	5
Tom Hanks	0.00	0.00	0.00	1
Tom Ridge	0.62	0.62	0.62	8
Tommy Franks	0.00	0.00	0.00	8
Tommy Thompson	0.00	0.00	0.00	3

```
         Tony Blair     0.47    0.68    0.56      37
         Trent Lott     1.00    0.25    0.40       4
     Venus Williams     0.00    0.00    0.00       5
       Vicente Fox      0.75    0.60    0.67      10
     Vladimir Putin     0.33    0.67    0.44       9
    Walter Mondale      0.00    0.00    0.00       2
        Wen Jiabao      1.00    0.60    0.75       5
      Winona Ryder      0.25    0.50    0.33       2
   Yoriko Kawaguchi     1.00    0.33    0.50       3

        avg / total     0.51    0.51    0.47    1081
```

So there are 2 clear issues:

- Decrease of scores in case of using more images in the dataset and less images per person for model training. But still the derived model have scores >0.5.
- The training of the model is taking >10 minutes which makes it hard to reeducate the model after adding the images to the dataset.

The solution of those problems could significantly improve the project.