AI documentation:

1 – Problem Definition:

Main idea is to recognize a human face from 400 different grayscale digital images to 40 distinct people Everyone has 10 images in a data set called Olivetti data set.
Differentiate between those images is made by labeling the faces Found in the image then split those faces “labels” -if there are more than one person in the image but if one person only found in the Image then we have one label “face” - and by applying the Random Forest classifier algorithm which tests and classify labels then predict Which one is relevant to definite person from those 40 persons based on the person's facial textures and shape which are previously existed? In the data set.

2-Algorithm used: Random Forests.

Our project run on the following steps sequentially: -

1- Get the data from Olivetti data set.
2- label the faces from the images in sub-windows.
3- Split the labels.
4- Train and test.
5- Applying Random forest classifier algorithm.
6- input the data to predict
7- Start prediction which label is relevant to who from the data set.
8- Output & results.

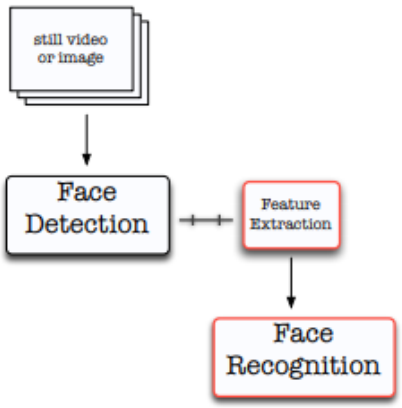


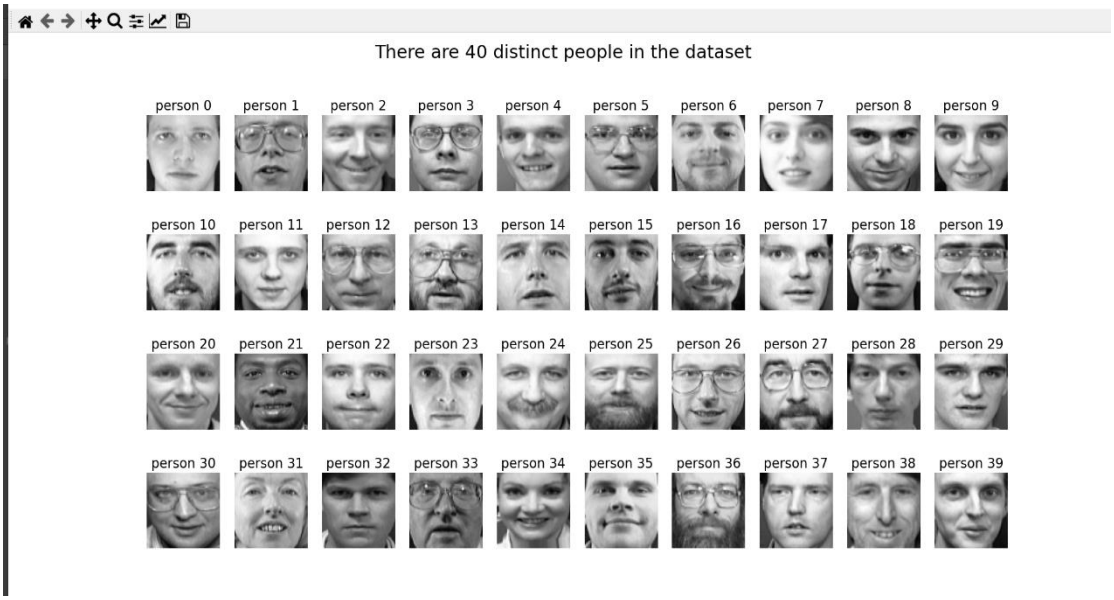
fig-The three main stages of a face processing system

3-Input Explanation:

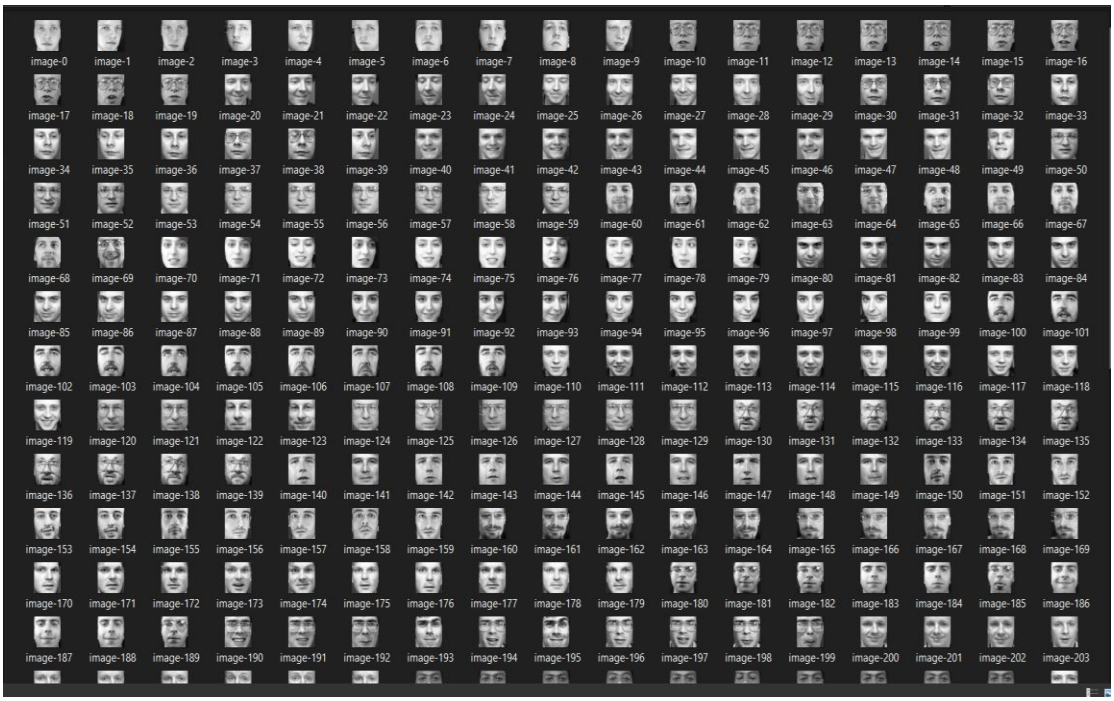
1-the Dataset employed (preferably a publicly available dataset):

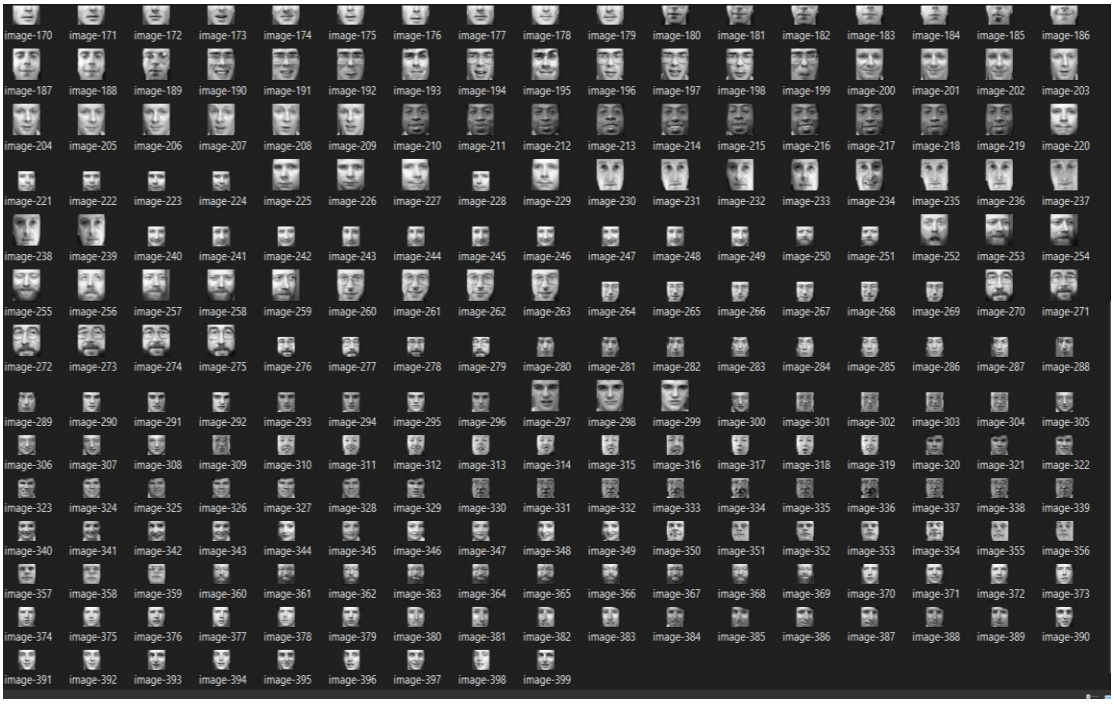
face recognition was performed using the face images in the Olivetti data set

There are ten different image of each of 40 distinct people



There are 400 face images in the dataset





Face images were taken at different times, varying lighting, facial expression and facial detail All face images have black background

1-The images are gray level.

2-Size of each image is 64x64.

3-Image pixel values were scaled to [0, 1] interval.

4-Names of 40 people were encoded to an integer from 0 to 39.

And you will select an image and I will view a name of the selected picture

---

Main functionalities

face recognition was performed using the face images in the Olivetti data set. The steps for face recognition are as follows:

1.Principal components of face images were obtained by Olivetti Dataset.

```
st.py > ...
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.image as mpimg
import os

print(os.listdir("D:/ai_project/Face-Recognition-by-Random-forest/olivetti_faces.npy"))
pics=np.load("D:/ai_project/Face-Recognition-by-Random-forest/olivetti_faces.npy/olivetti_faces.npy")
labels= np.load("D:/ai_project/Face-Recognition-by-Random-forest/olivetti_faces.npy/olivetti_faces_target.npy")
print("pics: ", pics.shape)
print("labels: ", labels.shape)
```

2. Show 40 Face Images of Selected Target

```
fig = plt.figure(figsize=(20, 10))
columns = 10
rows = 4
for i in range(1,columns*rows+1):
    img = pics[10*(i-1),:,:]
    fig.add_subplot(rows, columns, i)
    plt.imshow(img, cmap = plt.get_cmap('gray'))
    plt.title("person {}".format(i-1), fontsize=16)
    plt.axis('off')

plt.suptitle("There are 40 distinct people in the dataset", fontsize=22)
```

3. Split data and target into Random train and test Subsets

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(Xdata, Ydata, test_size = 0.241, random_state=45)

print("x_train: ",x_train.shape)
print("x_test: ",x_test.shape)
print("y_train: ",y_train.shape)
print("y_test: ",y_test.shape)

x_train = x_train.reshape(x_train.shape[0], x_train.shape[1]*x_train.shape[2])
x_test = x_test.reshape(x_test.shape[0], x_test.shape[1]*x_test.shape[2])

print("x_train: ",x_train.shape)
print("x_test: ",x_test.shape)
print("y_train: ",y_train.shape)
print("y_test: ",y_test.shape)
```

4.According to random forest classifier taking a (train-x and train-Y) data score were obtained, Algorithm take an image and compare data and view a name using label

```
from sklearn.ensemble import RandomForestClassifier

rf = RandomForestClassifier(n_estimators=400,random_state=1)
rf.fit(x_train, y_train)
RF_accuracy = round(rf.score(x_test, y_test)*100,2)
print("RF_accuracy is %", RF_accuracy)


image=mpimg.imread("D:/ai_project/Face-Recognition-by-Random-forest/images/image-0.png")
image=image.reshape(1,-1)
print(image.shape)
y_pred= rf.predict(image)
print("Person",y_pred)
fig = plt.figure(figsize=(4,5))
img = pics[10*(y_pred[0]),:,:]
plt.imshow(img, cmap = plt.get_cmap('gray'))
plt.axis('off')
plt.title("person {}".format(y_pred), fontsize=16)
plt.show()
```

4-Output Explanation:

5- Details of the algorithm & The Results of the experiments:

Algorithm applied is Random Forest algorithm. Recognition with random forests was compared to SVM and log-linear models.
Each learning method includes some methods of randomization.
Such aim is to choose the best option at every step. Very famous is random forest. Random forest learns built randomized decision tree. For each iteration the algorithm often produces excellent predictors .
Random Forest basic idea is to find the average value of noise. Very complex interaction trees can capture. Complex input space can be computed into simpler space and it's the aim of decision tree. Random forests are collection of decision trees. In has shown that the collection of random forest, decision trees trained randomly. Therefore available data reduces the over fitting in comparison. Therefore, RF methodology is extension of bagging classification tree.
It is a parallel learning process. It achieves a high accuracy and has fast training phase.
The advantages of random forest are :
-Local representation
-Classification with occlusion
-Parallelization
-Fast Training Time
For the Face detection best and worst performance cases :-

- True Positive Rate = true positives/(true positives + false negatives) = number of faces detected/ total number of faces in the input image
- False positive rate
  = false positives/(false positives + true negatives)

= (number of sub-windows incorrectly labeled as faces) / (all sampled sub-windows in the input image)

For the results & output of the experiments:

Here is a snapshot after running the code to recognize that the face In the input image is relevant to whom in the data set . It found that It was belonging to the first Person in the data set “Person[0]” .

```
In [2]: image=mpimg.imread("E:/3rd Year/AI/project/Face-Recognition-
by-Random-forest/images/image-8.png")
   ...: image=image.reshape(1,-1)
   ...: print(image.shape)
   ...: y_pred= rf.predict(image)
   ...: print("Person",y_pred)
   ...: fig = plt.figure(figsize=(4,5))
   ...: img = pics[10*(y_pred[0]),:,:]
   ...: plt.imshow(img, cmap = plt.get_cmap('gray'))
   ...: plt.axis('off')
   ...: plt.title("person {}".format(y_pred), fontsize=16)
   ...: plt.show()
(1, 4096)
Person [0]
```

5-Source Code url:

https://github.com/yousefkhaled282/Face-Recognition-by-Random-forest

6-References:

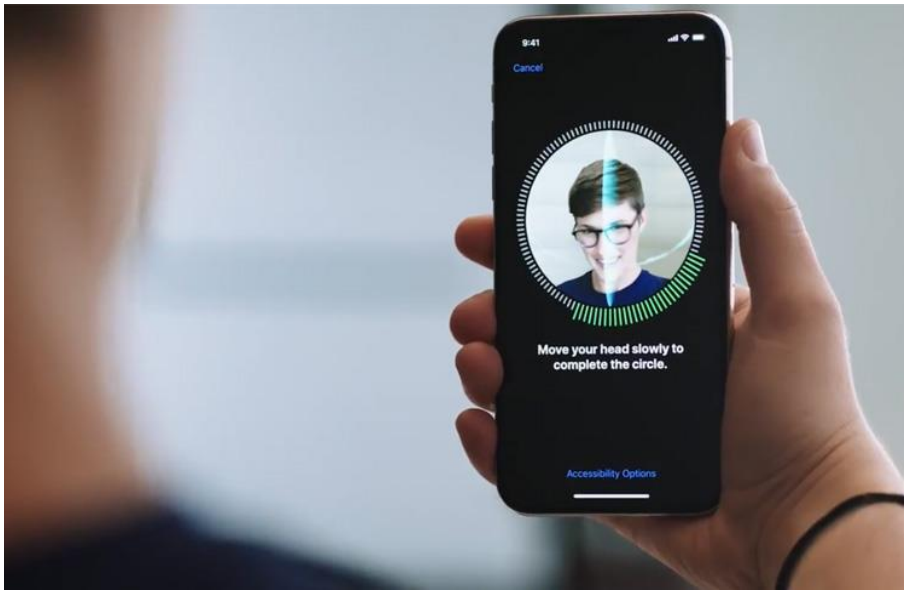
1- https://www.kaggle.com/serkanpeldek/face-recognition-on-olivetti-dataset/notebook

2- https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

3- https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html

7-Similar applications in the market:

1. UNLOCK PHONES



A variety of phones including the latest iPhone are now using face recognition to unlock phones. This technology is a powerful way to protect personal data

and ensure that, if a phone is stolen, sensitive data remains inaccessible by the perpetrator.

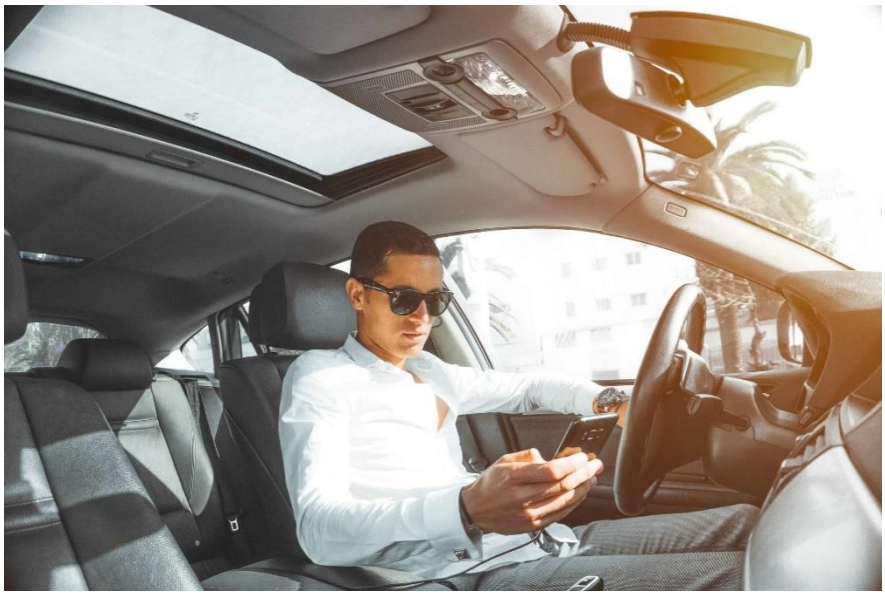
2.Avatar API:

Transform users' faces into 3D animated avatars that can emote.

3. Baby Prediction API:

Predict the future of the users by creating a photo of their potential child

4. RECOGNIZE DRIVERS

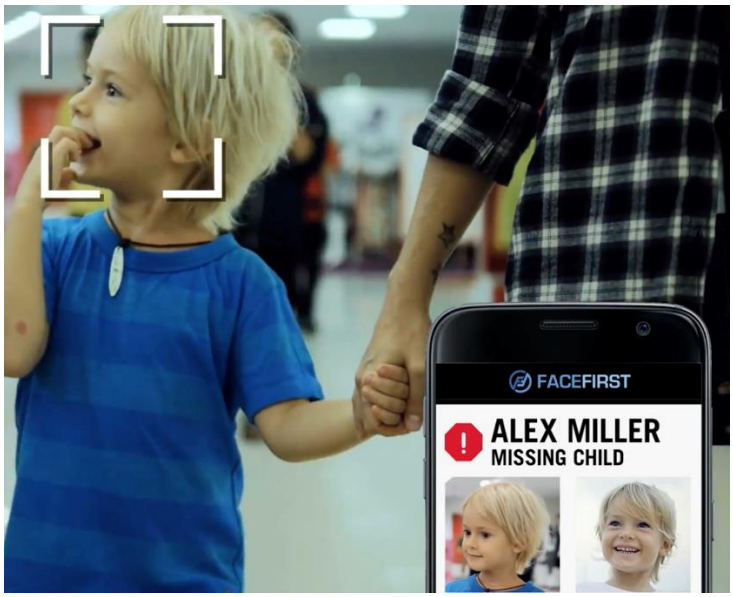


One use of face recognition for automobiles is using a face to replace a key as a means of starting a car. Face recognition can also be used to change radio stations and seat preferences based on who is driving. Face recognition can even make drivers safer by recognizing and alerting drivers if they are drifting off or not focusing on the road.

5.IDENTIFY PEOPLE ON SOCIAL MEDIA PLATFORMS

Facebook uses face recognition technology to automatically recognize when Facebook members appear in photos. This makes it easier for people to find photos they are in and can suggest when particular people should be tagged in photos.

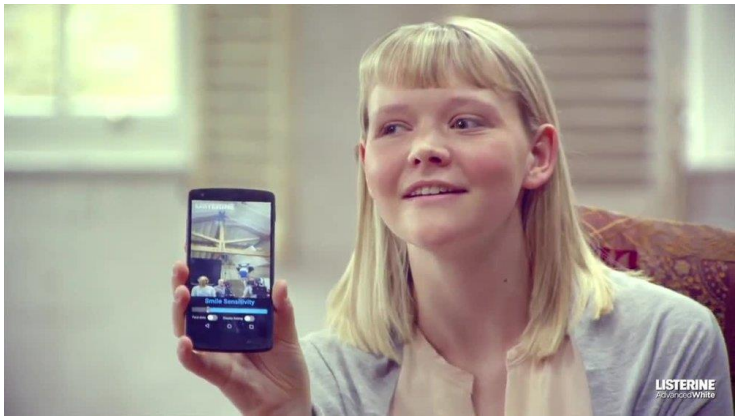
6.FIND MISSING PERSONS



Face recognition can be used to find missing children and victims of human trafficking. As long as missing individuals are added to a database,

law enforcement can become alerted as soon as they are recognized by face recognition—be it an airport, retail store or other public space. In fact, 3000 missing children were discovered in just four days using face recognition in India!

7.HELP THE BLIND



Listerine has developed a ground breaking facial recognition app that helps the blind using face recognition. The app recognizes when people are smiling and alerts the blind person with a vibration. This can help them better understand social situations.

8. AppLock

Sensory, a software company based in California, developed a face recognition app which also works on biometric technology. AppLock allows apps to unlock using facial recognition features as well as voice recognition

8-An initial literature review of Academic publications (papers) relevant to the idea:

Key words:
Random Forest Algorithm, Face detection, Face recognition, Artificial intelligence, Image processing.

Five Academic Papers discussed the same idea:

Paper no.1 :" Detection and Recognition of Human Faces using Random Forests for a Mobile Robot "
RHEINISCH-WESTFÄLISCHE TECHNISCHE HOCHSCHULE AACHEN
-**Supervisor**: PROF. GERHARD LAKEMEYER, PH. D.
-**Second supervisor**: PROF. ENRICO BLANZIERI, PH. D.
-**Advisers**: STEFAN SCHIFFER, THOMAS DESELAERS
-Abstract: This paper present a new framework for the automatic detection, recognition and learning of human faces with random forests. The principal motivation for their work is towards the design of a fast face processing system, especially applicable to mobile robots. Given a high amount of social interaction between a robot and human beings, the experience is rendered more natural if the robot is able to detect faces and recognize corresponding identities. If the identity of the person is not known then it is learned by the robot.
- Conclusion: Thesis presents an integrated framework to the detection, recognition and learning of faces with random forests. Random forests are structurally homogeneous to classical decision trees but randomize the decision criteria. Random forests sample random rectangular regions and learn features from training data. The advisers have an implementation of their framework that is compared to state-of-the-art approaches. Particularly, they compare the performance of face detection with random forests to adaptive boosting. Random forests outperform adaptive boosting when they are trained on the same data. Random forests perform comparatively to adaptive boosting using the training configuration supplied as part of the OpenCV installation. We however note that face detection with random forests suffers from alignment errors. The alignment errors are not tied theoretically to random forests but lie in the error prone merging techniques developed. Merging is influenced by their search for the neighbor range. If the neighbor range is large, multiple faces in the immediate vicinity will be falsely considered to be the same face. If the neighbor range is small, multiple sub-windows for the same face collected from the

multi-resolution analysis and multiple trees will be falsely considered as multiple faces. Thus, They are typically left with a large number of nominated detections in neighboring locations making our post-processing non-trivial. They presented comparisons of face recognition with random forests to face recognition with SVM and log-linear models. They considered 3 test collections and observed that random forests performed within a range of 9%, on average, to both SVM and log-linear models. A critical characteristic of random forests is the training time. Due to the fact that completely random features are sampled, the random tree growth is fast. We reported in Section 6.2.5 that the training of 3 random trees to a depth of 800 nodes with L = 50 takes a millisecond. This feature allows the rebuilding of trees in feasible time. A critical parameter is the forest size. Allowing an increase in the number of trees grown increases the classification accuracy and training time. A trade off is necessary depending on the time constraints.They have shown that with 20 random trees, high classification rates are obtained.

Paper no.2: " Face recognition and detection using Random forest and Combination of LBP and HOG features "

International Conference on Smart Computing and Electronic Enterprise.(ICSCEE2018)

Huda Mady
Dept. Computer Science
Almadinah International University
Kuala lumpur, Malaysia
Huda.lib22@gmail.com

Shadi M. S. Hilles
Dept. Computer Science
Almadinah International University
Kuala lumpur, Malaysia
shadihilless@gmail.com

- Abstract: the effective facial recognition method should perform well in unregulated environments based on video broadcast to satisfy the demands of applications in real-world
However, this still remains a big challenge for most current face recognition algorithms that will affect the accuracy of the system.

This study was conducted to develop face recognition method based on video broadcast under illumination variation, facial expressions, different pose, orientation, occlusion, nationality variation and motion. Viola-Jones algorithm was applied to improve face detection which is these method have proven to detect the faces in an uncontrolled environment in the real world simply and high accuracy. A combination of Histograms of Oriented Gradients (HOG) and Local Binary Pattern (LBP) descriptors was conducted for faces features extraction purpose. These descriptors have proven to be lower computational time. The latest and accurate technique was applied for face classification based on Random Forest classifier (RF). To evaluate the efficiency of the Random Forest classifier, compared it with Support Vector Machine classifiers (SVM) is done with different existing feature extraction methods. Four experiments were implemented on Mediu staff database and excellent results have reported the efficiency of proposed algorithm average recognition accuracy 97.6% The Computer Vision and Image Processing MAT LAB 2016b Toolboxes was used for coding the desired system, data set based on videos.

Paper no.3 : "Performance of Random Forest and SVM in Face Recognition"

Authors: Emir Kremic , Abdulhamit Subasi
Faculty of Engineering and Information Technologies,
International Burch University,
Bosnia and Herzegovina
emir.kremic@ibu.edu.ba , abdulhamit.subasi@ibu.edu.ba

- Abstract: The facial recognition plays a role in computerized system, and its aim is in recognizing individuals by using computerized system. Pattern recognition, machine learning, artificial recognition represents emerging approaches in training of smart mobile computers for human recognition. This is very complex issues, therefore it consist of various different features . In was proposed implementation of PCA for

mobile recognition. This paper we discuss the novel results, whose purpose for the future study and exploration is to perform on mobile machine with much higher rate of accuracy. In here, we are presenting the result of random forest (RF) compared with Support Vector Machine (SVM). Human brains are highly capable of recognizing humans instantly. Emphasize of this research is in training the machine system to be capable to learn and to recognize with the ability that will be closest in accuracy as human. Today, there exists higher need for artificial recognition using face, more than just a finger prints or other biometric features. In general face recognition consists of feature vector matching as described in Principle Component Analysis. In this study we have used 40 human faces and for each faces we used 20 different facial images for each person, which in total are 800 images of 205x274 sizes. The purpose of this research is to bring well designed approach for face recognition; therefore it can be incorporated and tested on real time application for mobile security. In studies conducted before, the images recognition has been done by applying the statistical model of PCA with feature vector extraction or SVM feature vector extraction. The methodology of this research is: images were read from database, 3D color image RGB are read from database then is performed skin color detection, then RGB to gray, histogram, SVM/RF, classify and authentication. The results obtain by accomplishing these steps for SVM 97.94% and Random Forest 97, 17%.

Paper no.4 : " Human face recognition using random forest based fusion of à-trous wavelet transform coefficients from thermal and visible images "

Ayan Seal , Debotosh Bhattacharjee , Mita Nasipuri
Department of Computer Science and Engineering,
PDPM Indian Institute of Information Technology,
Design and Manufacturing Jabalpur, India
Department of Computer Science and Engineering,
Jadavpur University,
Kolkata, India

- Abstract: This paper presents a new image fusion algorithm based on the visible and thermal IR face images, which exploits the advantages of both kinds of images. The proposed fusion algorithm utilizes the translation invariant à-trous wavelet transform and random forest (RF) classifier to decide the contribution of the visible and thermal IR face images in the formation of fused images. The Universal Image Quality

Index is used to evaluate the quality of fused images and results are quite satisfactory. The fused face images are recognized by RF classifiers. The recognition performances of the proposed fusion scheme are 99.07% and 100% for UGC-JU and IRIS benchmark face databases respectively, which is better than those if only visible or thermal IR faces are used.

Paper no.5 : "A Random Decision Forests Approach to Face Detection "

Authors: M. Hassaballah , Mourad Ahmed
Department of Mathematics, Faculty of Science,
South Valley University, Qena 83523, Egypt
m.hassaballah@svu.edu.eg , mourad.ahmed@sci.svu.edu.eg

- Abstract: Face detection has been considered one of the most important areas of research in computer vision due to its wide range of use in human face-related applications. This paper addresses the problem of face detection using Hough transform employed within the random forests framework. The proposed Hough forests-based method is a task-adapted codebooks of local facial appearance with a randomized selection of features at each split that allow fast supervised training and fast matching at test time, where the codebooks are built upon a pool of heterogeneous local appearance features and the codebook is learned for the face appearance features that models the spatial distribution and appearance of facial parts of the human face. Experimental results are included to verify the effectiveness and feasibility of the proposed method.

9-Development Platform:

This Project was developed on a Spyder IDE and was made using python language version 3.7.