Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 08, August - 2017

Implementation and Performance Analysis of Face Recognition System

K. Mary Niharika
ECE Department
Vidya jyothi institute of technology
Hyderabad, India

G. Ravi Kumar
ECE Department
Vidya Jyothi Institute of Technology
Hyderabad, India

***Abstract*— Face recognition system is a computer application for automatically identifying a person from still image or video source. A modern face recognition pipeline consists of four stages: detect face part, face alignment, representation of facial image, and recognition or classification. Though there is a lot of research going on the field of face recognition, achieving high recognition accuracy is still a challenging task for current approaches due to variations in the face images [1]. Recently, deep convolution neural networks set a new trend in the field of face recognition by improving state of art performance [4]. This project addresses deep convolution neural network based face recognition system using OpenFace tool and performance analysis of the system based on pose, illumination variations in the image, and changing size of training dataset. We use histogram of oriented gradients for face detection, aligning faces based on face landmark estimation algorithm and support vector machine for classification.**

Keywords*— *Face representation, face detection, face recognition, training, testing.

I. INTRODUCTION

Face Recognition is a term that includes several sub problems. The input of a face recognition system is always an image or video stream. The output is an identification or verification of the subject or subjects that appear in the image or video. The main aim of face recognition system is to find out efficient and discriminative features irrespective of inter and intra personal variation in the images. Face representation or feature extraction plays a dominant role in the performance of face recognition system. We use deep convolution neural network to get 128 measurements from each face. These measurements are also called as embeddings or features. Euclidean distance between these embeddings directly corresponds to face similarity. Faces of the same person have small distances and faces of distinct people have large distance [4]. General face recognition is as shown in Fig. 1.

Face detection is defined as the process of extracting faces from image or video. It has many applications like face tracking, pose estimation or compression. Face alignment is used to make eyes and mouth as centered as possible. The next step is feature extraction or face representation which involves obtaining relevant facial features from the data. These features could be certain face regions, variations, angles or measures, which can be human relevant (e.g. eyes spacing). This phase has applications like facial feature tracking or emotion recognition. Finally, the system recognizes the face. In an identification task, the system would report an identity from a database. This phase involves a comparison method, a classification algorithm and an accuracy measure.

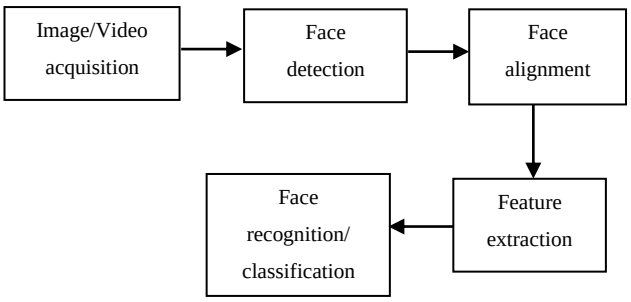

Fig. 1 General face recognition

Image/Video acquisition is the process of collecting real time face images from video source or camera. The image is represented in the form of matrix in digital computer. Preprocessing of image is done if necessary to remove noise or some other enhancements, and then given as input to face detection.

II. FACE DETECTION

Algorithm used to detect face part in image is histogram of oriented gradients (HOG) where each pixel in image replaces by a gradient magnitude and direction [2]. Steps involved in histogram of oriented gradients algorithm is as follows.

A) Compute gradients
B) Cell orientation histograms
C) Block normalization
D) HOG feature

A. Compute gradients

The first step for generating the HOG descriptor is to compute the 1-D point derivatives Gx and Gy in x and y direction by convolving the gradient mask with the gray scale image.

$$Gx = Mx * I, \quad Mx = [\ -1\ 0\ 1\ ]$$

$$Gy = My * I, \quad My = [\ -1\ 0\ 1\ ]T$$

Where I = an input image.

Base on the Gx and Gy values, the gradient magnitude μ and orientation angle θ for each pixel are computed. The

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 08, August - 2017

magnitude and the angle are expressed respectively towards the gradient pixel as follows.

$\mu = \text{sqrt}\ (G_x(x,y)^2 + G_y(x,y)^2)$

$\tan\theta = G_y(x, y)/G_x(x,y)$

At every pixel, the gradient has a magnitude and a direction. For color images, the gradients of the three channels are evaluated. The magnitude of gradient at a pixel is the maximum of the magnitude of gradients of the three channels, and the angle is the angle corresponding to the maximum gradient.

B. Cell orientation histograms

Divide window into non- overlapping cells of size 8*8 pixels. In each cell, compute a histogram of the gradient orientations binned into 9 bins. A pixel whose orientation is close to the bin boundary contributes gradient magnitude between two adjacent bins. Fraction of gradient magnitude decreases linearly with the distance of that pixel's gradient orientation from the two bin centers.

C. Block normalization

Divide the cells into overlapping blocks of size 2*2 cells, so each block has size 16*16 pixels. Consecutive horizontal or vertical blocks overlap by two cells that is block stride is 8 pixels. As a consequence, each internal cell is covered by four blocks. Concatenate the four cell histograms in each block into a single block feature b and normalize the block feature by its Euclidean norm.

D. HOG feature

Calculate image feature vector by multiplying total number of horizontal and vertical blocks in an image, number of cells in each block and total number of bins.

III. FACE ALIGNMENT

Aligning faces in unconstrained scenario is still a difficult problem due to pose, facial expressions. These problems can compensate by using sophisticated alignment technique. Face landmark estimation algorithm [3] is used to extract 68 specific points called landmarks which exist on every face at the top of the chin, the outside edge of each eye, the inner edge of each eyebrow, etc. After detecting the eyes and mouth parts of face, simply rotate and scale the image using affine transformation so that eyes and mouth are centered as best as possible as shown in Fig. 2.

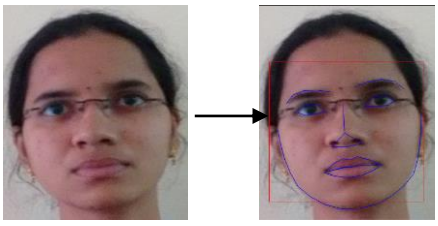

Fig. 2. Face alignment

IV. FACE REPRESENTATION

Today's top performing face recognition techniques are based on convolution neural networks. For example Facebook's DeepFace [5] and Google's FaceNet [4] systems yield the highest accuracy, which are used in this paper. And building blocks of network architecture are,

A) Convolution layer
B) Non linear function
C) Max pooling layer
D) Fully connected layer

A. Convolution layer

The convolution layer performs convolution operation between input image and kernel in the neural network. The main purpose of the convolution operation is to extract features from the input image. The output of the convolution operation is convolved feature or feature map. Convolution preserves the spatial relationship between pixels by extracting image features using small squares of input data.

B. Non lineary

ReLU stands for rectified linear unit and it is non linear operation performed after every convolution. This is an element wise operation (applied per pixel) and replaces all negative pixel values in the feature map by zero.

Output = Max (Zero, Input)

The purpose of ReLU is to introduce non linearity in network architecture, since most of the real-world data would be non-linear.

C. Pooling

Dimensionality of feature map reduced by using operation called pooling. Pooling is also called as sub-sampling or down-sampling. There are mainly three types of pooling: Max, Average, and Sum pooling.

Consider 2*2 window of the image, all elements in that window replaced by maximum value of them in case of Max pooling. Average and sum pooling calculates average and sum of the elements and replace the window elements by that value respectively.

D. Fully connected layer

The Fully Connected layer is a traditional Multi Layer Perceptron that uses support vector machine in the output layer for classification. The term "Fully Connected" implies that every neuron in the previous layer is connected to every neuron on the next layer. The purpose of the Fully Connected layer is to use these features for classifying the input image into various classes based on the training dataset.

This kind of networks has to train before using them. Training of neural network requires a lot of data and computer power. It takes hours to train a network. ImageNet and DeepFace neural networks are trained by OpenFace team and published for direct use [4] [6]. They trained the network by using 3 face images.

a. Take a training face image of a known person
b. Take another picture of the same known person
c. Take a picture of a totally different person

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 08, August - 2017

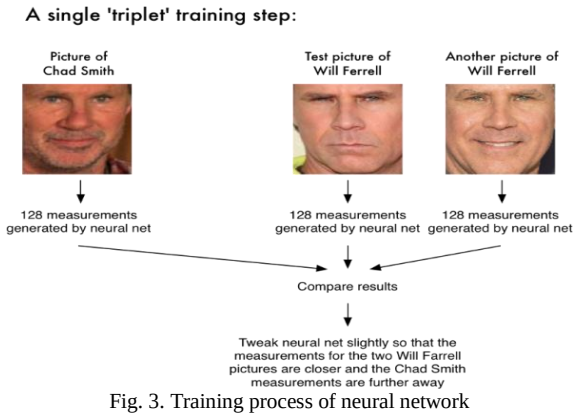

Fig. 3. Training process of neural network

The algorithm generates 128 measurements for each image as shown in Fig. 3. The measurements generated for #1 and #2 are slightly closer while making sure the measurements for #2 and #3 are slightly further apart. This step is repeated for millions of time with billions of images until the neural network learns to reliably generate 128 measurements (embeddings) of each image. Now the network is able to generate 128 measurements for our image as shown below.

128 measurements for my image using trained neural network

These embeddings are represented as points in a sphere as shown in Fig. 4. The purpose of representation is to visualize the embeddings of same person are close compare to other person embeddings (with color coding).

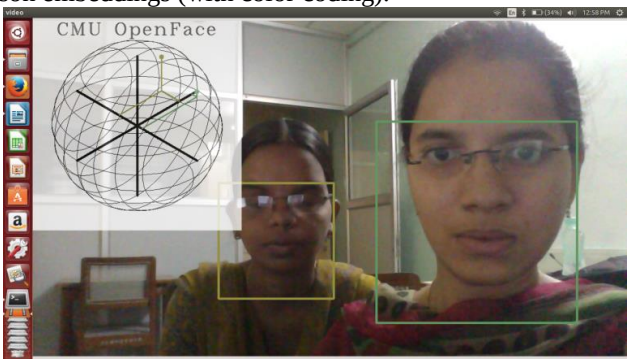

Fig. 4. Embedding represented as points in a sphere

V. CLASSIFICATION OR RECOGNITION

The embeddings are input to the classifier. The classifier is trained to get measurements from unknown image and find the closest match from known person in the database. Training classifier takes milliseconds and output of the classifier is name of the person.

The nearest neighbor classier calculates Euclidean distance (D) between the embeddings of two persons and lower score indicates two faces are more likely same person [4] [6]. The distance between images of Niharika and Prasanna is tabulated in following Fig. 4.

D = getRep(img1) - getRep(img2)

Image 1	Image 2	Distance
Niharika 1	Niharika 2	0.300
Niharika 1	Prasanna 1	0.812
Niharika 1	Prasanna 2	1.037
Niharika 2	Prasanna 1	0.970
Niharika 2	Prasanna 2	1.048
Prasanna 1	Prasanna 2	0.633

Fig. 5. Distance between embeddings of two persons

After comparison process, the system predicts the name of the person from closest match as shown in Fig. 5.

Fig. 5 Displaying name of the person

VI. PERFORMANCE ANALYSIS

The performance of face recognition system is analyzed based on three parameters. They are

A) Training data
B) Pose variation
C) Illumination variation in images

A. Training data

The objective of analyzing system by varying data (training-images) is to know the time taken by the system for training the data. This phase includes alignment, feature extraction and training models. Here training is done for 10 persons with 100 images each, 20 persons and so on up to 100 persons with 100 images each. Time is noted while training 1000 images, 2000 images and so on 10,000 images as shown in Fig.6.

S. No	No of persons	Total no of images		Training time (min)	Testing time (min)
		Train	Test		
1	10	1,000	100	8	4
2	20	2000	200	12	8
3	30	3,000	300	19	12
4	40	4,000	400	26	16
5	50	5,000	500	37	20
6	60	6,000	600	41	24
7	70	7,000	700	45	28
8	80	8,000	800	59	32
9	90	9,000	900	75	36
10	100	10,000	1,000	82	40

Fig.6. Performance of the system based on training database

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 08, August - 2017

B. Pose variation

The performance of the system based on pose variation is done by using dataset of 10 persons. The accuracy is calculated for 10 training images and 50 training images of each person which is tabulated as shown Fig. 7.

S. No	**Person name**	Number of faces recognized for		**Accuracy (%)**
		10 trained images	50 trained images	
1	Achala	9	9	90
2	Harika	10	10	100
3	Majid	10	10	100
4	Prasanna	10	10	100
5	Pratap	9	9	90
6	Praveen	7	7	70
7	Rakesh	7	7	70
8	Ranadheer	10	10	100
9	Sharath	5	5	50
10	Srinu	7	7	70

Fig.7. Performance of the system based on pose variations

Remarks:

a. Face shifted to 90°'s so half part of face is invisible.
b. Face turned towards left side and focusing upward direction or downward direction.
c. Face turned towards right side and focusing upward direction or downward direction.
d. Lips covered by mustache & widely opened mouth.
e. Black shaded eyes & closed eyes.

Note: Increasing the training data does not improve the recognition accuracy.

C. Illumination variation

To calculate recognition accuracy for face images showing illumination variations, considered 100 images for training and 100 images for testing. Recognition accuracy is tabulated as shown in Fig. 8.

S. No.	Person name	Recognized faces	Accuracy (%)
1	Achala	10	100
2	Harika	10	100
3	Majid	10	100
4	Prasanna	6	60
5	Pratap	10	100
6	Praveen	9	90
7	Rakesh	6	60
8	Ranadheer	10	100
9	Sharath	10	100
10	Srinu	10	100

Fig. 8 Performance of the system based on illumination

Remarks:

a. Images taken under dim light
b. Half part of face is dark
c. Dark background

Note: Increasing the training data does not improve the recognition accuracy.

VII. CONCLUSION & FUTURE SCOPE

Implementation of face recognition system is done by using deep neural network based feature extraction method successfully. The performance of the system is calculated based on training data, pose and illumination variations. The system gives 100% accuracy for frontal view faces.

There are a number of problems involved in a uni-modal biometric systems such as Noisy sensed data, Intra-class variations etc. The solution is using a multimodal biometric system which uses multiple biometric traits to offer robust decision-making.

REFERENCES

[1] Ming-Hsuan Yang, D. J. Kriegman, N. Ahuja. Detecting faces in images: a survey. s.l. : in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002. pp. 34-58. Vol. 24

[2] Navneet Dalal, Bill Triggs. Histograms of Oriented Gradients for Human Detection. France : IEEE, 2005. pp. 886-893. Vol. 1.

[3] Vahid Kazemi, Josephine Sullivan. One millisecond face alignment with an ensemble of regression trees. Columbus : IEEE, 2014. 978-1-4799-5118-5.

[4] Florian Schroff, Dmitry Kalenichenko, James Philbin. FaceNet: A Unified Embedding for Face Recognition and Clustering. Boston : IEEE, 2015. 1063-6919.

[5] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich. Going Deeper with Convolutions. USA : IEEE, 2015. 1063-6919.

[6] Brandon Amos, Bartosz Ludwiczuk, Mahadev Satyanarayanan. OpenFace: A general-purpose face recognition library with mobile applications. Pittsburgh : CMU-CS-16-118, 2016.