

IDENTIFYING INDIVIDUALS FROM A VIDEO SEQUENCE BY COMBINATION OF DCT AND ARTIFICIAL NEURAL NETWORKS

Kamal Shah¹, Kekre H.B.²

¹Department of Information Technology, St Francis Institute of Technology, Mumbai, India

²Department of Computer Engineering, Thadomal Shahani College of Engineering, Mumbai, India

E - Mail : shah.kamal@yahoo.com

Abstract

The goal of this project is to create neural network-based face detection system to identify people from a video sequence. A database containing the images of the people to be identified in the video sequence is created. Using the face detection code the faces inside the video sequence are cropped for recognition. It will eliminate background influence as much as possible. The face space is described by a set of feature vectors obtained using the discrete cosine transform (DCT). This DCT feature vector is then fed into Artificial Neural Networks(ANN) which are used to recognize the faces through learning correct classification.

Key words: Artificial Neural Network, DCT, Back Propagation Algorithm, Motion Detection, Head Tracking

I. INTRODUCTION

In today's world, the need to maintain the security of information or physical property is becoming both increasingly important and increasingly difficult. From time to time we hear about the crimes of security breaches at offices, banks and airports. It has become a necessity to develop an effective face detection system to keep a track of people entering and exiting a particular premise in order to recognize these criminals.

Some possible applications for automatic face detection and identification can be summarized on supervision and security applications, videoconferences, animation of facial expressions, as well as remote control camera applications. Image processing algorithms joint with Artificial Neural Network (ANN) technology act as valuable assistance tools, which can render feasible advanced solutions in the problem of face detection.

Many approaches to the overall face recognition problem have been devised over the years. This project aims at devising a new face recognition approach by using the Discrete Cosine Transform (DCT) and Artificial Neural Networks (ANN) classification model. The DCT is employed to extract the input features to build a face recognition system, and the flexible neural tree is used to identify the faces.

This project aims at identifying individuals present in a video database from a video sequence . Upon inputting an unknown face image extracted from the video, our face recognition system seeks to quickly and effectively determine whether or not it matches a known individual present in the database.

II. PREPARATION OF DATABASES

A. Video set 1

Description: this is the NRC-IIT video database that contains pairs of short video clips each showing a face of a computer user sitting in front of the monitor exhibiting a wide range of facial expressions and orientations as captured by a Intel webcam mounted on the monitor of computer . The video capture resolution is kept to 160 x 120. With the face occupying 1/4 to 1/8 of the image (measured by width).

Setup: For this Database, two video clips of each person are taken one after another, under approximately the same illumination conditions, the same setup and almost the same background, for all persons in the database. This database is thus most suited for testing the recognition performance with respect to such inherent to video-based recognition factors as low resolution, motion blur, out-of focus factor, facial expression variation, facial orientation variation and occlusions.

Characteristics

- ☞ Resolution: 160x120.
- ☞ Average file size: 780KB
- ☞ Average duration: 5secs
- ☞ Video type and compression: colour AVI, 20.0 fps. Intel webcam-provided codec compressed at 481 Kbps.
- ☞ Average total number of frames in a clip: 100.
- ☞ Total no of individuals:11
- ☞ Average total number of frames in a clip: 100.



Fig .2 Testing video set1 (video not n database)

B. Video set 2

Description: This is our database. It contains pairs of short video clips each showing a face exhibiting a wide range of facial expressions and orientations as captured by a 5.1 megapixel digicam. The video capture resolution is 640x480. With the face occupying 1/3 to 1/4th of the image (measured by width).

Setup: For this Database, two video clips of each person are taken one after another, under approximately the same illumination conditions, the same setup and almost the same background, for all persons in the database.

Characteristics:

- ☞ Resolution: 640 x 489
- ☞ Average file size: 1.5 MB
- ☞ Average duration: 10 - 20 secs.
- ☞ Total no. of frames in a clip: 300.
- ☞ Total number of video clips: 6



Fig .5 Database of video set2

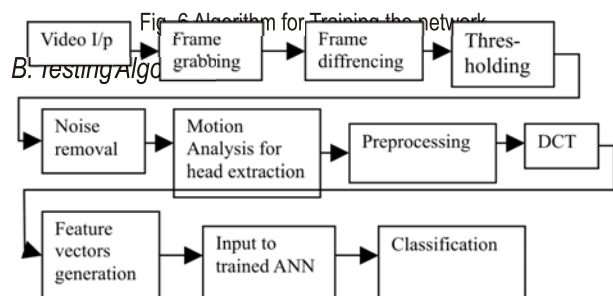
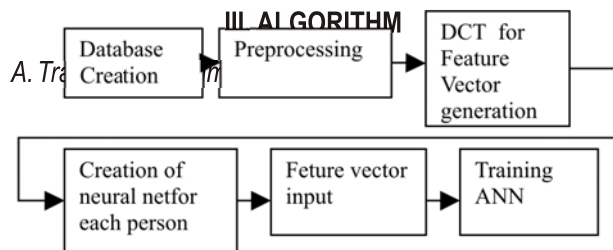


Fig .7 Algorithm for testing the network

TRAINING PHASE' flow chart description

1. *Database:* It contains face images of individuals to

be recognized. Different facial expressions per person taken under varying illumination conditions.

2. **Preprocessing:** Database images are normalized and resized.
3. **DCT:** Image divided into subblocks and blockwise DCT computed. High frequency coefficients discarded to ensure data compactness.
4. **Feature vector generation:** DCT coefficients are arranged in a zig-zag manner and feature vectors of the images are obtained.
5. **Neural network:** A MLF Neural network is created for each person in the database. Number of input neurons is equal to the number of elements in the DCT feature vector of a face and a single output neuron. Backpropagation algorithm is used for training.

B. 'TESTING PHASE' flow chart discription:

1. **Video I/p:** A video sequence is acquired. The video sequence has minimal illumination changes.
2. **Frame grabbing:** In this block the video sequence is cut into frames. The interval between two frames is predefined by the programmer.
3. **Motion estimation:** This consists of four sub-blocks.
 1. **Thresholding:** A threshold value is assigned by the programmer and a difference of pixel value greater than this threshold is considered
 2. **Noise removal:** noise from the frames is removed .
4. **Motion analysis:** The individuals present in the video are located
 1. **Head location and extraction:** In this block using an appropriate transformation head region is located. Once the head region is obtained head is extracted from it for further identification.
 2. **Pre-processing:** The face detected is normalized in a similar way as the database images and is given to DCT block for generating DCT coefficients.
5. **DCT:** DCT coefficients of the test image is computed in a similar manner.
6. **Feature extraction:** the DCT coefficients obtained from the DCT block are used for generating feature vectors.

Neural network as classifier. The feature vectors are given in the form of 1D matrix as an input to the trained neural network block. The MLF-ANN acts as a classifier and identifies whether the test image is in the database or not.

Here the extracted test face of video 02-1 is fed into the

ANN. ANN calculates its weights using the training function TRAINSCG and compares it with the weights of the trained faces. If the face exists in the database then the maximum weights matching will correspond to the test face else if the face is not in database the value of maximum matching weights will fall below threshold .In this case face 02-1 has the max weight match.

IV. RESULT ANALYSIS

A. Video set1

A.1 Face from database

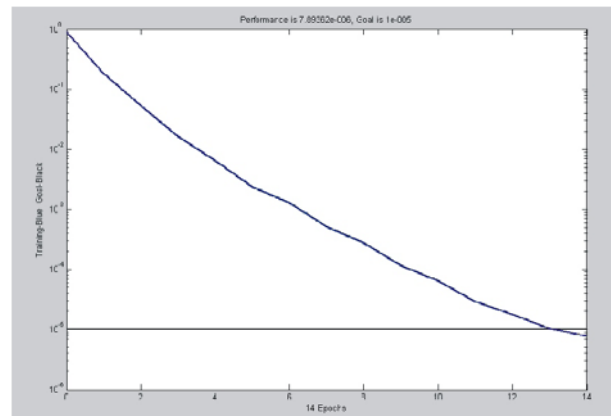
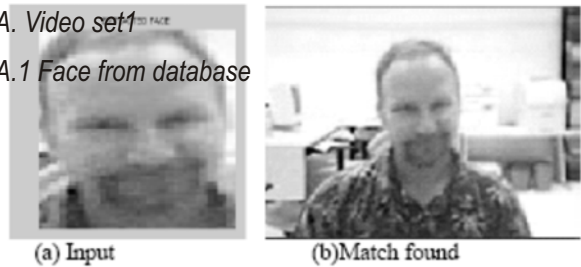


Fig .9 Performance Analysis (No. of epochs V/S Training goal

A.2 Face not from database



Fig .10 Match found with face02

R. Video set2

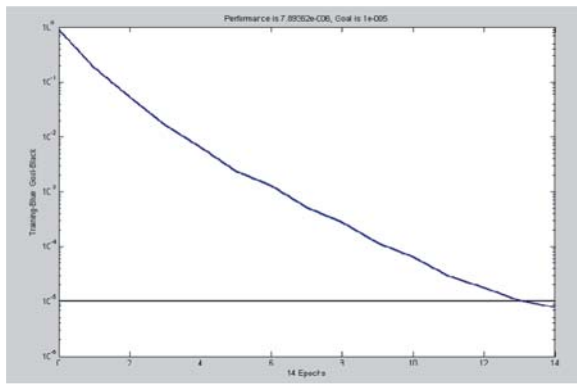
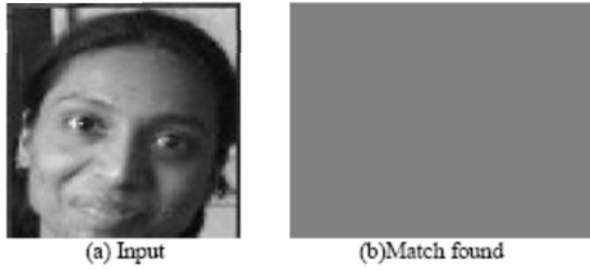


Fig .12 Performance Analysis (No. of epochs V/S Training goal

B.2 Some special cases and conditions

B.1 Case1: Here image from video set2 is selected and artificial noise is introduced as shown in fig 13 (a) & (b)

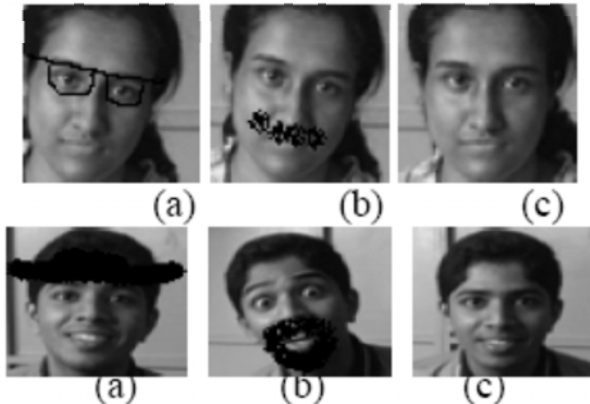


Fig .13 (a) & (b) Occluded image (c) original image

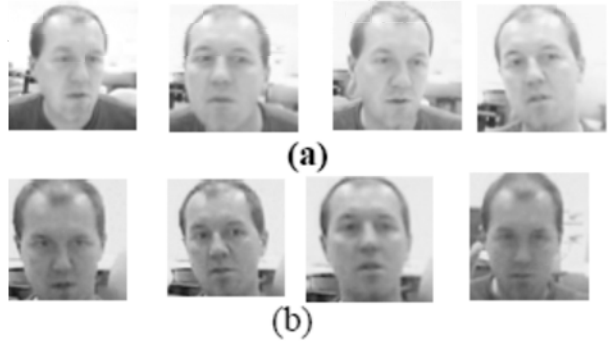


Fig .14 (a) images from training set (b) test images with different illumination

B.4 Case 4. In video set1 face01 has only frontal facial



Fig .15 (a) & (b) test images with head rotation (c) Match found with video set1 face01 (d) Match found with video set1 face05

V. CONCLUSION

Neural networks are an efficient, pervasive, and powerful means of computation. Neural networks are pattern classifiers. They do not store "knowledge" in a memory bank. The information is distributed throughout the network and is stored in the form of weighted connections. The most valuable characteristics of neural networks are adaptability and tolerance to noisy data.

We have built a face recognition system based on local appearance based face recognition approach, which utilizes the block based discrete cosine transform features with neural networks as the classifier. This method was

found to be robust enough to account for changes in facial expressions and addition of accessories .

REFERENCES

- [1] Conrad Sanderson, Kuldip K. Paliwal “.Features for robust face-based identity verification’,. Signal Processing 83(2003) 931-940.
- [2] R.Chellappa,C.L.Wilson, and S.Sirohey “,Human and machine recognition of faces ..A survey”, Proceedings of the IEEE 83(5):705–740, 1995.
- [3] Henry A. Rowley, Shumeet Baluja, Takeo Kanade “.Neural Network Based Face Detection.”,In Computer Vision and Pattern Recognition, 1996.
- [4] “Jun Wang, Mohan S Kankanhalli, Philippe Mulhem, Hadi Hassan Abdulredha “,Face Detection Using DCT Coefficients in MPEG
- [5] Christine Podilchuk and Xiaoyu Zhang . Face

recognition using dct-based feature vectors. Signal Processing Research Department, AT&T Bell Laboratories



- [6] Gaurav Aggarwal , Amit K. Roy Chowdhury, Rama Chellappa “.A System Identification Approach for Video-Based Face Recognition.”

Ms. Kamal Shah - Assistant Professor- Department of Information Technology, St.Francis Institute of Technology, Mumbai, India. Her research topic is Artificial Neural Networks. She has presented 2 papers in national and international conferences.