

Facial Expression Recognition Using Artificial Neural Networks

Deepthi.S¹, Archana.G.S², Dr.Jagathy Raj.V.P³

¹(Head of Section, Model Polytechnic College, Karunagappally, Kerala, India)

²(Lecturer in Electronics, Model Polytechnic College, Karunagappally, Kerala, India)

³(Professor, Cochin University of Science and Technology, Kochi, Kerala, India)

Abstract: In many face recognition systems the important part is face detection. The task of detecting face is complex due to its variability present across human faces including colour, pose, expression, position and orientation. So using various modeling techniques it is convenient to recognize various facial expressions. In the field of image processing it is very interesting to recognize the human gesture by observing the different movement of eyes, mouth, nose, etc. Classification of face detection and token matching can be carried out any neural network for recognizing the facial expression. Facial expression provides vital cues about the emotional status of a person. Thus an automatic face expression system (FER) that can track the human expressions and correlate the mood of the person can be used to detect the deception among humans. Other applications of automatic facial expression recognition system include human behavior interpretation and human computer interaction. This paper proposes a method using artificial neural networks to find the facial expression among the three basic expressions given using MATLAB (neural network) toolbox

Keywords: Artificial Neural Networks, Image processing, Discrete Fourier Transform (DCT)

I. Introduction

A facial expression is a visible manifestation of the affective state, cognitive activity, intention, personality and psychopathology of a person. It plays a communicative role in interpersonal relations. Facial expression constitutes 55 percent of the effect of a communicated message and is hence a major modality in human communications.

Automatic recognition of facial expression can be an important component of human machine interfaces. It may also be used in behavioral science and clinical practices. In order to enhance the communication between man and human like robots, it is important to recognize and understand facial expressions.

1.1 Image processing:

Image processing a form of signal processing in which the input is an image, and the output can be either an image or a set of characteristics or parameters related to that image. A digital image is composed of a finite number of elements, each of which has a particular location and value. These elements are referred to as picture elements, image elements or pixels. An image is digitized to convert it to a form which can be stored in a computer's memory or on some form of storage media, such as a hard disk or CD ROM. This digitization procedure can be done by a scanner, or by a video camera connected to a frame grabber board in a computer. Once image has been digitized, it can be operated upon by various image processing operations.

Some applications of Image processing are

1. Robotics
2. Medical field
3. Graphics and animation
4. Satellite imaging

1.2 Image processing techniques

Image processing techniques are used to enhance, improve or alter an image and to prepare it for image analysis. Image processing is divided into many sub processes, including Histogram analysis, thresholding, masking, edge detection, segmentation and others.

1.3 Real time image processing

In many of the techniques considered so far the image is digitized and stored before processing. In other situations, although the image is not stored, the processing routine requires long computation times before they are finished. This means that in general there is a long lapse between the time an image is taken and the time a result is obtained. This may be acceptable in situations in which the decisions do not affect the processing. In other situations there is a need for real time processing such that the results are available in real time or in a short enough time to be considered real time processing

1.4 Image processing toolbox:

Matlab is complemented by a family of application specific solutions called toolboxes. The image processing toolbox is a collection of MATLAB functions (called m-functions or m-files) that extend the capability of the MATLAB environment for the solution of digital image processing problems .Other toolboxes that used to complement the image processing tool boxes are the signal processing,NeuralNetworks,Fuzzy logic and wavelet Toolboxes

1.4.1 Image Types:

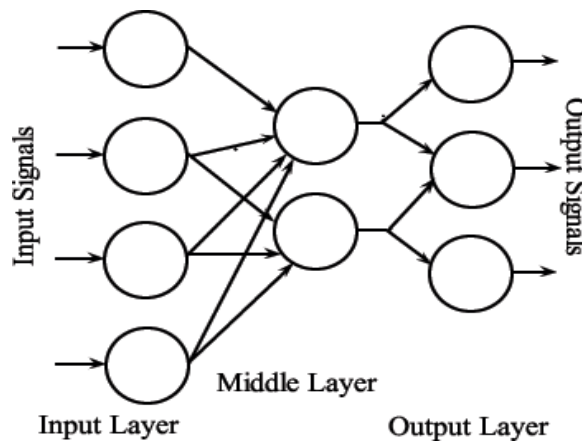
The toolbox supports four types of images. They are grey scale images, binary images, indexed images and RGB images. A gray scale image is a data matrix whose value represents shades of gray. A binary image is a logical array of 0s and 1s.An indexed image does not explicitly contain any colour information .It pixel values represent indices into a colour lookup table .Colours are applied by using these indices into a colour lookup table (LUT).Colours are applied by using these indices to look up the corresponding RGB triplet in the LUT.An RGB image is a three dimensional byte array that explicitly stores a colour value for each pixel.RGB image array are made up of width,heightand three channels of colour information.

II. Artificial Neural Network

An Artificial Neural Network often just called a neural network is mathematical model inspired by biological neural networks. A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. In most cases a neural network is an adaptive system that changes its structure during a learning phase. Neural networks are used to model complex relationships between inputs and outputs or to find patterns in data.

2.1 Architecture of typical artificial neural networks

Artificial neural networks can be defined as model reasoning based on the human brain. The brain consists of a densely interconnected set of nerve cells or basic information processing unit called neurons. An artificial neural networks consists of a number of very simple processors, also called neurons, which are analogous to the biological neurons in the brain



2.2 The Perceptron

The Perceptron is a type of artificial neural network. It can be seen as the simplest of feed forward neural network; a linear classifier. The perceptron is a binary classifier that maps its input x (a real valued vector) to an output value f(x) (a single binary value) across the matrix.

$$F(x) = 1, \text{ if } w \cdot x + b > 0$$

$$= 0, \text{ elsewhere}$$

w is a vector of real valued weights and w.x is the dot product (which compute the weighted sum) .b is a “bias” a constant term that does not depend on any input value. The value of f(x)(0 or 1) is used to classify x as either a positive or a negative instance in the case of a binary classification problem .The bias can be thought of as offsetting the activation function ,or giving the output neuron a base level of activity. If b is negative, then the weighted combination of inputs must produce a positive value greater than -b in order to push the classifier neuron over the 0 threshold

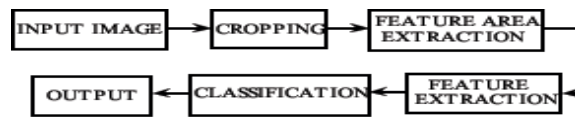
2.3 Advantages of Neural Networks

- * A neural network can perform tasks that a linear program can not
- * When an element of neural network fails, it can continue without any problem by their parallel nature.
- * A neural network learns and does not need to be reprogrammed.
- * It can be implemented in any application
- * It can be implemented without any problem

2.4 Disadvantages

- * The neural needs training to operate
- * The architecture of neural network is different from the architecture of microprocessors therefore needs to be emulated
- * Requires large processing time for large neural networks

III. Scheme of Work



Fig(2) Scheme of work

3.1 Input Image

Images used for facial expression recognition are static images. With respect to the spatial, chromatic and temporal dimensionality of input images, 2D monochrome facial image sequences are the most popular type of pictures used for automatic expression recognition. A data base of facial expression images is collected. Three expressions happy, sad and neutral is taken and identified. Fig(3)shows some sample images .3



Fig (3) sample images

3.2 Cropping

Cropping of image is done in order to obtain the specific portions required for expression recognition. The regions including eye and mouth are cropped out for this. Fig(4) shows some cropped images



Fig (4) cropped images

3.3 Feature area extraction

Two feature areas are selected to find out expression of image. Here we select the eye and mouth areas Fig (5)shows an example



Fig (5) Extracted eye and mouth areas

3.4 Feature extraction

The ultimate aim of the image processing application is to extract important features from image data. 2D DCT is used for feature extraction. In transformed image significant number of coefficients has small magnitudes and can discard entirely without losing image features.

IV. Discrete Cosine Transform

A Discrete Cosine Transform (DCT) expresses a sequence of finitely many data points in terms of a sum of cosine functions oscillating at different frequencies. DCTs are important to numerous applications in science and engineering, from loss compression of audio and images, to spectral methods for the numerical solution of partial differential equations. The use of cosine rather than sine function is critical in these applications: for compression, it turns out that cosine functions are much for efficient, where for differential equations the cosines express a particular choice of boundary conditions.

Face images have high correlation and redundant information which causes computational burden in terms of processing speed and memory utilization. The DCT transforms images from the spatial domain to the frequency domain. Since lower frequencies are more visually significant in an image than higher frequencies, the DCT discards high-frequency coefficients and quantizes the remaining coefficients. This reduces data volume without sacrificing too much image quality.

The DCT and in particular the DCT II is often used in signal and image processing especially for loss data compression, because it has an energy compaction. Most of the signal information tends to be concentrated in a few low frequency components of the DCT.

The DCT can be regarded as a discrete time version of the Fourier cosine series. It is a close relative DFT, a technique for converting a signal into elementary frequency components. Unlike DFT, DCT is real valued and provides a better approximation of a signal with fewer coefficients.

The formulae for a 2D DCT:

$$: C_u = \begin{cases} \frac{1}{\sqrt{2}}, & \text{if } u = 0 \\ 1, & \text{else} \end{cases}$$

$$: C_v = \begin{cases} \frac{1}{\sqrt{2}}, & \text{if } v = 0 \\ 1, & \text{else} \end{cases}$$

$$F_{UV} = \frac{1}{4} C_u C_v \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} S_{xy} \cos(u\pi \frac{2y+1}{2N}) \cos(v\pi \frac{2x+1}{2N}) \tag{1}$$

4.1. IDCT

The IDCT function is the inverse of the DCT function. The IDCT reconstructs a sequence from its discrete cosine transform (DCT) coefficients. To rebuild an image in spatial domain from the frequencies obtained, we use the IDCT formulae:

$$: C_u = \begin{cases} \frac{1}{\sqrt{2}}, & \text{if } u = 0 \\ 1, & \text{else} \end{cases}$$

$$: C_v = \begin{cases} \frac{1}{\sqrt{2}}, & \text{if } v = 0 \\ 1, & \text{else} \end{cases}$$

$$: S_{xy} = \frac{1}{4} \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} C_u C_v F_{uv} \cos(u\pi \frac{2y+1}{2N}) \cos(v\pi \frac{2x+1}{2N}) \tag{2}$$

The 2-dimensional DCT of all the cropped images are computed using the Matlab DCT2. Then a 20*20 matrix consisting of comparatively higher coefficients are selected and finally from that matrix, 20 values are selected. This is done both for the eye portions and for the mouth portion. Table (1) shows the DCT values for selected portions extracted from the sample picture below



Table.1

Extracted Area	DCT Values
EYE PORTION	10674,-1087,-2922,-306,907,80, 1473,46,348,8,238,41, 80,28, -14,149,380, -2922,-136,-344
MOUTH PORTION	6602.1,-882.7, 842.3, 653.1, 287.1,-21.4, -180.2,-55.4,-64.2, 134.9, 441.4, -9.2, 54.6, 48.6, 24.7,-10.4, 842.3,-18.5,-83.2,-228.1

All these values are appended together to form a single matrix, whose values are then scaled down.

V. Classification

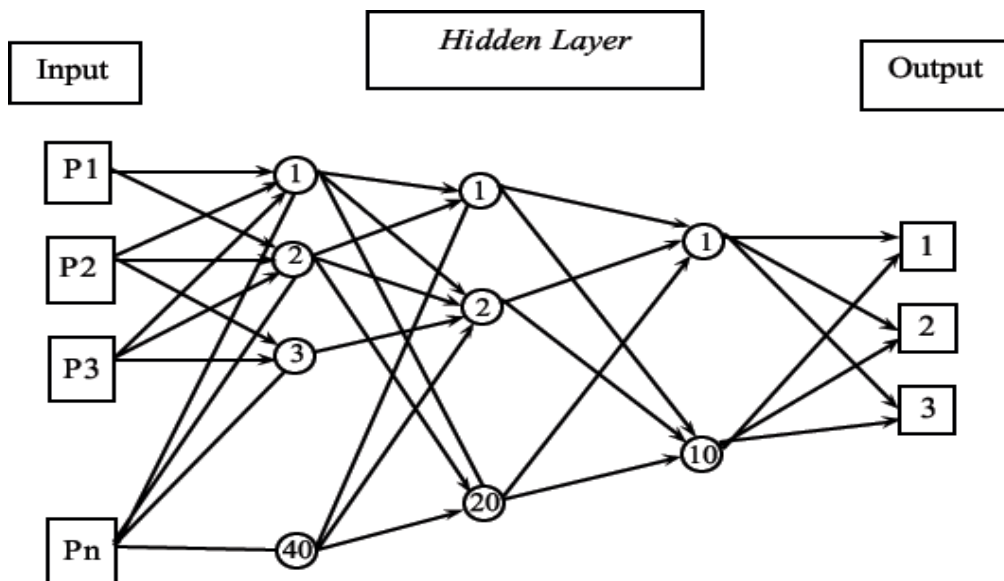
The neural network is a five layered perceptron where the output layers have three neurons for happy, sad and normal classes each. The number of neurons is 40, 20, 10 and 3 respectively. The network is trained in supervised manner using input and target. Here inputs are 40 coefficients of DCT of mouth and eye of happy, neutral and sad images. Targets are user defined for three expressions .We used [100] for happy,[010]for neutral and [001] for sad. Multilayer layer perceptron have trained with conjugate gradient method.

VI. Network Architecture

The number of neutrons is 40, 20, 10(hidden) and 3respectively.Standard back propagation is a gradient descent algorithm, in which the network weights are moved along the negative of the gradient of the performance function .Back propagation refers to the manner in which the gradient is computed for nonlinear multilayer networks.

There are generally 4 steps in training process. They are

1. Assemble the training data
2. Create the network object
3. Train the network
4. Simulate the network response to new input



VII. Conjugate Gradient Method

The basic back propagation algorithm adjusts the weights in the steepest descent direction (negative of the gradient).This is the direction in which the performance function is decreasing most rapidly. It turns out that, although the function decreases most rapidly along the negative of the gradient, this does not necessarily

produce the fastest convergence. In the conjugate directions, this produces generally faster convergence than steepest descent directions. There are different variations of conjugate gradient algorithms.

In mathematics, conjugate gradient method is an algorithms for the numerical solution of particular systems of linear equations, namely those whose matrix is symmetric and positive definite. The conjugate gradient method is an iterative method, so it can be applied to sparse systems that are too large to be handled by direct methods such as the Cholesky decomposition. Such systems often arise when numerically solving partial differential equations.

The conjugate gradient method can also be used to solve unconstrained optimization problem such as optimization. The biconjugate gradient method provides a generalization to non symmetric matrices. Various nonlinear conjugate gradient methods seek minima of nonlinear equations.

VIII. Training

Once the network weights and biases have been initialized, the network is ready for training .The training process requires a set of examples of proper network behavior inputs “p” and target outputs”t”.During training, the weights and biases of the network are iteratively adjusted to minimize the mean square error –the averaged squared error between the network outputs and the target outputs “t”

IX. Output

By testing of image we compare the outputs of output layer neurons with the predetermined targets and if they match, corresponding expressions is displayed. Targets are user defined for three expressions .We used [100] for happy, [010] for neutral and [001] for sad.

X. Conclusion

This paper has briefly overviewed automatic facial expression.2-D monochrome facial images are the most popular type of pictures used for automatic expression recognition. In this project expression recognition is done using neural networks. Neural networks tend to be black box, they will train and achieve a level of performance but we cannot easily determine how they are making the decision. Neural networks are invaluable for applications where formal analysis would be difficult or impossible, such as pattern recognition and nonlinear identification and control.

In the proposed expression recognition system, the image should be such that the head must be straight. The image should be properly illuminated. We can't the expression of an individual having beard. This project is designed for only three expressions, happy, sad and neutral.

The project can have numerous applications in the field of man- machine interactions and in emotion related research to improve the processing of emotion data.

REFERENCES

- [1] “Facial Expression Recognition: A Brief Tutorial overview”, C.C. Chibelushi and F. Bourel 2002.
- [2] F. Bourel, C.C. Chibelushi, A.A. Low, "Robust Facial Expression Recognition Using a State-Based Model of Spatially-Localized Facial Dynamics", Proc. Fifth IEEE Int. Conf. Automatic Face and Gesture Recognition, pp.106-111, 2002
- [3] Facial expression recognition technique using 2D DCT and k means algorithm, Liyang Ma and K.Khorasani, IEEE Transactions on Systems, Man and Cybernetics, Vol.34, No.3 June 2004
- [4] V. Bruce, "What the Human Face Tells the Human Mind: Some Challenges for the Robot-Human Interface", Proc. IEEE Int. Workshop Robot and Human Communication, pp. 44-51, 1992
- [5] “Facial Expression Recognition using Back propagation”, A. Sulistijono, Z. Darojah, A. Dwijotomo, D.Pramdihanto 2002
- [6] Zhang Z. Feature-Based Facial Expression Recognition Experiments With a Multi-Layer Perceptron. International Journal of Pattern Recognition & Artificial Intelligence. 1999; 13:893–912.
- [7] Ekman P, Friesen WV. The Facial Action Coding System A Technique for the Measurement of Facial Movement. San Francisco: Consulting Psychologists Press; 1978.
- [8] Sebe N, Lew MS, Sun Y, Cohen L, Gevers T, Huang TS. Authentic Facial Expression Analysis. Image and Vision Computing. 2007; 25:1856–1863.