Face Recognition and Emotion Detection Using Restricted Boltzmann

CAROLINE EL FIORENZA[1]

SIKTA SARKAR[2],SANCHAREE CHATTOPADHYAY[3],DEVI SUBADRA.V[4]

[1]Guide,[2,3,4]Student

Department of Computer Science and Technology ,SRM Institute of Science and Technology,Ramapuram,Chennai-600089,India.

ABSTRACT The new era of machine learning and artificial intelligence has led to the mushrooming of a new and fascinating field of face recognition which is simultaneously calling for the need of emotion and motion detection. The main driving force behind the enormous interest in this field is chances of approaching utmost security and accuracy in real time applications .In the proposed article detection of human faces and basic human emotions using face recognition and image processing has been proposed. The system promises to provide utmost level of recognition ratio and security irrespective of climate changes, temperature changes , optical changes and distance. It initially deals with intake of the input and analysing it with the help of Restricted Boltzmann machine(RBM) and MMI dataset which contains the samples of human emotions . After the proper execution, the application responds by detecting the emotion and expressions of the user and would also protect the domain from unauthorised access.

INDEX TERMS MMI dataset,Facial RecognitionTechnology(FERET) database ,Restricted Boltzmann Machine(RBM).

INTRODUCTION Face recognition has been a matter of extensive research for several decades.The field deals with robotics , biometry , pattern recognition and computer vision.

The mathematical algorithms of biometric facial recognition follow several stages of image processing : capture , the first step for the system to collect physical and

behavioural samples in predetermined conditions and during a stated period of time .Then comes extraction of the gathered data from samples to create templates based on them .The process of comparison comes into play in which the collected data will be compared with the existing templates which are already there in the data base .The final stage of face detection technology is to make a decision whether the face features of a new sample was collected and will check whether matching with the one which are already there in the facial database or not .It usually takes only a second .[1]

Face recognition algorithms has two main sub-divisions :the Analytic or Feature based approach and Holistic approach . The analyzes of features based approach computes a whole set of geometrical features from the face such as the eyes ,nose and the mouth .The holistic or appearance based approach considers the global properties of a human face pattern .In this ,the face is recognised as a whole without using only certain fiducial points obtained from different regions of the face .Despiteof automatic face recognition techniques ,recognition based only on the visual spectrum has difficulties performing in uncontrolled operating environments .It highly depends on illumination condition ,angle of view ,evenness of light on face ,makeup ,shadows and glint .Hence ,thermal IR is used .Thermal IR images represent the heat patterns emitted by veins and tissue structure of a face ,hence it remains unique for each and every face .Thermal IR imagery is independent of ambient illumination since the human face and body is an emitter of thermal energy . [2]

Information fusion utilizes a combination of different sources of information mainly for two purposes - first being generation of one representational format and second to reach a decision .Information fusion deals with committee machines ,consensus building , integration of

IJSER

multiple sensors , multi-modal data fusion and a combination of multiple experts/classifiers .The concept can be divided into three main parts : sensor data level(low level) , feature level(intermediate level) and decision level(high level) .Data fusion(low level) deals with combining several sources of raw data to give a new data that is more informative than the inputs .Feature level fusion (intermediate level) combines various features such as edges , lines ,corners ,texture parameters and combine in a fused feature map used for segmentation and detection .Decision fusion level(high level) deals with combining decisions received from several experts .This includes ranked list combination , AND fusion,OR fusion and majority voting .[2]

The field of face recognition is triggering the need of emotion and motion detection .There are seven basic types of emotions universally recognised across different cultures : disgust ,fear ,anger ,sadness ,happiness ,contempt ,surprise .More complex emotions are often a combination of these seven basic emotions .The main problems faced by emotion detection are : first ,classification of emotion might be difficult depending on the state of the input image that is static or in transition frame and second being existence of large databases containing training or sample images .[3]

In this article ,Restricted Boltzmann Machine (RBM) algorithm has been used to detect human faces and emotions . RBM is an algorithm used to reduce dimensionality ,filter collaboratively ,regress ,learning feature and topic modelling .RBM is a two layered neural net ,the first layer also called as the visible or the input layer and the second being the hidden layer .They together constitute the building blocks of deep-belief networks .Each circle in the figure below shows a neuron like unit called the node .Nodes are the places where calculations actually take place .Cross layered nodes are inter-connected but same layered nodes are not linked .The term ‘Restricted’ in RBM is because of the absence of any communication within the intra layer .[5]

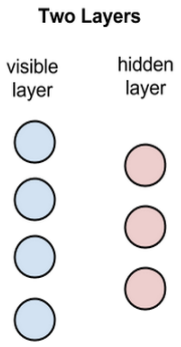

FIGURE 1. Two layered basic RBM structure

Let us assume that x is a single pixel value through the two layered net .At the first node of hidden layer ,x is multiplied by a weight(w) and added to bias (b) .The result hence derived serves as an input of an activation function that produces the nodes output or the strength of the signal through it ,where input is x .[5]

activation f((weight w* input x)+bias b) = output a

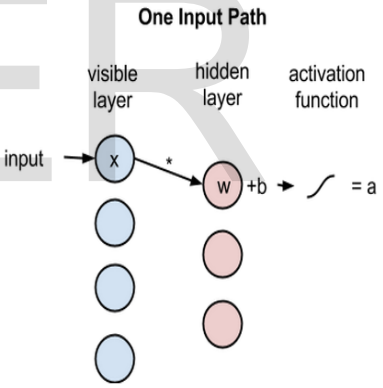

FIGURE 2. Basic working of RBM(Single input path)

In multiple input format ,each hidden node receives the given number of inputs multiplied by their respective weights(w) .The sum of these weights are added to the bias(b) ,the result passes through the activation algorithm ,giving rights to one output for each hidden node .[5]

IJSER

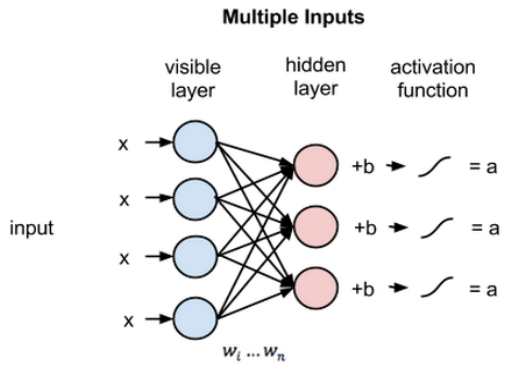

FIGURE 3. Multiple input format in RBM

MATERIALS AND METHODS

A. RESTRICTED BOLTZMANN MACHINE(RBM) IN FACIAL RECOGNITION

The RBM constitutes of a layer of visible units and a layer of hidden units with no communication between the hidden nodes .It is much easier and efficient than general Boltzmann machines as there is no need to perform any iterations to determine the activities taking place in the hidden units .[6]

Here ,we consider s_j as hidden states and s_i as the visible states .The distribution of the hidden states (s_j) can be explained by standard logistic form function:

$$p(s_j = 1) = \frac{1}{1 + exp\,(-\sum w_{ij}\, s_i)} \qquad (1)$$

The approximate gradient of the contrastive divergence ,where Q^1 is distribution of one step reconstruction of data produced by first pick of binary hidden states from their respective conditional distribution provided the data and their picking binary visible states from conditional distribution are given in the hidden state ,is :

$$\Delta w_{ij} \propto <s_i s_j>_Q^0 \; - <s_i s_j>_Q^1 \qquad (2)$$

We can increase the representational power without changing learning procedures and inference by imagining each visible unit , i ,has 10 replicas and ways same as the hidden units .Only active replicas are counted .During reconstruction from the hidden activities ,each replica shares the computational probability ,p_i ,of turning on .We then select n replicas to get a probability of $\binom{10}{n} n^{p_i} (10 - n)^{1-p_i}$.[6]

For simplicity ,we assume neurons have identical weights .It has been seen that when a single neuron is stimulated over an interval of time ,it may produce multiple spikes that may constitute of a rate code .We also assume that visible units can produce atmost 10 spikes and hidden units can produce a maximum of 100 spikes .To use RBMrate ,we train a single model on pairs of different images of the same individual and then we decide using this model of pairs ,which gallery image matches the best with the test image .We define the fit between two faces f_1 and $f_2$as$G(f_1 , f_2) + G(f_1 , f_2) - G(f_1 , f_1) - G(f_2 , f_2)$where$G(v_1 , v_2)$is the goodness score of the image pair v_1 , v_2under the model .The goodness score is a negative free energy which is the result of additive function on total input received from each hidden unit .However , to preserve the symmetry , image pairs of same individual $v_1$and $v_2$from the training set are set to reversed pair $v_2$and v_1.We trained the model with 100 hidden units for 1000 image pairs(500 distinct pairs) , 2000 iterations in set of 100 and learning rate of 2.5 x 10^{-6} for the rates , learning rate for biases being 5 x 10^{-6} and momentum being 0.95 .[6]

B. FACIAL RECOGNITION TECHNOLOGY(FERET)DATABASE

The FERET database is used for facial recognition system and serves as standard database for facial images and is used in developing algorithms and results of the reports .[7]

Here we consider 1002 frontal face images of 429 individuals collected over a period of time in variable optical conditions as the input of the FERET database .Out of these ,818 are used in both gallery and training set and the remaining 184 divided over 4 disjoint tests sets :

Δ**expression** test consists of 110 images of different individuals .Different images of the same individuals are also present in the training set taken at same time under same optical conditions but have different expressions .

Δ**days** test consists of 40 images from 20 individuals All these individuals have 2 more images from the same session present in the training set .

Δ**months** test is similar to Δ**days** test except the fact that the time between the sessions was 3

IJSER

months or more and was under varying optical conditions .This contains 20 images of 10 individuals .

Δ**glasses** test consists of 14 images from 7 individuals .All these individuals have 2 images in the training set taken at some time interval on the same day .The difference between the training and the test pairs is that one has glasses and the other does not .[6]

FIGURE 4. Examples of sample faces

FIGURE 5. Examples of processed faces

C. COMPERATIVE RESULTS

Comparison of RBMrate has been compared with the simplest method known as **correlation** . It returns the simplicity score which is the angle between two images represented by vectors of pixel intensities .The performance can be further increased by using additional layers of hidden units leading to better feature detection activities .[6]

D. RESTRICTED BOLTZMANN MACHINE(RBM) IN FACIAL EMOTION DETECTION

The RBM is a double layered undirected graphical model which has no lateral connections .The first layer ,the visible layer ,v ,and the second being hidden layer ,h .Each of its nodes are stochastic binary units and each configuration of **v** node and **h** node is characterized by an energy given by the following function :

$$E(v, h) = -\sum_{i,j} v_i w_{i,j} h_j - \sum_j b_j h_j - \sum_i c_i v_i \quad (3)$$

This can be further interpreted probabilistically as follows:[4]

$$P(v,h) = \frac{exp(-E(v,h))}{Z} \quad (4)$$

Considering the visible units as real , energy function can be defined as follows :[4]

$$E(v, h) = \frac{1}{2}\sum_i v_i^2 - \sum_{i,j} v_i w_{i,j} h_j - \sum_j b_j h_j - \sum_i c_i v_i \quad (5)$$

The **h** nodes are conditionally independent provided the visible layer and vice versa .The conditional probabilities are as follows :[4]

$$P(v|h;\ \theta) = \Pi ip(v_i\ |h)\ ,\ P(h|v;\ \theta) = \Pi_j p(h_j|v) \quad (6)$$

$$p(h_j = 1|v) = sigmoid(P_i W_{ij} + b_j)\ ,$$ For binary visible layer, (7)

$$p(v_j = 1|h) = sigmoid(P_i W_{ij} h_i + c_j)\ p(h_j = 1|v) = sigmoid(P_i W_{ij} v_i + b_j) \quad (8)$$

For real valued visible layer, we have,

$$p(v_j = 1|h) = N(P_i W_{ij} h_i + c_j, sigma) \quad (9)$$

The learning of RBM parameters can be done efficiently by maximizing the log-likelihood for the training data using gradient ascent .The exact gradient is intractable ,hence we use contrastive divergence which works fairly well .The exact gradient can be approximated by [4]

exact gradient : $\frac{\partial log p(v)}{\partial Wij} = < v_i h_j >^0 -$

$$< v_i h_j >^\infty \quad (10)$$

CD approximation :$\frac{\partial log p(v)}{\partial w_{ij}} = < v_i h_j >^0 -$

$< v_i h_j >^n$ *(11)*

D. MMI FACIAL EXPRESSION DATABASE

The MMI facial expression database aims to store large volumes of facial expressions in the form of visual data .It addresses a number of key omissions in other databases storing facial expressions .It also contains the recordings of full temporal patterns of facial expressions ,from neutral through a series of onset ,apex and offset phases again coming back to neutral face .It works better than other databases as it not only focuses on the basic 6 emotions but also contains prototypical expressions and expressions with single FACS Action Unit (AU)activated .[8]

FIGURE 6. MMI database containing basic human expressions .

E. ARCHITECTURE DIAGRAM AND MODULES

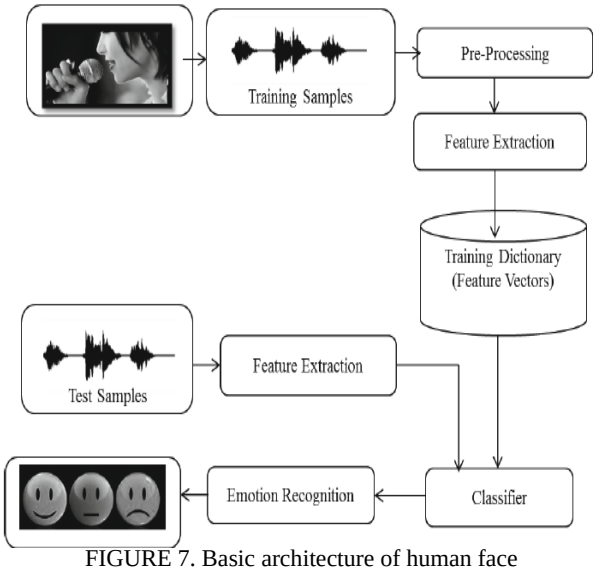

FIGURE 7. Basic architecture of human face recognition and emotion detection

The process starts with accepting inputs from the user which is kept in training samples .The pre-processing deals with rotation of the input image ,if tilted ,followed by cropping of some portion of the face along with the entire background .Feature extraction deals with converting the picture into an oval shape which contains only those features which are unique to a person .The hence received data is sent to training dictionary for feature verification where face recognition takes place and partial emotion detection .The result of face recognition is directed to classifier .Now ,the partial result of emotion detection is compared with the test samples and features are extracted .The final result is then sent to the classifier .The combined final result is ultimately sent for emotion detection .Ultimately ,if the result of face recognition matches with any of the pre-fed samples in database ,the access is granted followed by interpreted result of user emotion is declared .

FUTURE PREDICTIONS OF EMOTIONAL INTELLIGENCE AND CONCLUSION

The Artificial Emotional intelligence will rule the next 20 years of the human history .It plays a significant role in the normal lay man's life also .The integration of the artificial intelligence with the emotional intelligence has been of a great purpose in the aviation, gaming and biomedical fields.

Presently, the chat bots and mobile applications have been able to effectively respond to the people with the various emotions like happiness, frustration ,sadness ,anger and irritation thus, helping to improve the customer care .It is in the part of the phone with the Animoji and other security purpose things in the cooperate fields . The future predictions of the Emotional Intelligence(EI) is the idea of multimodal emotion recognition and an emotion chip .Big data also gives an edge to emotional intelligence .Artificial intelligence uses

IJSER

humans as organic robot ,to make that happen Human API needs to be created.

The EI will most influenced in the hospitals without the approach to hospitals the patients can be able to consult doctors and even it will be more easy for the doctors to detect the disease and to build trust in patients. Using USC the virtual therapist can be able to develop so that the emotional triggers in the patients who are suffering from (PTSD).Even the road maps will be using the emotional intelligence for the better servicing to the customers .Even in the self driving cars also the EI bots is used to detect the emotions.

Already Tabatha which can able to recognise the emotions .Even Eion , Musk ,Google ,Facebook and Microsoft have been continuously investing billions of billions in the EI field for the further development of EI.

IJSER

REFERENCES

1. http://www.upwork.com/hiring/for-clients/pros-cons-facial-recognition-technology-buisiness/

2. Fusion of visual and thermal face recognition techniques a comparision study(author-Jingu Hea october,2003).

3. Facial emotion detection in real time-Dan Duncan,Gautam Shine,Chris English.

4. Facial emotion recognition using deep learning. -Ankit Awasthi

5. http://Skymind.ai/wiki/restricyed-boltzmann-machine

6. Rate-coded Restricted Boltzmann Machines for Face recognition.

7. http://en.wikipedia.org/wiki/FERET_database

8. http://mmifacedb.eu