SURVEY PAPER ON METHODS OF EMOTION DETECTION FROM STILL IMAGES

[1]Shrutika C. Dhargave ,[2]Apurva Sonak ,[3]Vandana Jagtap
[1,2]M.E. Student, CE Dept, [3]Assistant Professor, CE Dept.
MAEER'S MIT, Pune
Email:[1]shrutika.dhargawe@gmail.com, [2]apurvasonak26@gmail.com
[3]vandana.jagtap@mitpune.edu.in

Abstract: **Reading human mind has always remained a mystery for science and so is the case with human emotions. Many methods of emotion recognition such as emotion detection from text or from human face have been discovered. Here, we present a survey on the novel approach of emotion recognition which is emotion detection from still images. The techniques which are used to do so are also discussed in brief.**

Keywords: Discrete Cosine Transform, Pattern Recognition Network, Expression recognition, 3-D muscle mesh, feature extraction, motion estimation, biometrics, adaptive median filter, bacteria foraging optimization, feature detection, facial expression

I. Introduction

Detecting emotion has been an increasingly popular research topic in recent year. Recent research is done on emotion detection from text.Their applications range from advertisement and commercial purposes to medical patient behavior analysis. Applying this in a social network setting, it can be a powerful tool to gain knowledge about how individuals, social circles, communities, or cities feel about current events or other such topics.

Emotion Recognition with the help of images is a growing research field. Emotion from images is to detect changes in facial expressions inaccording to an individual's internal emotion state and intentions.Human face is a significant place for detecting emotions, six emotions detected from human face. They are Happy, Surprise, Anger, Sad, Fear and Neutral.

This paper has covered some techniques which are used for emotion recognition from images. The techniques are discussed below in brief.

II. Emotion Detection using Pattern Recognition Network

The paper presented by Shivani and Shashwat at IJERA [1] uses the technique of pattern recognition network. The recognition of emotions is done by deploying an intelligent system using neural network, signal processing and image processing toolbox of Matlab. The network classifier is pattern recognition network which is actually a feed forward neural network that will be trained for some images bearing different emotions. The trained network is then simulated to test the new data for recognizing different emotions.

The technique involves following steps:

1. *Acquiring Images:*
2. Images are taken for sample data. They can be saved images or can be captured with the help of camera.
3. *Processing Images:*
4. The acquired image is resized and converted to gray scale for further processing.Gray scale images are very common and sufficient for various tasks.

5. *Feature Area Extraction:*
   From gray scale images eyes and mouth portions are extracted since these areas carry the more essential emotion information.
6. *Feature Extraction:*
   Feature extraction is done using 2D Discrete Cosine Transform (2D DCT) which changes the image data from the spatial to the frequency domain.The feature vector, which is created using the DCT matrix, is used to train the Pattern Recognition Network.
7. *Training Classifier:*
   Training classifier is Feed-forward neural network which uses back-propagation algorithm to train the classifier for input data against given target data i.e. all six emotions.
8. *Simulation:*
   The trained classifier is then simulated to test new real world data and identify all six basics emotions i.e. Happy, Sad, Anger, Fear, Surprise and Neutral.

More real life applications can be explored like driver monitoring and studying human psychology. In the proposed system the feature areas are extracted manually, automatic feature area extraction can be done. Mix Emotions like Happy and Surprise, Sad and Fear from an image can also be detected.

III. Emotion Recognition using 3-D model

Kostas compares muscle contraction and expansion from the human face to relevant data taken from a 3-d model of a head [2]. This takes place at curve level, with each curve being drawn from detected feature points in an image sequence or from selected vertices of the polygonal model. The result of this process is identification of the muscles that contribute to the detected motion; this conclusion is then used in conjunction with neural networks that map groups of muscles to emotions. The notion of describing motion with specific points is also supported in MPEG-4 and the relevant encoded data may easily be used in the same context.

It observes the temporal movement of facial keypoints, which reveals the type of emotion taking place in a video sequence. The comparison is based on synthetic 3-D generated prototypes for the corresponding points.

Ekman conceived the notion of an action unit, which is in essence the recognizable result of the flexing of a single or a small group of facial muscles. The muscles that are related to expressions are superficial, i.e. near the surface of the head, and are responsible of the general shape of the head. Most of these muscles work collectively; as a result, it is difficult to separate the margins between the areas of influence of distinct muscles.

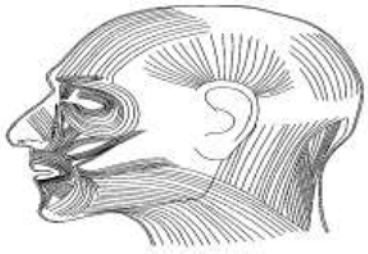

Fig 1. Structure and position of facial muscles

The grouping of areas of vertices are done with respect to the facial feature to which they correspond and the muscles responsible for their transformation. This mapping is a result of anatomy surveys.The outcome of the modeling process is a library of possible muscle actions. Some of these emotions are termed universal.
This system utilizes automatic feature extraction and motion estimation techniques, along with 3-d face models to compare motion data to pre-defined prototypes, which results in muscle activation information which is mapped to groups of emotion.

IV. Emotion Recognition using Matlab

The study done by Neha [3] presents a computationally efficient approach for edge detection which further leads to classification of facial expression recognition from static facial images. The various algorithm are used to produce the characteristic features such as lips and eyes. Firstly the images will be loaded in the train folder and test folder. After this these images are analyzed by series of algorithms and techniques to enhance the image input, maintain intensity and removing noise from image. Second algorithm detects the edges of image.
From the edge points various distances between features is

calculated and principal component analysis is used for data reduction and next algorithm detects the face.

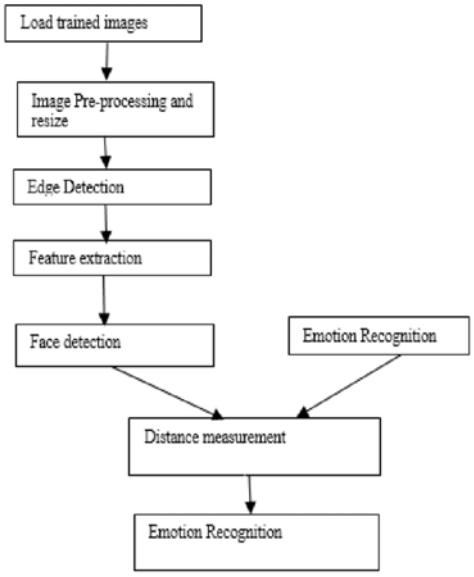

Fig 2.Functional block diagram of the facial recognition system.

They proposed an accurate and high speed emotion detection system. The color and featurebased detections were adopted to find skin-color fast and selected candidate blocks carefully. We used lighting compensation to improve the performance of color-based scheme, and reduce the computation of feature-based scheme. The major contribution of this paper is that the proposed method can detect edges of the images and from that edges distance between various features is calculated by using Euclidean distance Formulae. This distance is different for every image posing different emotions. On the basis of this distance emotions are classified.

The proposed approach can be applied to hardware implementation. Due to the proposed method has simple structure, it is suitable to be implemented in hardware to achieve very high performance and low power system.

V. Emotion Recognition through Facial Expression using Desktop Devices

Considering the importance and the escalating need for advanced Human Computer Interaction, the paper revolves around generating the mood or emotion of humans based on their facial expression. Furthermore, it explains the necessity of Real-Time Systems [4] in order to achieve high levels of interactions with the machines. It is of prime importance that the interaction of humans with the computers should be latency free, thus taking it to the level of face-to-face communication.

The development of a robust real-time perceptive system, which will take into account the facial expressions, detect human face and code the expression dynamics.

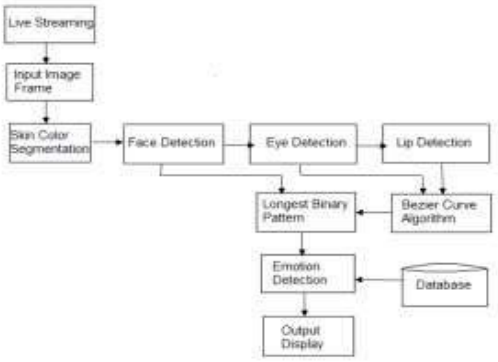

Fig 3.Steps for emotion recognition using Desktop Devices

Their paper defines a system that automatically detects human emotions on the basis of facial expressions. The techniques followed are more suited thus limiting the latency in response. Moreover, it further takes a step to improvise the emotion detection technique using Cubic Bezier Curve Implementation which is more adaptive and resurface as the ones with utmost importance in various other fields like that of robotics, computer graphics, automation and animation. The system works well for faces with different shapes, complexions as well as skin tones and senses basic six emotional expressions. The system even handles face rotations across x-axis and fetches accurate results for horizontal rotations.

However this flexibility is not extended over vertical rotations of image. Although it is unable to discover compound or mixed emotions.The system accurately detects the emotions for a single face, however this correctness reduces prominently for multiple faces. The system can further be upgraded to include multiple face detections. This would indeed form a big step in areas of Human Computer Interaction, Image Processing, Animation, Automation and Robotics.

VI. Emotion Detection using Patches

Face is detected and then physically crop from record images as a resolution of 256*256 pixels [5]. Then, an entire set of patch are extract from the imagery. Then a patch harmonizing process is takes place. Here change the extract patches into distance facial appearance. Todetain the facial group of patches.Where the least amount rule is use to discover the accurate matching features in space. Then a group of "main" patch are preferred .After giving a test image similar patch harmonizing process is performed using "most important" patch. The consequential distance facial appearances are given to the random forest classifier to recognized six essential emotions

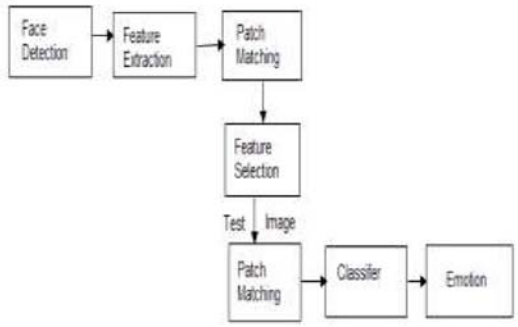

Fig 4.Steps for emotion detection using patches

1. *Face detection and feature extraction:*

A face detected from the whole image is then manually cropped from the image. A set of different patches with different size is extracted from the image. The patch set is collected by scanning all rows and columns inthe face region. PCA algorithm is used for feature extraction process.

2. *Patch matching:*

Patch matching is used to find the majority similar features located within a space for action categorization and object identification. The matching helps to reduce the error. Patch matching performed to change the extract patch into reserve facial appearance. This distance features are finding by harmonizing thesepatches among every patches with its corresponding training image. Here taking two as input values and select one minimum value as final. The least value is selected for finding emotions *3. Feature selection and patch matching:*

A group of most important patch only chosen for reduces the computation time. After giving a test image,equal patch harmonizing is perform on the salient patches in testing images, resulting in distance features arefed into the random forest classifier.

4. Random forest classifier:

Random forest has many categorization trees. To categorize a new thing from an input, give the input underevery tree in the forest. Each tree provides a classification, and says the tree "votes" for that class. The randomforest had chosen the most votes for categorization.

VII. Emotion detection using Genetic Algorithm

Ekman and Friesen developed the most comprehensive system for synthesizing facial expression based on what they called as action units [6]. Double structured neural network has been applied in the methods of face detection and emotion extraction. In this, two methods are proposed and carried out; they are lip detection neural network and skin distinction neural network. Facial action coding is given to every facial point. For example, code 23 is given for lip funnel, code 4 for eye brow lower, code 10 for chin raise etc. The cods are grouped for a specific facial emotion. In order to determine the category of emotion, a set of 15 facial points in a face-profile sequence has been recommended.

The human lip shape is a combination of two ellipses, which can be called as an irregular ellipse. The word 'irregular' is used as the ellipse has two different minor axes whereas the major axes remains the same. Thus the edge detected lip mage is considered as an irregular ellipse. Lengths of minor axes of the lip feature for each emotion are computed.

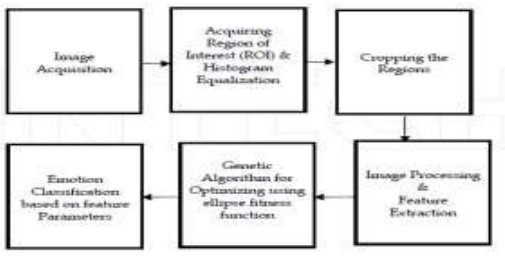

Fig5. Process Flow of Image Processing

Fig6. Edge detected and whitened lip image

Figure 6 shows the whitened area of edge detected lip image for a particular emotion. The major axis is "2a" (considered as fixed) and two minor axes are "2b1" and "2b2" (to be computed). This is shown in Figure 7 with b1 and b2 suitably substituted for top and bottom portions respectively.

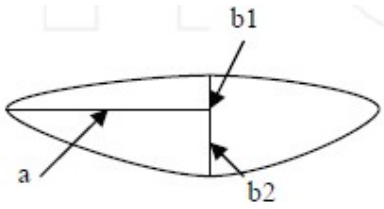

Fig7. The Irregular Ellipse

Algorithm:

1. Represent the problem variable domain as chromosome of a fixed length and population, with suitable cross over probability and mutation probability
2. Define a fitness function to measure the performance, or fitness of an individual chromosome in the problem domain
3. Randomly generate an initial population of chromosomes.
4. Calculate the fitness of each individual chromosome.
5. Select a pair of chromosomes for matting from the current population. Parent chromosomes are selected with a probability related to their fitness. Highly fit chromosomes have a higher probability of being selected for mating compared to less fit chromosomes.
6. Create a pair of offspring chromosomes by applying the genetic operators – crossover and mutation
7. Place the created offspring chromosomes in the new population
8. Repeat from step 5 until the size of new chromosome population becomes equal to the size of the initial population
9. Replace the initial chromosome population with the new population
10. Go to step 4, and repeat the process until the termination criterion is satisfied.

A feed forward neural network is used to classify the emotions which is based on optimized ranges of 3-D data of top lip, bottom lip and eye [7]. The optimized values of the 3-D data are given as inputs to the network as shown in Figure 8.

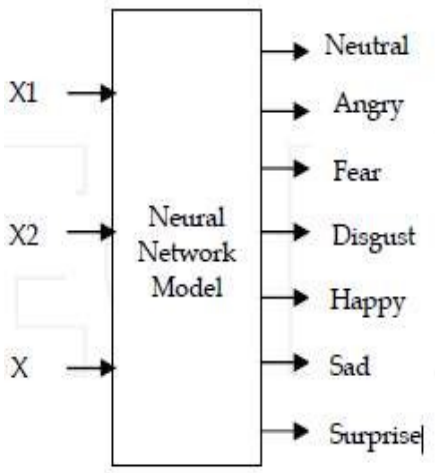

Fig8. Emotion recognition using neural network

After getting the inputs, the neural network model recognizes the emotion and gives the name of the emotion as the output.

The software package can be developed as an expert emotion classification system for a personalized face. The applications of this emotion classification system are many such as from identifying criminals through a police enquiry to helping bedridden disabled dumb patients.

VIII. Emotion Detection using Gabor Filter

Shubhangi [8] discusses the application of Gabor filter based feature extraction in combination with neural network for the recognition of five different facial emotions. Facial expressions provide rich information about human emotion and play an important role in human communication. The recognition system considers Eyebrows, Nostrils, Lip Line, Wrinkles, Jaw Line and Eyelids to detect the emotions. The neighboring region of a pixel is described by the response of a group of Gabor filters in different frequencies and directions, which have a reference to the specific pixel. The

Gabor Filtering method generated 24 features values accompanied by variance for each feature. The Face Emotion Recognition (FER) is achieved in two parts; Image processing and classification.

The first part, that is the image processing stage, investigates a set of image processing methods suitable for recognizing the face emotion. The acquired images then undergo few preprocessing methods including Gaussian smoothing, conversion to YCbCr and grey scale conversion.

The second part is of classification, which using all the grey scale images classifies and recognizes the emotions.

IX. Emotion Detection using Deep Belief Network

With the help of recent discoveries, it is possible to effectively train artificial neural networks with multiple hidden layer units which allow hierarchical, incrementally abstracted representation. The performance of a Deep Belief Network is a powerful deep learning architecture for the task of classifying the face emotion [9]. The DBN architecture learns representationally useful and human interpretable features of the input autonomously even with unlabeled data.

The general idea which is loosely inspired by the computational properties of biological neurons involves a layered, forward-connected network of nodes composed of an input and one or more hidden layers and output layer. The nodes at the input layer correspond to the low level information from which the network is trying to learn from the information. The output layer represents the learning ability such as the category of an image or proper motor output. The hidden units are not actually observed components of the learning problem and so are less easily interpreted. But before sending the input they allow the network to re-represent it to the output. They provide representation at a level of abstraction in between input and output and thus greatly increase the expressivity of Artificial Neural Networks.

The breadth and depth of the network should increase with the complexity of the learning task and optimizing the size parameters is often a process of trial and error. However, a single sufficiently large hidden layer is theoretically capable of learning any problem that can be learned by an arbitrarily deep network. Moreover, researchers have shown that traditional ANN learning algorithms like backpropagation are not effective for training deeper networks.

In particular, Hoch Reiter described the vanishing gradient problem, which analyzed how error gradients decay exponentially so lower-level layers receive little to no information on how to change weights once the network becomes reasonable good at a task; in effect, the lower levels are dispensable in deep architectures trained with back-prop. Because of these properties, most ANN approaches have favored shallow networks.

A Deep Belief Networks (DBN) is a common way to implement a deep neural net. DBNs incrementally creates layer-by-layer the network topology which starts with the input then hidden layers. Each layer is trained with a Restricted Boltzmann Machine (RBM), a stochastic neural network which is composed of a visible layer and a hidden layer as illustrated in Figure 9. A RBM learns the probability distribution of a set of observations over the hidden units. Visible units correspond to the attributes of an observation and hidden units model the dependencies between different aspects of the observed data.

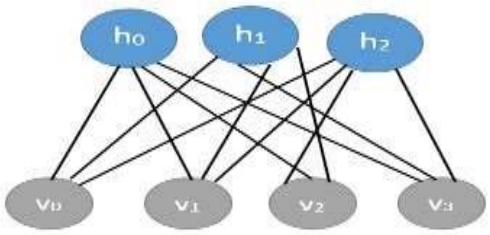

Fig9. An illustration of the layers and connections of an RBM

DBNs are formed by stacking a series RBMs, starting with the input layer and first hidden layer. Figure 10 shows an example of the DBN architecture. Once the first RBM is trained, its hidden layer becomes the visible layer of a second RBM and we add a new layer as the hidden layer of the new RBM.

This entirely unsupervised pertaining can be continued indefinitely up to the final hidden layer. After this process is complete, connecting the final hidden layer to an output layer results in a deep neural network that can be trained with gradient descent on the error from the target

output. In this way, a DBN iteratively learns increasingly abstract, hierarchical representations by composing the lowerorder features of preceding layers.

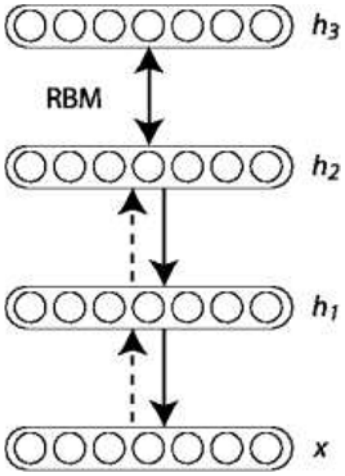

Fig10. A simple graphical illustration of a DBN with 3 hidden layers

Here, it explore the design and performance of DBNs through a series of experiments in facial emotion image classification. We believe that the levels of abstraction involved in mapping a matrix of pixel values to an emotion category make it particularly well suited for studying deep architectures. Moreover, given the compositional nature of faces (where individual features like eyes, lips, and forehead are combined in a more holistic structure), we expected that the learned intermediate representations would be more qualitatively interpretable.

We expect to see our Deep Belief Network develop low-level feature representations of the face with and combine these into higher-order representations to compose entire faces. Once these features have been learned in unsupervised training, we expect that the model will efficiently learn the particular classification task. Moreover, we anticipate that deeper architectures will outperform shallow networks, since our task was chosen so promote hierarchical representations. We have less refined predictions with regards to the influence of the size of hidden layers, but generally anticipate that forcing information compression by reducing the number of units in the hidden layer will result in better abstractions.

X. Emotion Detection using Hybrid Technique

Renu presents novel approach for the detection of emotions using the cascading of Mutation Bacteria Foraging optimization and Adaptive Median Filter in highly corrupted noisy environment [10]. The approach involves removal of noise from the image by the combination of MBFO & AMF and then detects local, global and statistical feature form the image. The Bacterial Foraging Optimization Algorithm (BFOA), as it is called now, is currently gaining popularity in the community of researchers, for its effectiveness in solving certain difficult real world optimization problems.

At first we investigate the emotionally intelligent computers which can perceive human emotions. In this research paper four emotions namely anger, fear, happiness along with neutral is tested from database in noisy environment of salt and pepper. Very high recognition rate has been achieved for all emotions along with neutral on the training dataset as well as user defined dataset. The proposed method uses cascading of MBFO & AMF for the removal of noise and Neural Networks by which emotions are classified.

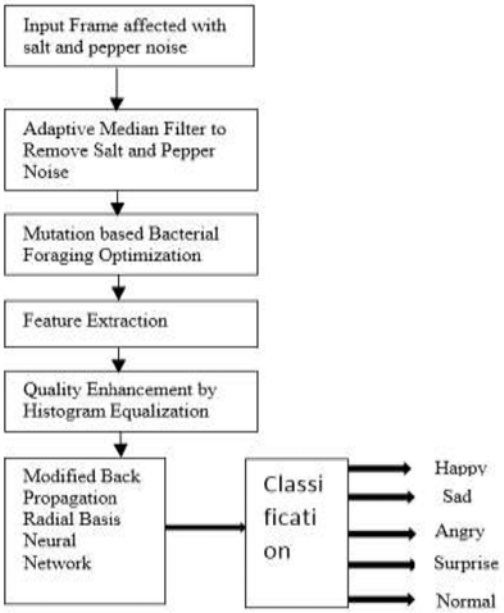

Fig11. Block Diagram of Hybrid Technique

1. *Input Noisy Image:*
   The input image is picked randomly from the database which is used for training and evaluated for the recognition accuracy and sampling.
2. *Adaptive Median Filter:*
   Adaptive median filter is very good at preserving image detail. So to run a median filter the steps are as follows:

a. First consider each pixel in the image

b. Based on the intensities, sort the neighboring pixels into order.

c. Then replace the original value of the pixel with the median value from the list calculated. Following is the algorithm of Adaptive Median Filter:

(i). Initialize w = 3 and Zm = 39.
(ii). Then compute Z min, Z max and Z med (iii). If Zmin < Zmed < Zmax, then go to step (v).
Otherwise w = w+2.
(iv). If w ≤ Zm, go to step (ii).
Otherwise, replace Zxy by Z med (v).
If Zmin< Zxy< Zmax, then Zxy is not noisy else replace Zxy by Zmed.

3. *Mutation Bacteria Foraging Optimization:*
   During the first stage the input noisy image which has a salt and pepper noise of varied densities from 0.05 to 0.9 is applied to the adaptive median filter. In the second stage, both the noisy and adaptive median filter output images are passed to next step as search space variables in the Bacteria Foraging Optimization (BFO) technique to minimize errors which occurs due to differences in filtered image and noisy image.
4. *Pre-Processing and Feature Extraction:*
   In this phase, the face image passed is transformed to operational compatible format where the image is resized to uniform dimension, the data type of the image sample is then transformed to double precision and passed for feature extraction to extract the features for recognition. In the first step of image processing, the region of interest that is of a lip and an eye or only lips region or only eyes region have been selected independently from the transformed images with the help of the mouse. The ROI image is then converted into grayscale image for recognition.
5. *Histogram Equalization:*
   Histogram equalization improves the contrast in the grayscale and its main goal is to obtain a uniform histogram from the image. This method also helps to reorganize the intensity distributions of the images. New intensities are not introduced into the image but are mapped with the existing intensity values then to the actual number of intensity pixels in the resulting image will be equal or less than the original intensity value. In the new image sequence, the histogram equalized image is filtered using filters in order to make the image smoother.
6. *Modified Back Propagation and Radial Basis Neural Network:*
   In this phase epochs and errors are calculated of particular face region. Two parameters are used and the total numbers of epochs and errors are calculated in neural network. Based on the input parameters extracted two types of neural networks were trained [11]. They are:
   (i) Back Propagation Neural Network
   (ii) Radial Basis Neural Network
7. *Classification:*
   A representative set of the facial features are determined from the inputs into this stage. After this training, the new images that is the facial features are processed and entered for identification into the recognition stage. Then in the recognition stage the emotions are detected as angry, happy, sad, surprise and normal.

With fixed step size the BFO requires more computation time with less accuracy. The speed of BFO is increased by enhancing the accuracy in terms of quality of images due to the mutation. This technique can be used as robust face emotion detection algorithm.

XI. Conclusion

We studied emotion recognition by using various techniques. The paper gives an overview of the few techniques available for emotion detection and recognition. It will help to understand these techniques in short and to choose from them which can be well suited for future evaluation depending on the individual's project.

Though there are various techniques available but the key concept remains the same. We have to first select that portion of the images which helps in detecting the emotions. These feature extraction can be performed using various techniques such as 3-D model techniques, Matlab, patch based feature extraction. After extracting the features it should be analyzed with the database to recognize which type of emotion it is. These is done using neural networks, deep belief networks, pattern recognition method, etc. and the final output is produced giving the emotions from the images

References:

[1]Shivani Gangwar, (2013, Sept.). On Human Emotion Recognition by Using Pattern Recognition Network at Int. Journal of Engineering Research and Applications. Vol. 3, Issue 5, pp.535-539

[2]Kostas Karpouzis**,** George Votsis, George Moschovitis and Stefanos Kollias on Emotion Recognition Using Feature Extraction and 3-D Models.

Available: http://www.image.ntua.gr

[3]Neha Gupta and Prof. Navneet Kaur (2013, Jul-Aug) On Design and Implementation of Emotion Recognition System by Using Matlab at International Journal of Engineering Research and Applications Vol. 3, Issue 4, pp.2002-2006

[4]P. M. Chavan, Manan, Jinal, Aditi, Pooja (2013, May) On Real Time Emotion Recognition through Facial Expressions for Desktop Devices at International Journal of Emerging Science and Engineering, Vol. 1, Issue 7 [5]Ramya S on Emotion Recognition Using Facial Images at International Journal of Engineering and Technology (IJET)

[5]M. Karthigayan, M. Rizon, R. Nagarajan and Sazali Yaacob, "Genetic Algorithm and Neural Network for Face Emotion Recognition" Universiti Malaysia Perlis (UNIMAP).

[6]J.-J. Wong and S.-Y. Cho, "A Face Emotion Tree Structure Representation with Probabilistic Recursive Neural Network Modeling," Neural Computing and
Applications, vol. 19, pp. 33-54, 2010.

[7]Shubhangi Giripunje, Preeti Bajaj and Ajith Abraham, "Emotion Recognition System Using Connectionist Models" at Machine Intelligence Research Labs, India Center, GHRCE, Nagpur, India.

[8]Kevin Terusaki and Vince Stigliani (2014, May), "Emotion Detection using Deep Belief Networks"

[9]Renu Nagpal, Pooja Nagpal and Summet Kaur (2010, Dec.), "Hybrid Technique for Human Face Emotion Detection" at International Journal of Advanced Computer Science and Applications Vol.1 No. 6

[10]Ruba Soundar Kathavarayan and Murugesan Karuppasamy, (2010) "Preserving Global and Local Features for Robust Face Recognition under Various Noisy Environments", International Journal of Image
Processing (IJIP) Volume (3), Issue (6)