International Journal of Computer Applications (0975 – 8887)
Volume 158 – No 6, January 2017

Human Affect Recognition System based on Survey of Recent Approaches

Shweta Malwatkar
Department of Computer Engineering
MITCOE, Kothrud, Pune

Rekha Sugandhi
Department of Computer Engineering
MITCOE, Kothrud, Pune

Anjali R. Mahajan, PhD
Department of computer and IT
Govt. Polytechnic, Nagpur

ABSTRACT

In recent years, the analysis of human affective behavior has been a point of attraction for many researchers. Such automatic analysis is useful in various fields such as psychology, computer science, linguistics, neuroscience etc. Such affective computing is responsible for developing standard systems and devices, useful for recognition and interpretation of various human faces and gestures. The emotions are categories as anger, disgust, fear, happiness, sadness and surprise. Such emotion recognition system involves three main steps: face detection, feature extraction and facial expression classification. Hence, there is a need for standard approaches that solve the problem of machines understanding the human affect behavior. This survey paper presents some recent approaches that recognize the human affective behavior, with their advantages and limitations. This paper also presents some basic classifiers such as SVM, ANN, KNN and HMM, used for emotion classification and audiovisual databases with their emotion categories. Based on the survey, an affect recognition system has been proposed that adopts a cognitive semi-supervised approach.

Keywords

Affective computing, facial emotions, classification, image processing, machine learning.

1. INTRODUCTION

Emotion is an important factor in social interaction. It is also useful in human intelligence and interpretation, etc. As in social environment, interpretation and experience of emotion are important for communication, it is essential to understand the emotion in human's daily life. In recent, various computer-based technologies have emerged to understand the emotions. This research comes under the affective computing. In the area of affective computing, various systems and devices are being developed, which are useful for emotion recognition and their interpretation. Efforts in the development of communication systems have resulted in great progress in applications. But the research work in implicit communication i.e. emotion based communication is underappreciated. Recently, the research in affective computing attracts researchers because of its use in various applications such as psychiatric diagnosis and screening, customer relationship management, children's behavior analysis, etc. Emotion classification is a very challenging problem because emotions are constructs with fuzzy boundaries in labels. The emotion is expressed through various aspects like speech and non-speech vocalizations, gestures, facial expressions, physiological signals, and many others. Such technologies are useful in various computer supported communicational application devices such as robots, hand-held devices, mobile phones etc. In various human-computer interaction (HCI) systems, emotion recognition system plays very important role for providing easy interaction between computers and humans. Therefore, it becomes a need to develop the good quality of techniques that easily identify and classify the emotion of humans.

Mostly, Facial Expression Recognition basically performed in three major steps [2]:

- Face Detection
- Feature Extraction
- Facial Expression Classification

In this procedure of emotion detection, initially, various devices are required to capture the physical behavior of the user. For example, the video camera is used to capture the facial expressions, posture of body and gestures, and to capture the speech, microphones are needed. Some sensors can identify the cues by measuring the data like skin temperature and conductance. Then from this gathered data, various meaningful patterns are extracted to classify the emotions. This is done by various machines learning techniques such Speech Recognition, Natural Language Processing, facial expression detection, and classifications of various face emotions.

1.1 Human Emotions

Here some kinds of emotions are described in figure 1. In this figure emotions are categorized into various categories such as:

- High and Low Arousal Positive Affects
- High and Low Arousal Negative Affects
- High and Low Arousal Neutral Affects
- Moderate Arousal Positive and Negative Affects

Figure describes total 16 emotions of human including alert, excited, elated, happy, contented, serene, relaxed, calm, fatigued, lethargic, depressed, sad, upset, stressed, nervous, and tense.

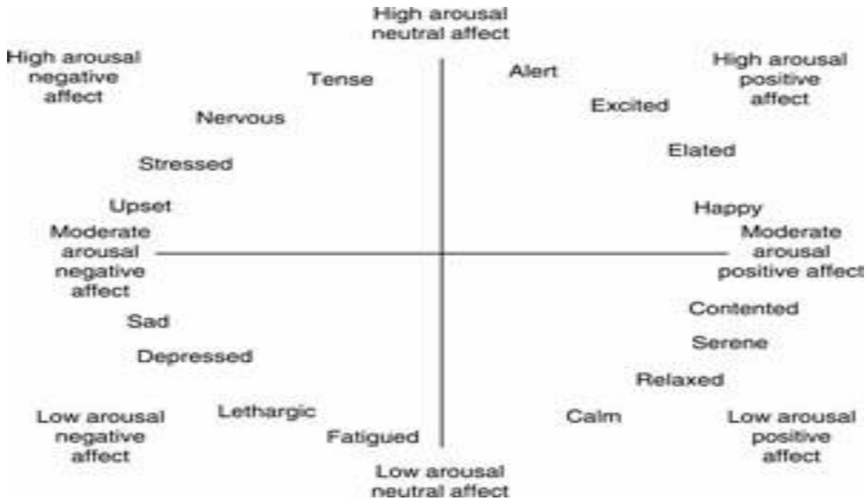

Fig 1: Bi-axial plotting of Human Emotions based on degree of arousal

2. LITERATURE SURVEY

2.1 Affect Recognition System

R. Suresh et.al proposed the automatic facial expression recognition system [4]. This system was the combination of two main steps: feature extraction with Contourlet Transform (CT) and classification with K-Nearest Neighbor (KNN) classifier. For feature extraction phase, expressive images were considered as an input. These images were taken from JAFFE database. These images were then decomposed up to four levels using contourlet transformation. The contourlet transform scheme is consist of two iterated filters named as Laplacian Pyramid (LP) and Directional Filter Bank (DFB). LP can detect point discontinuities of an image and these point discontinuities are linked into a linear structure by using DFB. The main purpose of this procedure is to identify the edges of images for segment detection. In this work, various combinations of filters were used for feature extraction. Combination one contains 9/7 and Ladder filter and Combination 2 contain Maxflat and regular linear phase bi-orthogonal filter. These extracted features were then used to train the classifier. In classification phase, KNN classifier is used to recognize the patterns from given input test image. That is test image is classified based on closet feature space. To achieve high facial expression recognition rate, correlation measure is used as a distance metric. A system with combination 1 filter type achieves maximum classification Accuracy.

Hao Tang et.al proposed the gender and person independent facial expression recognition system [5]. This system executed in two phases: feature extraction with properties of line segments and classification with Support vector Machine (SVM). In feature extraction phase, a set of features is constructed using properties of line segments those connecting certain facial feature points on 3D facial expression image. In this phase, the set of 96 features is extracted from each image. This feature set consists of normalized distance and slope of line segments connecting a subset of 83 feature points. This set of features is used to train the SVM classifier. Author constructed multi-class SVM classifier using one against One Approach. This system is analyzed with BU-3DFE database and system achieves 87.1 % of emotion recognition accuracy rate. But this system fails to classify the neutral emotion because a fiducially measure of neutral emotion is removed from six universal facial expressions during the preprocessing step.

One more system with LDA classifier is proposed in [6]. This system performs feature extraction process same as explained in [5] on BU-3DFE database. But this system with LDA classifier only achieves 83.6% of the average expression recognition rate. With SVM, surprise emotion recognition reaches to 99.2% but for fear it slightly reduces by 0.8%.than LDA.

The human facial expression recognition method based on ANN is proposed in [7]. This is the automatic technique that detects 22 important facial feature points and generates facial feature vector by using Euclidian distances between particular points. The Multi-Layer Perceptron (MLP) neural network with back Propagation learning algorithm issued as the network classifier to classify the facial expression. This expression is classified from a set of seven basic expressions like happy, sad, surprise, fear, anger, disgust and neutral. This system achieves high accuracy for front face images and for some small left and right angle rotated face images. But the accuracy reduces for large angle rotation of face in an image and also reduces because of strong sidelight. The system is tested with 240 images taken from color FERET database. The system is trained with ANN feed forward back propagation classifier. The system got 100% accuracy for trained database and 85% accuracy for test database. It is quite difficult to detect the angry, disgust and sad emotions on a face as they are similar in some condition. The system fails to classify the emotion if the person is too old with many wrinkles on his/her face. A simple approach for automatic recognition of facial expression is presented in [8]. This approach uses HMM for recognition of features. This system successfully extracts the features and detects the emotions from given variable sized real time videos. This system is a combination of HMM, Code HMM, and ANN. The motion of the face is modeled by HMM as follows, first, according to the function of HMM in processing continuous dynamic signal and model recognition, and for the sample's overlap and similarity in the sample space, Code- HMM was made up respectively; then, inducted KNN and some discrimination rules by analyzing the output result. This system recognizes the happy expression with 94% accuracy. The system also recognizes the neutral and disgust expression very well.

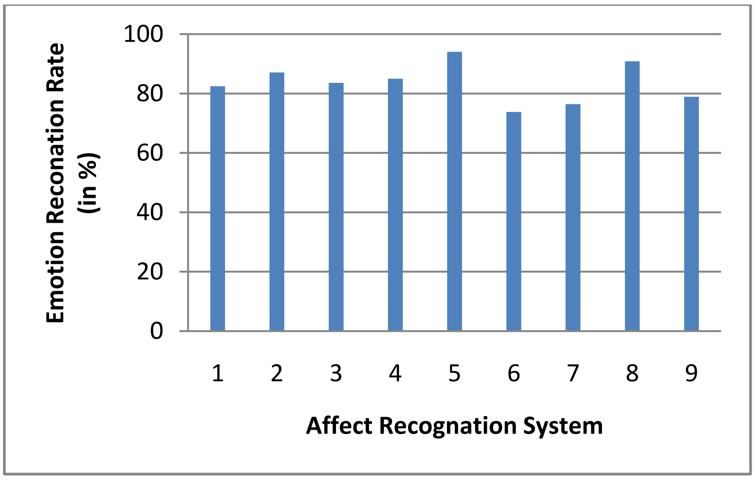

Fig 2: Emotion Recognition Rate of Basic Classification Systems

Table 1. Human Affect Recognition Systems

Sys tem	Classification	Feature Extraction	Database Used	Emotion Recognition Rate
1	K-nearest Neighbor [4]	Cotourlet Transformation	JAFFE	82.47
2	Multi Class SVM [5]	Line Segment Properties	BU-3DEF	87.1
3	LDA [6]	Line Segment Properties	BU-3DEF	83.6
4	MLP with Back Propagation [7]	Feature Vector by Using Euclidian Distances Between Particular Ppoints	COLOR FERET	85
5	Code HMM and ANN [8]	AMM	-	94
6	Deep Belief Network Models [15]	Forward Selection, Information Gain (IG), and Principal Component Analysis (PCA)	IEMOCAP	73.78%
7	SVM [19]	Drift and Noise Reduction With Band-Pass Filter and Feature Extraction With ANOVA Test	Database with Eye Gaze Data and EEG Signals	76.4 %
8	Optimized Kernel Laplacian Radial Basis Function Neural Classifier [21]	Principal Component Analysis (Pca) and Linear Discriminant Analysis (LDA)	Own Recorded Video Database	90.83
9	SVM [14]	Open-Source Audio Feature Extractor, openSMILE	IEMOCAP	78.4

Authors of [14] examined the use of differing timescales in constructing emotion classifiers. They used the Interactive Emotional Dyadic Motion Capture (IEMOCAP) database, which contains audiovisual information. The fusion method is presented which makes use of the individual classifier's output that is trained using multi-dimensional inputs with varying temporal lengths. The classification task was performed over three emotional dimensions: valence, activation, and dominance. The classifiers with 400 ms-analysis length with multi-model multi-temporal approach outperformed the utterance level classifiers in all cases but the differences are not significant. Also, the system proves that SVM is better than GMM classifiers but it is also insignificant. This system shows significant improvement if speech is analyzed at the different timescales, along with fusion. The fusion of the two-timescale analysis shows the accuracy up to 7.3 % and 6.9 % for a 12-way classifier. For activation classification, 5.5 and 4.4 % improvements were observed and for valence, 3.8 and 1.9 % points' improvements were observed.

Yelin Kim et.al proposed a deep belief network model for feature selection from audio video input data in [15]. An automatic emotion recognition system predicts high-level affective content from low-level human-centered signal cues. The paper shows high improvements in classification because of the used feature selection method. Authors focus on deep learning techniques in which the system captures complex nonlinear feature interactions in multimodal data. The system demonstrates that these models improve the performance of emotion classification over baselines that do not employ deep learning. The DBN for the non-prototypical data achieve 56.70% to 57.70% accuracy and for prototypical data accuracy is up to 70.46% to 73.78%. The performance gap between the maximal UAs of proposed model and baseline models is 0.40%. From this, the system suggests that for emotionally clear utterances, supervised feature selection is a good choice instead of unsupervised feature selection.

[16]Physiology-based detectors of non-basic affective states (e.g., boredom, confusion, curiosity) were trained and validated on naturalistic data. Data is collected during interactions between 27 students and author. Two types of detectors were developed, user-independent and user-dependent. But authors concluded that user independent detector is not feasible through physiological signals alone. The set of two feature selection methods and nine classification methods were applied to the problem of recognizing eight affective states such as boredom, confusion, curiosity, delight, flow/-engagement, surprise, and neutral. Therefore, this study evaluates the efficacy of affect detection using a host of feature selection and classification techniques on three physiological signals electrocardiogram (ECG), electromyogram (EMG), galvanic skin response (GSR) and their combinations. Within-participants 10-fold cross validation. A classification task x classifier matrix of kappa scores obtained by using 10-fold cross-validation. The results prove that the average scores associated with the specific emotion versus neutral discrimination were 0.35. A performance was lower for the most difficult discrimination task involving all seven affective states with the neutral state. Minor changes to this classification task, such as including neutral but excluding infrequent states such as delight and surprise and excluding both neutral as well as delight and surprise, produced moderate improvements in classification accuracy. The results also show that for a three-factor (low, intermediate, and high) arousal discrimination was quite higher than accuracy for a two-factor valence (positive and negative)

The system in [17] consists of four major parts; speech acquisition, feature extraction at each timescale level, machine learning for all set of features, and information fusion to merge the information. In the proposed real-time emotion detection system, emotional information encoded at different timescales like supra- and Intra- frame level is extracted and fused as emotions are encoded at different speech levels, like as supra-frame, intra-frame level. A new information fusion algorithm which takes care of the characteristics of individual machine algorithms associated with different timescale features is introduced. The system shows the Equal Error Rate for spectral and prosody features under different settings and variable speech segment. EFR is low for prosody features for short segments. Whereas for spectral features EFR is low when number if Gaussian mixture is 65 for all segment lengths. Therefore there is need of proper adjustment of k according to segment length in real-time applications while the number of GMM can be fixed.

In the paper [18] most common and famous systems which are used for face recognition are compared these are face-based affect recognition and one which involves the utilization of brain-computer interfaces. According to authors, it will be preferable to the estimations retrieved from face-base affect recognition framework by making use of a BCI. In opposite case it is best to BCI values by making use of the values from the face-based affect recognition framework creates models with low correlation. So authors concluded that both the systems are not reliable.

Paper [19] has introduced a new computational system by making use of video content as well as brain human brain's activity in reaction to videos for video arousal recognition. Authors investigated the low-level audio-visual features as well as fMRI-derived features taken out from fMRI data when subjects are watching the videos for the finding the arousal levels of videos. The system is not able to find emotions in various dimensions like valence, dominance. With modality fusion strategy and a support vector machine, classification accuracy reached to 68.5 % for three labels of valence and 76.4 % for three labels of arousal.

An audio-visual emotion recognition system is proposed in [21]. It is the combination of rule-based technique and machine learning technique. It improves an efficiency of audio and video recognition. In video recognition, Bidirectional Principal Component Analysis (BDPCA) and Least-Square Linear Discriminant Analysis (LSLDA) techniques are used for the reduction of features and discrimination of class. These features are then sent to Optimized Kernel-Palladian Radial Basis Function (OKL-RBF). It is the neural based classifier. For audio recognition, two types are features are used such as prosodic and spectral features. After that to identify the similar emotions in an audio clip, these extracted features are used. Finally, the output of both modules is combined to improve the accuracy of audio video recognition system. The performance of visual emotion recognizer is evaluated using CK+ dataset. This system achieved 96.11% of recognition rate and it is better that other method such as SNMF, MFA etc. With eNTERFACE 05 and RML databases, recognition rate for a visual and audio path is 79% and 51.7% respectively.

2.2 Datasets used for Affect Recognition

- Here present some audiovisual emotion databases, for training and evaluation in emotion recognition systems. The collected databases are recorded in English language and classify the features based on various emotion categories [1].
- eNTERFACE: Six emotion categories such as anger, disgust, fear, happiness, sadness, surprise
- IEMOCAP: Five emotion categories include happiness, anger, sadness, frustration, and neutral); 3 dimensions as valence, activation, dominance.
- RML: Six emotion categories such as anger, disgust, fear, happiness, sadness, surprise.
- VAM: Three dimensions (valence (negative vs. positive), activation (passive vs. active), dominance (weak vs. strong))
- SAVEE: Seven emotion categories (anger, disgust, fear, happiness, sadness, surprise, neutral)

- TUM AVIC: Five level of interest; 5 non-linguistic vocalizations (breathing, consent, garbage, hesitation, laughter)

Surrey Audio-Visual Expressed Emotion (SAVEE) database: has been recorded as a pre-requisite for the development of an automatic emotion recognition system. The database consists of recordings of 4 male actors in 7 different emotions, 480 British English utterances in total. The sentences were chosen from the standard TIMIT corpus and phonetically-balanced for each emotion. The data were recorded in a visual media lab with high-quality audio-visual equipment, processed and labeled. To check the quality of performance, the recordings were evaluated by 10 subjects under audio, visual and audio-visual conditions. Classification systems were built using standard features and classifiers for each of the audio, visual and audio-visual modalities, and speaker-independent recognition rates of 61%, 65%, and 84% achieved respectively. [9].

TABLE II Comparative Analysis of Available Databases

Database / Properties	SAVEE	SEMAINE	JAFFE	IEMOCAP
Number of Subjects	10 Males In 7 Different Emotions	150 Participants From 7 Different Countries	10 Japanese Females	10 Subjects (Five Males And Five Females), And Two Subjects Form A Pair For Dyadic Conversions
Number of Samples	480 Utterances	959 Recordings	213 Images Of 7 Facial Expressions	12 Hours Of Audiovisual Data, Including Video, Speech, Motion Capture Of Face, Text Transcriptions.
Emotions Description	6 Basic Emotions (Anger, Disgust, Fear, Happiness, Sadness And Surprise) And Neutral Emotion	Fear, Anger, Happiness, Sadness, Disgust, Contempt, Amusement	Happy, Angry, Disgust, Fear, Sad, Surprise And Neutral	Anger, Happiness, Excitement, Sadness, Frustration, Fear, Surprise, Other And Neutral State
Published Year	April 2011	May 2014	Jan 2007	Des 2008
Availability	Freely Available	Freely Available	Freely Available	Not Freely Available
Download Link	Http://Kahlan.Eps.Surrey.Ac.Uk/Savee/Download.Html	http://semainedb. eu/.	http://www.kasrl.org/jaffe-download.html	http://sail.usc.edu/iemocap/iemocap_release.htm

SEMAINE:

The SEMAINE corpus consists of emotionally colored conversations. Users were recorded while holding conversations with an operator who adopts in sequence four roles designed to evoke emotional reactions. The operator and the user are seated in separate rooms; they see each other through teleprompter screens and hear each other through speakers. To allow high-quality recording, they are recorded by five high-resolution, high frame rate cameras, and by four microphones. Fully continues (i.e. time and value) annotations of 21 dimensions including valence, arousal, power, and the six basic emotion by between three and seven raters per session is included. A full transcription of the conversations is included as well [10].

IEMOCAP:

To support and expand the research in human-machine communication and to ease the experimental analysis, Interactive emotional dyadic motion capture database (IEMCOAP) is designed in [20]. Using the IEMOCAP database, gestures, and speech from different subjects can be analyzed for modeling personal styles. Facial expressions are detected with 10 actors who recorded three scripts in a fictitious scenario, by setting the markers (diameter=4mm) at head, face and also hands in dyadic interactions. Emotions like, neutral state, happiness, sadness, anger, frustration, surprise, fear, disgust and excited are captured.

Japanese Female Facial Expression (JAFFE) Database:

This database consists of 213 images of 7 facial expressions such as Happy, Angry, Disgust, Fear, Sad, Surprise and neutral. These images are pose by 10 Japanese females [11]. To recognize the facial expressions in images of JAFFE database, Gaussian Process (GP) Classification technique is used in. This classifier achieves the accuracy up to 93.43 %, without a need for feature selection and extraction process, as it makes use of leave-one-out cross-validation approach [12]. Fei Cheng et. al. has found one problem with image size of JAFFE database. They have resized the images to reduce the impact of background. Also, another problem is that, in some cases, distinct expressions may not be very distinguishable [12].

Therefore there is need of more rigorous examination so that the images of the same person with the same expression are not considered in the training phase. Dan Duncan et. al.have also identify some problems in JAFFE Database. In this database many different facial expressions were incorrectly classified as ‘fear ‘one issue is that the facial expressions in the JAFFE dataset are quite subtle, exacerbating the ability to differentiate emotions. Another issue is that there are few images labeled with ‘fear' and ‘disgust', making it difficult to train the classifier to recognize these two emotions correctly. Also, images of a database are in grey-scale format. The lack of RGB format can sometimes exacerbate the ability of the classifier to distinguish between important features and background elements [13].

3. PROPOSED SYSTEM

For given input video IEMOCAP dataset, generate uni-model multitemporal based fusion classifier by using Semi-supervised learning system. Generate the classifier in such a way that, it accurately and efficiently classify the emotions on a face in a given input test video, on the basis of three emotion dimensions such as activation, valence, and dominance. Figure 4 depicts the architectural view of a proposed system.

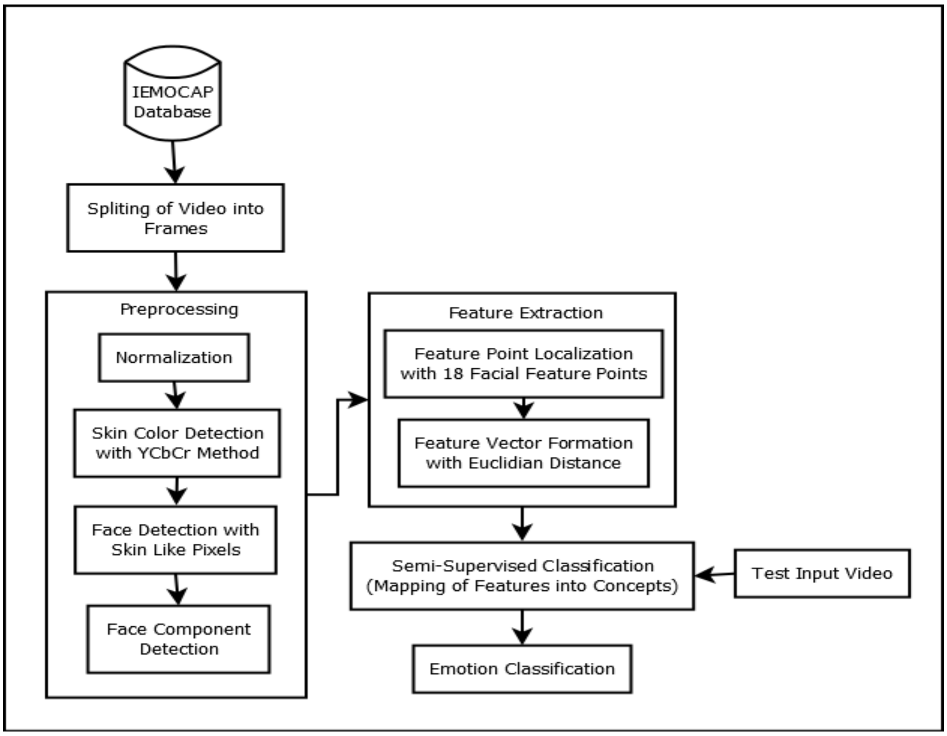

Figure 4: Architectural View of Proposed system Architecture

Steps of Proposed System Working

3.1 Video Framing

In this step videos are divided into equal sized frames at every 0.1th sec. For this framing, image cutter function is used. In this unction, initially lengths of video are calculated and start from initial point of video and cut the video after every 0.1th second. This is considered as a single frame or image from video. This procedure is repeated for all training and testing videos.

3.2 Preprocessing

Normalization:

All frames are normalizing to a size of [1000, 1000]. It provides a small detailed observation of all image values.

Skin Color Detection

The detection of skin color is very important for any kind of face detection. In this step, every type of skin color is detected changing from person to person [22]. For this purpose, YCbCr color space method [23] is used. In this method, pixels are classified by using Cb and Cr Components. Good skin color segmentation is segments the every skin color including blackish, yellowish, brownish, whitish. It also provides accurate results under different light conditions.

Face Detection

After the detection of skin like pixels, there is a need of main face image region detection. Sometimes background of image might contain skin like pixels [24]. But this is not the interested region. To get region of interest, connected region plays very important role. For this, initially maximum connected components [28], [31] of binary image are calculated. After this, object with maximum connected pixels that fits the face shape is kept and other pixels are considered as 0.

Eyes and Mouth Point Detection

After getting the maximum connected region in the image, to calculate the feature points there are a need to find out eyes and mouth boundaries [26]. The detection of eyes and mouth boundaries, depends on the color space components measurements [25] of an image.

3.3 Feature Extraction

Feature point localization:

After face parts boundary detection, feature vector is calculated [29]. Total 18 feature points are used, as shown in above figure 5.

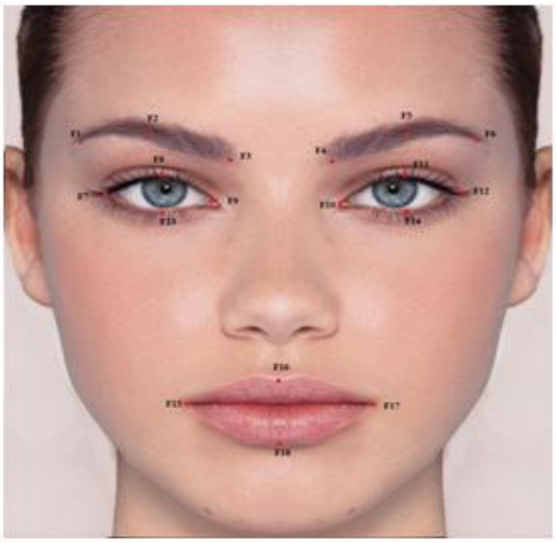

Figure 5: Facial Feature Points

Feature Vector Formation:

For each image frame, feature vector is generated, which contains 9 feature elements [7]. Distance between two pixels or two feature points of image is calculated by using Euclidian Distance [31].

Feature Vector: {D1, D2, D3, D4, D5, D6, D7}

Where,

D1= Distance between the lip outer comers

D2= Distance between the lip upper and lower comers

D3= Distance between the eyebrows inner comers

D4= Distance between eyebrow outer corners

D5= Distance between the eyebrows mid points

D6= Left eye height

D7= Right eye height

3.4 Semi Supervised Classification

In this phase, test video is classified, according to set of six emotions. In this phase, test video converted into frames, then all frames are preprocessed and features are extracted. By using these extracted features of test image and trained features, emotion is classified. For classification, this system makes use of Concept Map[32]. After feature extraction of all frames of input video, concept map is created. It shows the relationship between each feature, its value and particular identified emotion label. This is repeated for all number of frames. All this data is then considered as a training data. Now for actual affect recognition, test video is provided as input. Again for this test video, concept map is generated. This concept map of test video is classified against trained concept map. This procedure will return the emotion label of test video. Video is classified on the basis of six standard human emotions.

4. CONCLUSION

This survey presents the automated system that can recognize and interpret the emotions of humans face. The problem arises only when the user provides fake expressions on their face. This survey presents some recent approaches for human affect recognition systems. For given input video, frames are generated. Also some audiovisual databases can be used for analysis of such system have also explained in details. These frames are gone through various steps to identify the human face expressions or emotions. This paper presents the identified advantage and limitation of each approach along with the classifier used for emotion classification. This survey presents semi supervised classification for affect recognition. As a feature scope of the system we try to implement Affect recognition for multimodal affect classification at various temporal lengths.

5. REFERENCES

[1] Wu, Chung-Hsien, Jen-Chun Lin, and Wen-Li Wei, "Survey on audiovisual emotion recognition: databases, features, and data fusion strategies,"APSIPA Transactions on Signal and Information Processing 3 2014, pp.12.

[2] Mishral, Swati, and Avinash Dhole, "A Survey on Facial Expression Recognition Techniques," International Journal of Science and Research (USR) Index Copernicus Value 20 l3, pp. 6.

[3] Affective Computing. In Wikipedia, Retrived Sept 29,2016.From https://en.wikipedia.org/wiki/Affective_computing.

[4] Suresh, R., and S. Audithan, "Contourlets for facial expression analysis uses one nearest neighbor classifier," Current Trends in Engineering and Technology (ICCTET), 2nd International Conference on. IEEE, 2014.

[5] Tang, Hao, and Thomas S. Huang, "3D facial expression recognition based on properties of line segments connecting facial feature points," IEEE International Conference onAutomatic Face & Gesture Recognition, 2008.

[6] Jun Wang, Lijun Yin, Xiaozhou Wei, and Yi Sun, "3D Facial Expression Recognition Based on Primitive Surface Feature Distribution," CVPR 2006.

[7] Borah, Sagarika, and Sharmila Konwar, "ANN based human facial expression recognition in color images," International Conference onHigh Performance Computing and Applications (ICHPCA), IEEE, 2014.

[8] Punitha, A., and M. Kalaiselvi Geetha, "HMM Based Real Time Facial Expression Recognition," International Journal of Emergine Technology and Advanced Engineering 3.1 2013, pp. 180-185.

[9] Retrieved Oct 19, 2016, from http://kahlan.eps.surrey.ac.uk/savee/

[10] McKeown, Gary, et al., "The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent," IEEE Transactions on Affective Computing 3.1, 2012, pp 5-17.

[11] Retrieved Oct 19, 2016, from http://www.kasrl.org/jaffe.html

[12] Cheng, Fei, Jiangsheng Yu, and Huilin Xiong, "Facial expression recognition in JAFFE dataset based on Gaussian process classification," IEEE Transactions on Neural Networks 21.10, 2010, pp. 1685-1690.

[13] Duncan, Dan, Gautam Shine, and Chris English, "Facial Emotion Recognition in Real Time," Stanford University.

[14] Kim, Jonathan C., and Mark A. Clements, "Multimodal affect classification at various temporal lengths," IEEE Transactions On Affective Computing 6.4, 2015, pp. 371-384.

[15] Y. Kim, H. Lee, and E. M. Provost, "Deep learning for robust feature generation in audiovisual emotion recognition," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., 2013, pp. 3687–3691.

[16] O. AlZoubi, S. K. D'Mello, and R. A. Calvo, "Detecting naturalistic expressions of nonbasic affect using physiological signals," IEEE Trans. Affective Comput., 2012, vol. 3, no. 3, pp. 298–310.

[17] S. Kim, P. G. Georgiou, S. Lee, and S. Narayanan, "Real-time emotion detection system using speech: Multi-modal fusion of different timescale features," in Proc. IEEE 9th Workshop Multimedia Signal Process., 2007, pp. 48–51.

[18] J. Gonzalez-Sanchez et al., "Affect Recognition in Learning Scenarios: Matching Facial- and BCI-Based Values," 2013 IEEE 13th International Conference on Advanced Learning Technologies, Beijing, 2013, pp. 70-71.

[19] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," IEEE Trans. Affective Comput., 2012, vol. 3, no. 2, pp. 211–223.

[20] Carlos Busso, "IEMOCAP: Interactive Emotional Dyadic Motion Capture Database", Springer Science, 5 nov 2008.

[21] Seng, Kah, Li-Minn Ang, and Chien Ooi. "A Combined Rule-Based and Machine Learning Audio-Visual Emotion Recognition Approach." IEEE Transactions on Affective Computing, TAFFC 2016.

[22] Hong-Bo Deng, Lian-Wen .lin, Li-Xin Zhen, Jian-Cheng Huang, " A New Facial Expression Recognition Method Based On Local Gabor Filter Bank And PCA Plus LDA " , International Journal Of Information Technology Vol. I I No . 11 2005 ,Pp 86-87.

[23] Jure KovaT, Peter Peer, And Franc Solina, "Human Skin Colour Clustering For Face Detection", Digital Image Processing - January 29, 2013, Pp. 1-19.

[24] Borah, Sagarika, and Sharmila Konwar. "ANN based human facial expression recognition in color images." High Performance Computing and Applications (ICHPCA), 2014 International Conference on. IEEE, 2014.

[25] R. L. Hsu, M. Abdel-Mottaleb, A. K. Jain, "Face Detection In Colour Images", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 24, No. 5, Pp . 696-706, 2002.

[26] S .P.Khandait , Dr. R.C.Thool , P .D.Khandait , "Automatic Facial Feature Extraction And Expression Recognition Based On Neural Network" , (IJACSA) International Journal Of Advanced Computer 2011 Science And Applications, Vol. 2, No.1, January.

[27] Akshat Garg, Vishakha Choudhary, "Facial Expression Recognition Using Principal Component Analysis" International Journal Of Scientific Research Engineering &Technology (IJSRET),Volurne 1 Issue4 Pp 039-042 July 2012, ISSN 2278 – 0882

[28] C. A. Bouman, "Connected Component Analysis", Digital Image Processing - January 29, 2013, Pp. 1-19.

[29] Sagarika Borah, Nitumoni Hazarika, Dr.T.Tuithung , "22 Human Facial Feature Point Detection and Generation of facial feature vector in color images", IEEE International Conference on Advances in Engineering and Technology, ICAET20 14, 978-1- 4799-4949-6 .

[30] Sagarika Borah, Sharmila](onwar, Dr. T.Tuithung , Rahul Rathi,"A Human Face Detection Method Based on Connected Component Analysis", IEEE International Conference on Communication and Signal Processing-ICCSP'14, , April 3-5, 2014, India, 987-1-4799-3357-0,pp 748-751.

[31] Wang, Liwei, Yan Zhang, and Jufu Feng. "On the Euclidean distance of images." IEEE transactions on pattern analysis and machine intelligence 27.8 (2005): 1334-1339.

[32] Novak, Joseph D., and Alberto J. Cañas. "The theory underlying concept maps and how to construct and use them." (2008).