STUDY OF HUMAN EMOTION DETECTION USING VIDEO: A REVIEW

***Abstract*-Emotion Recognition has been a core area of research for improvement of Human-Machine Interactions (HMI). Recent works aim at smooth interaction between man and machine. It is necessary to develop algorithms for proper recognition of non-acted features with least possible complexity. The applications of this field of study include areas of medicine, education, voice recognition, customer feedback, lie detection and many more. The collection and detection human emotion can be done using voice, facial images, videos, physiological signals or the combinations of these. Features like Power Spectral Density of the signals, displacements of body extremities, facial expressions are used for feature extraction. The analysis can be done by selection of one or more effective features to gain higher accuracy level. The commonly used classifiers are Support Vector Machines (SVM) or Hidden Markov Models (HMM). Addition of a neural network classifier can further improve the recognition accuracy. This paper reviews contributions in the field of application of emotion recognition. It aims to help a driver, who is fatigued or under substance abuse. The proposed work is to recognize a driver's expression. If the expressions of fatigue or anger are recognized, then the job is to alarm the driver with a sound. This may help to avoid accidents.**

1 INTRODUCTION

Emotion detection is a challenging and important field of research. It has wide applications in the development of Human Machine Interaction [1]. The application fields include education and medicine. An interactive teaching method can be developed using this. Persons who cannot speak clearly can communicate using emotions. Head gesture recognition system can be developed which will be beneficial for persons with disabilities [2]. The system has the ability to identify, recognize and interpret the human gestures in order to control some devices. The camera used to capture the human face also has an important role. EEG signal analysis has been used for depression detection from brain waves. Video signal analysis has been used for facial feature detection, eye movement, facial emotion recognition etc. Diagnosing depression in the early curable stages may save the life of a patient [3]. In a clinical setting, pain can be detected automatically using a series of facial action units. The results are better compared to patient self-report [4]. Lie detection tests make use of interactional dissynchrony [5]. Smooth interaction between man and machine would enhance these applications. Prior works have been done in this field by using video, audio, physiological signals or their combinations as input [6]. Fourier transformation, Wavelet transformation, Gabor's transformation applied to the input data result in feature extraction. The extracted features are then classified into the different categories of emotions.

Emotion detection involves observing the responses of a person to a particular stimulus [7]. Various stimuli can be used to elicit response. The results to the acted stimuli under controlled environment are processed for emotion recognition. The recent field of Affective Computing (AC) aims to reduce the communicative gap between the highly emotional human and the emotionally challenged computer by developing applications that recognize and respond to the emotions of the user. [8]

Significant works in Literature have proposed emotion detection with the aid of video [9] [10] [11]. Volunteers are recruited to participate in the experiment. The resultant videos are processed. Existing datasets can be used for processing, for e.g. Cohn-Kanade Dataset, Geneva Multimodal Emotion Portrayals (GEMEP). The videos were segmented into various frames using different algorithms [12] [13] [14] [15]. The frames for which significant changes have been observed were saved in a buffer. For forensic analysis or surveillance applications, the enormous volume of video data is reduced by using an object-based video synopsis. This captures the exact frames where significant changes have been observed [16]. The values in the buffer are transformed, using Fast Fourier Transform or

Gabor's Transform. Skin detection or face detection algorithms are applied. A novel face detection system can identify and locate all faces irrespective of their position, scale, orientation, light conditions and expressions [17] [18]. Once the face has been successfully detected the facial feature tracking was performed. This was done using various models like Active Shape Model [19]. The facial features were then extracted for feature selection. Feature Selection was mainly done using statistical testing. Analysis of Variance (ANOVA) is applied on the two-dimensional valance arousal model [20]. Various classifiers were used for distinguishing the emotions. The commonly used classifiers are Support Vector Machines [21], Hidden Markov Models [22], Baum-Welch Algorithm [23]. The detected emotion can be used to buzz a signal to indicate that the driver should check himself.

II THEORIES OF EMOTION

Emotion is defined as a strong feeling derived from one's circumstances, mood or relationships with others [24]. Many researchers have tried to remain agnostic to the emotion theories. Different categories of emotions have been suggested by different psychologists. The six most widely accepted emotion categories are happy, sad, surprise, fear, disgust, anger. Various models of emotions have been proposed by various psychologists but the two mostly used models are discrete emotional model by Ekman [25] and two dimensional valance arousal model by Lang [26].

The discrete emotional model proposed by Ekman defines the six basic emotions of happiness, sadness, anger, fear, disgust and sadness to be universally present among humankind.

The two dimensional valance arousal model is characterized by the categorizations of valance and arousal. Valance defines the pleasantness and arousal defines the activation for a particular emotion.

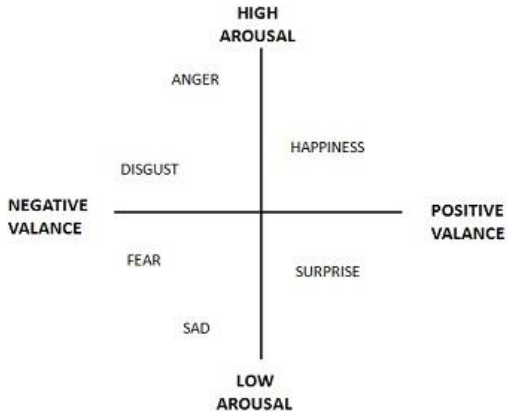

Fig 2.1: Basic emotions based on the two-dimensional valance arousal model

III IMPLEMENTATION

A. Emotion Eliciting Stimuli:

Videos collected from various modes can be used for emotion detection. Prior works have been done by collecting videos by acted emotions. Actors enact emotions under the guidance of a professional. Data has also been collected by showing different kinds of videos obtained from movies, plays or pictures [27] to the volunteers and recording their emotions. A participants' "zapping behavior" yields response to a particular commercial [28]. It is used to elicit emotion. Neutral expression data was collected as reference [29]. Prominent changes in expression provide good database for analysis. Non-acted [30] [31] emotions provide good data for analysis as they are more natural. Non-acted emotion detection is also a more challenging field of research.

B. Pre-processing:

The collected data is generally contaminated with noise and other interferences. To reduce noise different filters like Adaptive Filters [32], Elliptic Filters [33], Butterworth Filters [34] were used. Also, down-sampling [35] was done if physiological signals were also included to reduce memory occupied [36]. The beginning and end of the videos were omitted in order to reduce the effect of external interferences.

C. *Video Segmentation:*
Collection of videos was manually saved. The videos were segmented into frames. Numerous algorithms can be applied for video segmentation. The frame in which considerable change was detected was saved in a buffer. The changes are saved with respect to a threshold value, which is calculated at runtime. These frames are further analyzed.

D. *Face Detection and Facial Feature Tracking*:
Face detection is responsible for locating a face within an input image. Facial landmarks like position of eye, nose can be determined after face detection. For tracking the facial features, the eyes, nose, lips are tracked from each frame and their location is determined. This was used to find accurate location of facial features and track their deformation. These deformations are particularly used for determining the emotion. The feature points were individually searched to analyze the variations.

E. *Training*: This was done using various mathematical models. The methods of Learning using Privileged information or Fuzzy Logic System were applied. Action units are the smallest visible facial movements. They are generally local muscle movements. These were classified and matched with a dataset. This dataset has been used for training. Individual or combinations of action units determine the resultant emotion. This is based on the frequency of occurrence of the action units.

F. *Emotion Detection*: The emotion detected was used for developing the driver's assistance system. The "sleepy", "angry" or "fatigued" states of emotions were recognized for this. A signal is generated to indicate the driver to check himself. An additional speed-checker can be included in the system to keep the speed of the automobile in control. This aids in minimizing road accidents.

CONCLUSION

In this paper, the researchers have reviewed the different steps of human emotion detection with the help of videos. The motive is to utilize the emotion detection technique to minimize road accidents by building a driver's assistance system. Here it should be mentioned that no differentiation between acted and non-acted emotions has been included in this review. Acted emotions were omitted while developing this system as it is expected that a driver will be in a natural state of mind while driving. Emotion detection using video cannot be applied to detect masking of emotions. Multimodal emotion recognition using physiological [37] signals will be apt for that. Human emotion change is clearly observable for 3-15 seconds [38] [39]. Thus, if the videos can be captured during this time period, more accurate results may be obtained. Non-acted, neutral emotion recognition can be improved using a more vigorous feature selection and classification algorithm.

The future work of this review would be employed for developing a drivers' assistance system using human emotion detection. The emotion of the driver in the driving state is detected using video. The driver would be indicated to check himself by buzzing a signal in case any fatigue or sleepiness is detected [40].

REFERENCES

Ref No	Paper Name	Bio-signals Used	No. of Subjects	Emotions	Stimuli Used	Feature Extraction	Classification	%accuracy
2	Analyses of a Multimodal Spontaneous Facial Expression Database	None	215 (157 males, 58 females)	Happiness, anger, disgust, fear, sadness, surprise	Videos collected from the internet	Manual, Interrater Reliabilty	Manual	99%
11	Automatic Emotion Recognition in Video	None	User-dependant	Happiness, sadness, feat, disgust, surprise, anger	Video sequences	Displacement of facial features	Action Unit	>50%
19	Understanding Discrete Facial Expressions in Video Using an Emotion Avatar Image	None	MMI Dataset, Cohn-Kanade Dataset, GEMEP-FERA Dataset	Anger, fear, joy, relief, sadness	Emotion Avatar Image	Local Binary Pattern, Local Phase Quantization	Support Vector Machines	82.6%
20	Automatic Recognition of Non-Acted Affective Postures	Motion Capture Data	11 (6 females + 5 males)	Concentrating, defeated, frustrated, triumphant	Most expressive motion stimuli	Z Rotation, Y Rotation, X Rotation of low level features	Ground Truth	66.7%
23, 24	Multimodal Emotion Recognition in Response to Videos	Electroencephalogram signals, pupillary reflex, eye gaze data	24	Calm, medium aroused, activated	Multimodal (facial video, audio, vocal expressions)	Logarithms of Power Spectral Density	Ground Truth	68.5% (for three labels of valence) 76.4% (for three labels of arousal)
25	Emotion Recognition with the Help of Privileged Information	electroencephalogram	USTC-ERVS, MANHOB-HCI, DEAP	Valence, arousal	Multimodal	Logarithms of Power Spectral Density, lightning key, color energy, visual excitement	Support Vector Machines	78.68%, 77.66%, 76.65% (Valence) 73.10%, 73.10%, 76.14% (arousal)
26	Technique for automatic emotion recognition by body gesture analysis	EyesWeb XMI	Geneva Multimodal Emotion Portrayals(10 actors)	Anger, joy, relief, sadness	EyesWeb XMI Data	Displacement of bounding triangle of body extremities, energy	Manual	99% (BonFerroni corrected alpha value)

30	Emotion and Gesture Recognition with Soft Computing Tool for Drivers Assistance System in Human Centered Transportation	None	Manual	Anger, fear, sadness, fatigue	Video Sequences	Centroids of eyes, centroids of lips	Sugeno Fuzzy Inference System	91.66% (facial gesture detection) 90% (emotion detection)

[1] Gunnar Johansen, "CONTROL SYSTEMS, ROBOTICS, AND AUTOMATION" – VOL XXI – Human Machine Interaction

[2] Rushikesh T. Bankar and Dr. Suresh S. Salankar, "Head Gesture Recognition System Using Gesture Cam", 2015 Fifth International Conference on Communication Systems and Network Technologies

[3] Yashika Katyal, Suhas V Alur, Shipra Dwivede and Menaka "EEG Signal and Video Analysis Based Depression Indication, 2014 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)

[4]Patrick Lucey, Jeffrey F. Cohn, Iain Matthews, Simon Lucey, Sridha Sridharan, Jessica Howlett and Kenneth M. Prkachin, "Automatically Detecting Pain in Video Through Facial Action Units", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 41, NO. 3, JUNE 2011

[5]Xiang Yu, Shaoting Zhang, Zhennan Yan, Fei Yang, Junzhou Huang, Norah E. Dunbar, Matthew L. Jensen, Judee K. Burgoon, and Dimitris N. Metaxas, "Is Interactional Dissynchrony a Clue to Deception? Insights From Automated Analysis of Nonverbal Visual Cues", IEEE TRANSACTIONS ON CYBERNETICS, VOL. 45, NO. 3, MARCH 2015

[6]Shangfei Wang, Zhilei Liu, Zhaoyu Wang, Guobing Wu, Peijia Shen, Shan He, and Xufa Wang "Analyses of a Multimodal Spontaneous Facial Expression Database", IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, VOL. 4, NO. 1, JANUARY-MARCH 2013

[7] Jerritta S, M Murugappan, R Nagarajan, Khairunizam Wan, "Physiological Signals Based Human Emotion Recognition: A Review"

[8] Rafael A. Calvo and Sidney D'Mello, "Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications", IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, VOL. 1, NO. 1, JANUARY-JUNE 2010

[9] Donald Glowinski, Antonio Camurri, Gualtiero Volpe, Nele Dael, Klaus Scherer , "Technique for automatic emotion recognition by body gesture analysis"

[10] George A. Tsihrintzis, Maria Virvou, Ioanna-Ourania Stathopoulou, and Efthimios Alepis, "On improving visual-facial emotion recognition with audio-lingual and keyboard stroke pattern information"

[11] Yasira Beevi C P Dr. S. Natarajan, "An efficient Video Segmentation Algorithm with Real time Adaptive Threshold Technique", International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 2, No.4, December 2009

[12] Shao-Yi Chien, Yu-Wen Huang, Bing-Yu Hsieh, Shyh-Yih Ma, and Liang-Gee Chen , "Fast Video Segmentation Algorithm With Shadow Cancellation, Global Motion Compensation, and Adaptive Threshold Techniques", IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 6, NO. 5, OCTOBER 2004

[13] Patrick Chiu, Andreas Girgensohn, Wolf Polak, Eleanor Rieffel, Lynn Wilcox, "A GENETIC ALGORITHM FOR VIDEO SEGMENTATION AND SUMMARIZATION"

[14] M. EL Hassani, S. Jehan-Besson, L. Brun, M. Revenu, M. Duranton, D. Tschumperl, D. Rivasseau1, "A Time-Consistent Video Segmentation Algorithm designed for Real-Time Implementation"

[15] Tung-Chien Chen, "Video Segmentation Based on Image Change Detection for Surveillance Systems"

[16] Yogameena Balasubramanian, Kokila Sivasankaran and Sindhu Priya Krishraj, "Forensic video solution using facial feature-based synoptic Video Footage Record", IET Journals The Institution of Engineering and Technology

[17]P. Ithaya Rani and K. Muneeswaran, "Robust Real Time Face detection automatically from video sequence based on Haar features", 2014 International Conference on Communication and Network Technologies (ICCNT)

[18]Kenz Ahmed Bozed and Osei Adjei, Ali Mansour, "Detection of Facial Expressions based on Morphological Face Features and Minimum Distance Classifier", 14th international conference on Sciences and Techniques of Automatic control & computer engineering - STA'2013, Sousse, Tunisia, December 20-22, 2013

[19] Van Ginneken, "Active Shape Model"

[20], [26] SA Shankman, DN Klein , "The relation between depression and anxiety; an evaluation of the tripartite, approach-withdrawal and valencearousal models", Elsevier

[21] Andreas Christmann, Ingo Steinwart, "Support Vector Machines"

[22], [23]Zoubin Ghahramani, “An Introduction To Hidden Markov Models and Bayesian Networks”

[24] Paul R. Kleinginna, Jr., and Anne M. Kleinginna, “A Categorized List of Emotion Definitions, with Suggestions for a Consensual Definition”

[25] Paul Ekman , “An Argument for Basic Emotions”

[27] Janneke Wilting, Emiel Krahmer & Marc Swerts, “Real vs. acted emotional speech”, Communication and Cognition

[28]Songfan Yang, Mehran Kafai, Le An, Bir Bhanu, “Zapping Index:Using Smile to Measure
Advertisement Zapping Likelihood”, IEEE TRANSACTIONS ON AFFECTIVE COMPUTING,
VOL. 5, NO. 4, OCTOBER-DECEMBER 2014

[29] Pojala Chiranjeevi, Viswanath Gopalakrishnan, and Pratibha Moogi, “Neutral Face Classification Using Personalized Appearance Models for Fast and Robust Emotion Detection”, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 24, NO. 9, SEPTEMBER 2015

[31],[33],[34] Scot Douglas , “Digital Signal Processing Handbook “

[35] J. Kim, "Bimodal Emotion Recognition using Speech and Physiological Changes", Robust Speech Recognition and Understanding

[37] H. Gunes, "Automatic Dimensional and Continuous Emotion Recognition", International Journal of Synthetic Emotions

[38] Zeliang Zhang Fumei Liu and Wenliang Qu, “Review of the Visual Feature Extraction Research”, 2011 IEEE 7th International Colloquium on Signal Processing and its Applications

[39] Paul Modler Frank Hofmann loannis Zannos, “Gesture Recognition by Neural Networks and the Expression of Emotions”

[41] Marko Tkalčič, Ante Odić, Andrej Košir and Jurij Tasič, “Affective Labeling in a Content-Based
Recommender System for Images”, IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 15, NO. 2, FEBRUARY 2013

[42] Faisal R. Al-Osaimi, “A Novel Multi-Purpose Matching Representation of Local 3D Surfaces: A Rotationally Invariant, Efficient, and Highly Discriminative Approach With an Adjustable Sensitivity”, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 25, NO. 2, FEBRUARY 2016

[43] Priyank Rathod , Kiran George and Nikhil Shinde, “Bio-Signal based Emotion Detection Device”

[44] Tim K. Marks, John R. Hershey, and Javier R. Movellan, “Tracking Motion, Deformation, and Texture
Using Conditionally Gaussian Processes”, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 32, NO. 2, FEBRUARY 2010

[45] Manish Dixit, Sanjay Silakari, “A Hybrid Facial Feature Optimisation Approach using Bezier Curve”, 2015 International Conference on Computational Intelligence and Communication Networks