Performance Analysis of Visual and Physiological Techniques for Affect Recognition System

[1] Thamba Meshach W, [2] Hemajothi S, [3] Mary Anita E A
[1] Associate Professor, [2,3] Professor
[1]Department of Computer Science and Engineering, Prathyusha Engineering College, Chennai, India
[2]Department of Electronics and Communication Engineering, Prince Shri Venkateshwara Padmavathy Engineering College Chennai, India
[3]Department of Computer Science and Engineering, S. A. Engineering College, Chennai, India

Abstract : Affective computing is used to develop a system which can interpret and simulate human affects. This is a multidisciplinary field which explores the use of technology to enhance the communication between human and machines. This includes the field of engineering, cognitive science and psychology. The application of affect recognition system spans a wide range in the field of telemedicine, biometric surveillance, security, interactive education, video games, health monitoring, and social monitoring. The main objective of this work is to analyze the literature and understand the fundamentals and approaches in the design of affect recognition systems, its applications and also to identify the area which requires improvement and new developments.

***Index Terms* - Affective computing, Affect, Face Expression Recognition, Physiological Signals.**

I. INTRODUCTION

The terminology 'Affect' is defined in psychology as an external expression of emotions and state of mind. Thus the objective of Affective computing is to build a system which can interpret human affect state with the input from face, biological signals or voice. The main objective of developing an affect system is to reduce the gap between the human beings who are highly emotional and emotionless computer. The application of this system is in the field of health, games and academics. In academics the on-line system can be developed to be more interactive such that learning is more interesting and motivating [1]. The applications of affect computing is classified into three types by Picard. The first type is to determine the emotions of the user. The second type is to develop a system which predicts the emotion that a person should perceive. The third type is to develop a system that actually feels an emotion. From the literature it is evident that affect recognition system consists of four vital steps, which are data acquisition, pre-processing, extraction of features and classification.

II. EMOTIONAL MODELS

From the literature the different models for describing the affective behaviours are categorical model defined by Ekman et al. [2], in which affect state is selected from a list. The effective model is the valence arousal model [3] which is 2-dimensional model. The other model is the Facial Action Coding System (FACS) where affect is related to the facial actions. Dimensional model is advantage compared to the other models based on the literature work. The valence describes the relationship between the event and the nature of the event as positive or negative, whereas the arousal is related to the event whether it is high or low activated. The advantage of the dimensional model is that the affect is classified in a continuous domain.

III. FACIAL EXPRESSION RECOGNITION SYSTEM

For communication, facial expression are more important which may be verbal or non-verbal or a combination of both. Under non-verbal communications there are many types as contact between eyes, body language, gestures or facial expressions. Among these, facial expressions plays a major role in communication. For example, a smile express a person's happiness. The feeling of sadness express the loss which is expressed by skewed eyebrows. The anger is expressed with squeezed eyebrows, stretched eyelids which express the unpleasant condition. The disgust is expressed in face with pull down eyebrows. The surprise is identified by widened eyes and open mouth. Thus the face expressions based on different circumstances is based on geometric of eyes, eyebrows, nose and mouth. Thus human face reveals a lot of information and it can be of three types as static information, slow information and fast information. The skin colour and shape forms the static content, the muscle tone, skin texture give static information. The movements of different parts of face gives the fast information which helps in identification of facial expressions of a person.

IV. PHYSIOLOGICAL SIGNALS RECOGNITION SYSTEM

The facial expression is one of the important methodology to determine human affect and hence it plays a vital role in designing an automatic human affect system. The major drawback of this methodology in determining the human affect is that one can easily hide the expressions from the face, in which case determining the expression from face is not possible. The biological signals have been proved as a reliable source of information from which data can be acquired which cannot be altered or changed.

V. MODULES OF AFFECTIVE RECOGNITION SYSTEM

The different modules of affective recognition system are data acquisition, pre-processing, feature extraction and classification.

5.1 Data Acquisition

The input is obtained from standard various databases available like Japanese Female Facial Expressions (JAFEE, 2017, AR face database (AR, 2018), Taiwanese Facial Expression Image Database (TFEID, 2017). In literature the data for developing affect system is obtained from the databases. Though these databases are proved to be reliable, an efficient system can be built if the data could be acquired experimentally. For experimental setup, a reliable emotion elicit system is required to build an appropriate affect system. The different methodologies that are adopted are natural existence of feelings [4], film clippings [5], and musical audio [6]. The different types of signals used as input data are Electrocardiogram (ECG), Electromyogram (EMG), Blood Volume Pulse (BVP), Electro Dermal Response (EDR), Respiration Activity (RSP), Skin Temperature (ST), and Pupil Diameter (PD).

5.2 Pre-processing

Pre-processing of the image is performed such that it will increase the intelligibility of the image, enhances the image features such that the expressions could be identified easily. The image is resized and also the images are cropped and scaled for a proper dimension to standardize the process for a particular application [7]. Normalization performed by median filter is essential to reduce the illumination and also changes in the face images [8]. The size and location of the different parts of the face are identified using different algorithms like Adaboost algorithm, Haar transform [9]. Histogram equalization is an essential method to determine the illumination differences and thus improves the difference in intensities (Demir, 2014). Among the different methods determining the Region of Interest (ROI) is suitable and accurate to determine the different parts of the face. The feature extraction is important step which retrieves relevant distinguished information from the face. Based on the literature, there are five types of feature extraction method. The first method is texture based and the example of this method is Gabor filter which gives both magnitude and also phase content [10]. The second type is determining the edges las Line Edge Map (LEM) which develops the features of geometrical structure (Gao et al., 2003). An example of third type of method is Principal Component Analysis (PCA) which gets the global features and local features. Local Curvlet Transform (LCT) gets statistical information as mean, standard deviation [11]. Another efficient method to determine the features are based on distance in terms of patches [12].

5.3 Classification

The vital step in face expression recognition step is the last classification stage which discriminates between the different expressions. The Euclidean distance, Minimum Distance Classifier (MDC) are distance based classifier [13]. The kNN method estimates the relationship between adjacent pixels, based on which facial expressions are classified [14]. The statistical based model is Hidden Markov Model (HMM). Support Vector Machine (SVM) is one of the powerful supervised classifier method which gives good classification efficiency [15]. The other classifiers used for facial expressions are Decision Tree (DT), Learning Vector Quantization (LVQ), Bayesian neural network, Convolution Neural Network.

5.4 Performance Evaluation

The system performance is evaluated based on different parameters as complexity rate, classification efficiency or recognition accuracy. The efficiency of the system depends upon the efficiency of all the modules of the system. Thus reliable input data, efficient pre-processing step, robust feature extraction methodology and an effective classifier, all together develops an efficient facial expression recognition system.

Table 1 gives an overview of the recent work on affective computing using facial expressions and signals. The first five papers discusses the affect recognition system using facial recognition system. The papers from six to eleven discusses the affect recognition system using signals. The last paper is the affect recognition system for wheel chair navigation application.

Table 1: Literature Survey on Affective Computing using Facial Expressions and Physiological Signals

Reference	DB / DS	Feature Extraction Method	Classifier	Objective / Key Approach	RA (%)	Drawbacks / Challenges / Issues
[16] Hernandezmatamoros et al. (2015)	KDEF	ROI, Face Dimensions, Feature Vector (Gabor function)	SVM	In this work the face is detected from the image and it is segmented into two regions as forehead/eyes and mouth. Each of these regions is segmented into N x M blocks. These blocks are characterized by using 54 Gabor functions and it is correlated with one of the N x M blocks. The dimensions are reduced using PCA and the feature vectors are given as input to the classifier.	98	RR 98% is achieved if only one ROI is used. Only two ROI is considered, that is Forehead / Eye & Mouth
[17] Happy et al. (2015)	JAFE, CK+	Appearance Features (LBP) Local Binary Pattern (LBP)	SVM (Approach: One-against-one classification)	In this work facial patches are used which generates discriminative features that separates two expressions effectively. The features which have the maximum variation between pairs are selected and these selected features are projected into lower dimensional subspace and further classified using multi-class classifier and expressions are classified based on majority voting methodology.	93.3	-Partially occluded images are not addressed in this study. - This analysis is confined to databases without facial hairs. - Dynamics of expression in temporal domain is also not considered in this study.
[18] Hegde et al. (2016)	JAFE, Yale	Texture Feature [GF] Gabor Magnitude Feature Vector (GMFV) Gabor Phase Feature Vector (GMFV)	Euclidean distance (ED), SVM	The high dimensional image space is projected into low dimensional image subspace. This is performed by solving the singularity problems of linear discriminant analysis. For this transformation Kernel Locality Preserving Symmetrical Weighted Fisher Discriminant Analysis algorithm (KLSWFDA) is used.	88.58	The results found less compared to other approaches related to discriminative nature algorithms due to incomplete solution for singularity problem with lesser Fisher ratio value.
[19] Michael et. Al (2018)	JAFE, CK	Texture Features from noisy facial images. LDN and DGLTP descriptors are used for histogram feature extraction	SVM	The major work is to remove the pixels which are noisy in the face image by using the Enhanced Modified Decision Based Unsymmetric Trimmed Median Filter (EMDBUTMF) method	88.63	This FER-LDN-DGLTP method takes 260 ms for query image processing which is little more milliseconds than the existing methods compared in the paper.
[20] Nigam et al. (2018)	CK+, JAFFE and Yale	W_HOG (Discrete wavelet transform and Histogram of oriented gradients feature) : DWT, HOG	Multiclass SVM with one-verses-all architecture (OVA)	The work consists of four modules, they are: Face processing, Domain transformation, Feature extraction and Expression recognition. The Face processing step consists of face detection, cropping and normalization steps using viola-jones method. In domain transformation, spatial domain features are transformed into the frequency domain by applying discrete wavelet transform (DWT). Feature extraction is performed by retrieving Histogram of Oriented Gradients (HOG) feature in DWT domain which is termed as W_HOG feature.	75 (yale) 71.43 (JAFE) 88 (CK+)	For Training: Yale :60 images JAFE : 213 images CK+ : 420 images (70 images of each category – 6 categories) More Training images results in improved RA but RA vary for different emotion categories. RA is only 40% for the expression FEAR

[21] Jianhai Zhang et al (2016)	DEAP Emotions DB	EEG – Powers of 4 frequency band	PNN, SVM	Emotions were elicited by watching music videos from scalp EEG. PNN is a feed forward neural network based on the Bayesian strategy. PNN was applied in the EEG emotion recognition with the subband power features. Relief-based channel selection algorithm is used to reduce the number of channels used in classifying emotional states. Channels are ranked according to their weights.	81.76 for PNN 82.00% for SVM	Classification rate of the PNN was slightly lower than that of the SVM.
[22] Leila Mirmohamadsadeghi et, al. [2016]	DEAP	Slope of the phase difference of the RSA and the respiration : EMG,RSP	SVM	Emotions were stimulated by watching music video clips. The slope of the phase difference between the RSA and the respiration achieved the highest accuracy.	74% - valence 74% - arousal 76% - liking.	Using a larger number of features degrades the accuracy of the SVM classifier. When the features are combined with another (best two features) or several others (best four features), the accuracy dropped.
[23] Zied Guendil et al [2015]	AuBT	Wavelet Transform Features CWT: EMG, RESP, ECG, SC	SVM	Emotions recognised from physiological signals in a multiresolution approach. Physiological features were extracted from the most relevant wavelet coefficients and the feature vectors obtained from each signal were combined using multimodal fusion technique to construct one feature vector for each emotion.	95%	For SC signals the recognition rate was 65% and 70% for ECG.
[24] Hernan F. Garcia et al. [2016]	DEAP	Gaussian process latent variable models (GP-LVM): EEG, EMG, EOG, GSR, RSP, T, BVP	SVM	Dynamic affect recognition from multimodal physiological signals, which is based on learning a latent space using Gaussian process latent variable models (GP-LVM), which maps high dimensional data (multimodal physiological signals) into a low dimensional latent space. The dynamics are incorporated to the model by learning the latent representation, with associated dynamics.	89%	
[25] Gyanendra Kumar Verma et al [2014]	DEAP	EEG, Peripheral signals – EDA, GSR,EOG & SCR: Different Powers, STD and SE of detail and approximation coefficients	SVM	Multimodal fusion approach using Daubechies Wavelet Transform features for emotion recognition from physiological signals. DWT transform for multiresolution analysis of signal has been used. The 32 EEG channels are considered as independent modes and features from each channel are considered with equal importance	85.46%	RA is high only with EEG compared to when it mixed with peripheral signals
[26] Zhang, Y., Ji, X.et al, [2016]	DEAP	EEG- IMFS {intrinsic mode Function}: EMD	SVM	Emotion Recognition based on Empirical mode decomposition (EMD) and SampEn. The proposed method can recognize 4 categories of emotional states, such as HAHV, LAHV, LALV, and HALV, on the DEAP benchmark database. EMD method has been employed to decompose EEG signals only containing channels F3 and C4. A series of IMFs obtained by EMD are used to calculate SampEn values and to	93.2%	The classification performance is sensitive to the length of the segment, when the length of the segment is smaller, then the number of segments divided is more with this proposed method. It results in greater computational cost.

				form Feature vectors. These vectors are fed into SVM classifier.		
[27]: Lamti, H. A.,et. al 2016	40 healthy user	EEG: mean power P_m & Root mean square RMS (Welch and Wavelets)	SVM	The features extracted from band wave signals are the Power mean (Pm) and the Root mean square (RMS). Two different techniques have been adopted: the Welch method and wavelets. With the huge quantity of data in the extracted features due to the crossing of many parameters (band wave per time unit per sensor per subject) a selection phase is needed. For this purpose, Principal Component Analysis (PCA) and Genetic Algorithm (GA) were applied. SVM is used in classification phase. The system was tested on a virtual environment and assessed on four parameters: obstacles hit, navigation path deviation, execution time and outbound POG. For the first and second parameters, the introduction of emotions is efficient as it decreases the number of obstacles hit and the deviation rate from the optimal trajectory.	88%	The drawback is that for the third parameter, which the outbound POG are. Although the overall number decreases between the first and second trials, it still shows dependency on the scenario. When the scenario becomes more difficult, the number of outbound points of gaze becomes higher.

VI. Conclusion

The affective computing is a developing field which incorporates different domain areas like engineering, psychology, sociology. The solutions to this problems are mainly based on machine learning techniques. The existing database consists of limited affect model, contain few subjects or few samples for certain emotion categories. The applications with social relevance are wide which ranges from medical applications, security and behavior analysis. The main steps in development of an automatic affect recognition system consists of data acquisition, pre-processing, feature extraction, classification and all these modules contribute to improve the efficiency of the system.

References

[1] S. D'Mello, S. Craig, B. Gholson, S. Franklin, R. Picard, and A. Graesser, "Integrating Affect Sensors in an Intelligent Tutoring System,", Proc. Computer in the Affective Loop Workshop at Int'l Conf. Intelligent User Interfaces, pp. 7-13, 2005.

[2] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion", Journal of personality and social psychology, vol. 17, no. 2, p. 124, 1971.

[3] A. Russell, "A circumplex model of affect", Journal of Personality and Social Psychology, vol. 39, no. 6, pp. 1161–1178, 1980.

[4] R. Picard, E. Vyzas, and J. Healey, "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 10, pp. 1175-1191, Oct. 2001.

[5] C. Lisetti and F. Nasoz, "Using Noninvasive Wearable Computers to Recognize Human Emotions from Physiological Signals", EURASIP J. Applied Signal Processing, vol. 2004, pp. 1672-1687, 2004.

[6] A. Choi and W. Woo, "Physiological Sensing and Feature Extraction for Emotion Recognition by Exploiting Acupuncture Spots", Proc. First Int'l Conf. Affective Computing and Intelligent Interaction, pp. 590-597, 2005.

[7] Biswas, S., "An Efficient Expression Recognition Method using Contourlet Transform", Int. Conf. Percept. Mach. Intell. pp. 167–174, 2015

[8] Ji, Y., Idrissi, K., "Automatic facial expression recognition based on spatiotemporal descriptors", Pattern Recognit. Lett. 33, 1373–1380, 2012.

[9] Mahersia, H., Hamrouni, K., "Using multiple steerable fi lters and Bayesian regularization for facial expression recognition", Eng. Appl. Artif. Intell. 38, 190– 202, 2015.

[10] Bashyal, S., Venayagamoorthy, G.K.V., "Recognition of facial expressions using Gabor wavelets and learning vector quantization", Eng. Appl. Artif. Intell. 21, 1056–1064, 2008.

[11] Demir, Y., "A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering", Neural Comput. Appl. 27, 131–142, 2014.

[12] Zhang, L., Member, S., Tjondronegoro, D., "Facial expression recognition using facial movement features", IEEE Trans. Affect. Comput. 2, 219–229.

[13] Hegde, G.P., Seetha, M., Hegde, N., "Kernel locality preserving symmetrical weighted fisher discriminant analysis based subspace approach for expression recognition", Eng. Sci. Technol. Int. J. 19, 1321–1333, 2016.

[14] Poursaberi, A., Noubari, H.A., Gavrilova, M., Yanushkevich, S.N., "Gauss – Laguerre wavelet textural feature fusion with geometrical information for facial expression identification", EURASIP J. Image Video Process., 1–13, 2012.

[15] Hernandez-matamoros, A., Bonarini, A., Escamilla-hernandez, E., Nakano-miyatake, M., "A Facial Expression Recognition with Automatic Segmentation of Face Regions", Int. Conf. Intell. Softw. Methodol. Tools, Tech. 529–540, 2015.

[16] Hernandez-matamoros, A., Bonarini, A., Escamilla-hernandez, E., Nakano-miyatake, M., "A Facial Expression Recognition with Automatic Segmentation of Face Regions", Int. Conf. Intell. Softw. Methodol. Tools, Tech. 529–540, 2015.

[17] Happy, S. L., & Routray, A., "Automatic facial expression recognition using features of salient facial patches", IEEE Transactions on Affective Computing, 6(1), 1–12, 2015.

[18] Hegde, G. P., Seetha, M., & Hegde, N., "Kernel Locality Preserving Symmetrical Weighted Fisher Discriminant Analysis based subspace approach for expression recognition", Engineering Science and Technology, an International Journal, 19(3), 1321–1333, 2016.

[19] Michael Revina, I., & Sam Emmanuel, W. R., "Face expression recognition using LDN and Dominant Gradient Local Ternary Pattern descriptors", Journal of King Saud University - Computer and Information Sciences, 2018.

[20] Nigam, S., Singh, R., & Misra, A. K., "Efficient facial expression recognition using histogram of oriented gradients in wavelet domain", Multimedia Tools and Applications, 2018.

[21] Zhang, J.; Chen, M.; Hu, S.; Cao, Y.; Kozma, R., "PNN for EEG-based Emotion Recognition", In Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 9–12, pp. 2319–2323, 2016.

[22] Mirmohamadsadeghi, L.; Yazdani, A.; Vesin, J.M., "Using cardio-respiratory signals to recognize emotions elicited by watching music video clips", In Proceedings of the 2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP), Montreal, QC, Canada, 21–23, pp. 1–5, 2016.

[23] Zied, G.; Lachiri, Z.; Maaoui, C., "Emotion recognition from physiological signals using fusion of wavelet based features", In Proceedings of the International Conference on Modelling, Identification and Control, Sousse, Tunisia, 18–20,2015.

[24] Garcia, H.F.; Alvarez, M.A.; Orozco, A.A., "Gaussian process dynamical models for multimodal affect recognition", In Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society, Orlando, FL, USA, 16–20, pp. 850–853, 2016.

[25] Verma, G.K.; Tiwary, U.S., "Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals", Neuroimage, 162–172, 2013.

[26] Zhang, Y.; Ji, X.M.; Zhang, S.H., "An approach to EEG-based emotion recognition using combined feature extraction method", Neurosci. Lett., 633, 152–157, 2016.

[27] Lamti, H.A.; Ben Khelifa, M.M.; Alimi, A.M.; Gorce, P., "Emotion detection for wheelchair navigation enhancement", Robotica, 34, 1209–1226, 2016.