

A Survey on Human Emotion Recognition Approaches, Databases and Applications

C.Vinola*, K.Vimaladevi†

* *Department of Computer Science and Engineering, Francis Xavier Engineering College, Tirunelveli, Tamilnadu, India*

† *Department of Computer Science and Engineering, P.S.R Engineering College, Sivakasi, Tamilnadu, India*

Received 7th Aug 2015; accepted 30th Nov 2015

Abstract

This paper presents the various emotion classification and recognition systems which implement methods aiming at improving Human Machine Interaction. The modalities and approaches used for affect detection vary and contribute to accuracy and efficacy in detecting emotions of human beings. This paper discovers them in a comparison and descriptive manner. Various applications that use the methodologies in different contexts to address the challenges in real time are discussed. This survey also describes the databases that can be used as standard data sets in the process of emotion identification. Thus an integrated discussion of methods, databases used and applications pertaining to the emerging field of Affective Computing (AC) is done and surveyed.

Key Words: Emotion recognition, Affective Computing, Artificial Intelligence, Human Computer Interaction

1 Introduction

Artificial Intelligence (AI) [1] is a popular field of research involving computational intelligence. Emotion recognition or Affective Computing (AC) being the AI related area imparts intelligence to computers in recognizing human emotions.

With the advances in the field of emotion recognition, increasing number of investigations have been done and they have found a significant place in different applications involving human computer communication. The primary objective of an emotion recognition system is to interpret the input signals from different modalities and use them to convey information about the interpreted emotion. For example, the doctor may act based on patient's emotions recognized.

The nonverbal communication is an important consideration by all these systems. The significant facet of such systems is to study the mechanism behind the non-verbal communication by human beings to be taught in the computer so that the applications can interpret and connect with user's emotions.

Previously, publications on surveys have been published in this area where Zeng et al. surveyed the multimodal recognition methods including audio and visual input signals with a focus on spontaneous

Correspondence to: selvivino@gmail.com

Recommended for acceptance by Angel Sappa

<http://dx.doi.org/10.5565/rev/elcvia.795>

ELCVIA ISSN: 1577-5097

Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

expressions [6]. Rafael A. Galvo and Sidney reviewed the affect detection models with the methods and applications in an interdisciplinary approach [7].

This survey is structured to provide a better insight into the recent approaches along with the databases used and a brief outline of the application areas in which these systems could be successfully implemented and used.

2 Emotions and affective computing

2.1 Theories of Emotion

The work conducted by researchers in the recent decades was based on various theories of emotion stated by philosophers and psychologists [2]. Charles Darwin proposed the theory of evolution and insisted that emotional expressions must be understood in terms of survival value. Walter B. Cannon along with Philip Bard developed the Cannon-Bard Theory believing that emotions come first and then the bodily reactions. James-Lange Theory proposed by William James and Carl Lange states that an individual has a physiological response and then experiences an emotion.

Joseph E. LeDoux states that the physiological responses are hardwired in the brain and conscious emotions are irrelevant to them. Robert Plutchik believed that there were 8 classes of emotions namely anger, fear, sadness, disgust, surprise, anticipation, acceptance and joy and states that that all other emotions evolved from them. Paul Ekman, a psychologist studied the relation between internal emotions, facial expressions and body language. Ekman invented the Facial Action Coding System (FACS) to categorize facial expressions. Ekman addressed the issues to improve the emotional life of the people [5] including cross-cultural differences.

With the theoretical perspectives of emotions, researchers have attempted to contribute various approaches to emotion recognition including vision based methodologies.

2.2 Affective Computing

Being a multidisciplinary field involving computer science, psychology and cognitive science, AC has instigated the computer scientists in building up emotion detecting models to promote human-machine interfaces. In the book written by Picard [3] the reasons are stated for why the computers should feel and identify the emotion as humans can. According to him there are four states of a computer with respect to affective computing 1. Computers that cannot express and that cannot perceive 2. Computers that can express but not perceive 3. Computers that cannot express but perceive 4. Computers that can express and perceive.

Picard also has articulated the research challenges in affective computing responding to the criticism raised by the scientists [4]. Challenges reported are related to sensing and recognizing emotions, affect modeling and expression of emotion. Amidst the challenges in the domain of AC, the scientists have tried to explore the possible ventures to reach the ultimate goal in developing an affect receptive interface to model an effective emotion detection system.

Literatures on affective computing [8], [9], [10], [11] emphasize the researcher approaches and future enhancements that can be done to improve Human Computer Interaction (HCI).

3 Emotion recognition modalities and approaches

Emotions in humans are expressed in different modes which gives a lead to the development of affect recognition systems. Different modes, individually or in various combinations contribute to diverse approaches for emotion recognition. A possible emotion recognition framework is depicted in the figure 1. Features of the selected mode after extraction and processing is used for deciding upon the emotion by applying suitable classification method. Final decision is done by the fusion of different outcomes.

AC scientists have adopted these approaches and attempted to prove their efficacy. Each modality has its significance, which is explored in the systems developed. This section discusses the modality used in those systems with the focus on achievements.

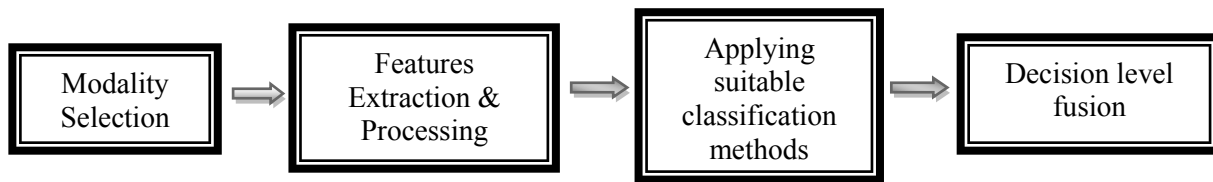


Figure 1: Emotion recognition framework

3.1 Facial Expressions

Facial Expressions are widely used and accepted mode in emotion detection as it is activated immediately by the experienced affect. Ekman and Friesen developed FACS where in the facial movements are described by the action units (AUs) [12]. Provoked by the system developed, many researchers have incorporated facial expressions in their systems as one of the very successful modes in understanding one's mental state. In the recent days, many different approaches for facial expression detection and recognition have been proposed. Automatic facial expression recognition methodology using neural classifiers has successfully brought in good results [13], [14]. Systems were trained with many different images including variety of facial gestures for improvement of testing accuracy. Neural classifiers involves computational load in recognition of emotion.

Appearance and shape features extraction for facial expression recognition is proposed by the researchers [15] wherein the decision fusion is done and the emotion is detected. Local descriptors are being used for the first time to enhance the performance. Dynamic facial actions were not taken into account in the work. Approaches involving facial elements and muscle movements [16] representing dynamic features have removed the limitations caused by the approaches using static features leading to higher CRR (Correct Recognition Rate). Processing time is effectively reduced in this approach comparatively but again it is not a video based approach involving multi frames which is mandatory for real time applications.

Dynamic texture recognition (combining appearance and motion features) using Local Binary Patterns [17] which is simple and robust in terms of grayscale and rotation variations, makes it very capable for real time problems. Ramchand et al. have proposed an emotion classification method based on Local binary patterns in person dependant and person independent mode in which person dependant mode produced more accurate results [18]. The degree of templates to be formed in the person dependant mode is more and it increases with the number of persons to be considered.

Video based approaches have emerged amidst the static image based approaches with a focus on developing systems to work in real time. Facial action parameters independent of the head pose in continuous videos are used in these methods. Dornaika [19], in contrast to the other approaches involving feature extraction and then deciding the expression introduces a new approach allowing the simultaneous retrieval of facial actions and expression using a particle filter adopting multi-class dynamics. In the first stage, the 3D head pose is retrieved using a deterministic registration technique based. In the second stage, the facial actions as well as the facial expression are simultaneously retrieved using a stochastic framework based on second-order Markov chains.

In [20], dynamic warping technique used ensures that the changes in the video rate or facial action duration do not influence the accuracy of recognition. Dynamics in the videos are used to explore the emotional state of a person by extracting local video patches in [21]. Manual Intervention for annotation of facial landmarks is avoided with no requirements of the preprocessing steps which increases the speed of the system. Different research works with their approaches and recognition rate achieved is mentioned in Table 1.

Systems that work on spontaneous databases have materialized to suit the real time environments. A robust emotion recognition system working in challenging situations like illumination and posture variations using a method termed ESL(Extreme Sparse Learning) is proposed by Seyedehsamaneh et al. [22]. But, as there are parameters to be optimized in this method the computational cost is the constraint.

Table 1: Emotion Recognition Using Facial Expressions

| Work Done by | Approach | | | |
|-----------------------|---|--|--|--------------------------|
| | Features used | Classification method | Database & Recognition rate | No of Classes Recognized |
| Zisheg LI [15] | I. Appearance II. Shape | I. Component based Bag of Words Method II. PHOG (Pyramid Histogram of Oriented Histogram) descriptors | C-K database, 96.33% | 6 |
| Ligang Zhang [16] | Patch based 3D Gabor features | Adaboost, SVM (Support Vector Machines) | JAFFE, 92.93% C-K database, 94.48% | 6 |
| Ramchand Hablani [18] | Local Binary patterns | Template Matching I. Person dependant II. Person independent | JAFFE, I. 94.44% II. 73.61% | 6 |
| Fadi Dornaika [20] | 3D Tracker based Facial actions | Dynamic Time Warping, PCA (Principal Component Analysis) + LDA (Linear Discriminant Analysis) | CMU database, Above 90% | 6 |
| Munawar Hayat [21] | Local video patches | Clustering algorithms, SVM | BU4DFE 94.34% | 6 |
| Seyedehsamaneh [22] | Pose-invariant features based on Optical Flow | I. Extreme Sparse Learning II. Kernel extreme Sparse Learning | ECK+ Database I. 90.36 II. 92.74 | 6 |

Facial gestures always play a prominent role in identifying the emotion of a person, acting as the base of an emotion recognition system with other modalities used to promote the accuracy of the system. Systems developed aim at increasing the accuracy but at reduced computational load.

3.2 Bodily expressions

In addition to the facial expressions, the body gestures provide an effective input on affect of a person. Many AC researches have attempted to use them along with the facial expressions to improve the recognition capacity of the system. Body language matches with the inner emotional state, thus helping in the emotion classification. Survey publications are available including discussions about the importance of the body gestures and the comparison of the methods used [23]. Table 2 mentions the research works done using body gestures as an AC mode.

Table 2: Emotion Recognition Using Bodily Expressions

| Work Done by | Approach | | | |
|-------------------------|--|--|---|--------------------------|
| | Features used | Classification method | Database & Recognition rate | No of Classes Recognized |
| Ginevra Castellano [24] | Body- QOM (Quantity of Motion) | I. DTW (Dynamic Time Warping) II. Meta-Feature Approach | -- | 4 |
| Hatice Gunes [25] | Face & Body- Displacement features | I. Feature Level fusion II. Decision Level fusion | I. 96% II. > 80% | 7 |
| Tadas Baltrušaitis [26] | Face&Upper Body- 22 Feature points | I. HMM (Hidden Markov Model) II. (DBN) Dynamic Bayesian Network | -- | 5 |
| Hatice Gunes [27] | Face & Body- Temporally segmented features | I. Video based II. Frame based | FABO, I. 85% II 82.6% | 12 |
| Shizhi Chen [28] | Facial Land mark points, HOG descriptors, Motion Information | I. Temporal Normalization II. Bag Of Words SVM with RBF Kernel | FABO Database, Improved recognition rate compared to the state of art. | 10 |

As an individual modality, body gestures and dynamics have been used in detecting four classes of emotions namely based on high, low arousal, positive and negative emotions from movement expressivity by Ginevra Castellano et al. [24]. The method was novel of its kind but only small data sets are used to test the system. Gunes and Piccardi have attempted to perform a feature level fusion (early) and decision level fusion (late) of the face and body modalities[25]. Performance evaluation showed that bimodal fusion outperformed the classification done using the facial modality alone. In [26], the system developed inferred emotions from upper body gestures in addition to facial expressions, including head and shoulder motions. A multilevel DBN (Dynamic Bayesian Network) models the emotional state depending on the probabilities of the gestures. Without changing the main architecture of the system it is possible to improve any module or add features to it but the limitation is properties like intensity, offset, and onset of the emotional state are not considered which if included could improve the performance further.

Research shows that dynamics of the modalities are important in the affective body interpretation and therefore methods considering the temporal segments are proposed [27]. System is able to process bimodal data and their temporal segments and fuse them at either the feature level or the decision level but the limitations are that it does not deal with missing data/multiple expressions in a single video sequence and also assumes that the background does not change which is less possible in real time.

Shizhi Chen et al. combined the local motion and appearance feature in a novel framework to model the dynamics of the face and body gesture [28]. Improvement in the recognition accuracy compared to the frame based approach is achieved. It outperformed the method already proposed in [26] in accuracy and also the latter involved complex features in the process. It also proved that the bimodality improved the system's outcome.

3.3 Speech Signals

Verbal Communication aids in recognizing the affect of the communicating person effectively. Many parameters of the audio signals, including tone of the voice and pitch exhibit the emotional state of the person. Audio parameters are used as inputs to the emotion classification system which recognizes the emotions with reference to the parameters. Importance of audio based AC can be identified by the various works carried out by the scientists. Survey literature focusing only on speech emotion recognition has been written by Ayadi et al. with a comprehensive view of the features, classification and databases of the systems developed [29].

Acoustic characteristics of speech like prosody in different permutations are used in the research works of identifying emotions. Different approaches in Speech Emotion Recognition and their findings is listed in Table 3. Analysis of different features of prosody, spectral envelope and voice quality is done by Iker et al. Evaluation of the emotion identification methods is done in the work where features based on spectral envelope outperformed the prosodic ones [30].

An IFN (Iterative Feature Normalization) approach has been proposed in which the acoustic differences in the natural speech of the different speakers are reduced [31]. Due to normalization done in the technique the difference between the emotional classes is preserved. It is an attempt made to convey out a speaker independent system.

Convolutional Neural Networks have been used by Qirong Mao et al. [32] to learn affect salient features for Speech Emotion Recognition wherein there are two learning phases, simple features learnt in the lower layers and salient features learnt in higher layers and achievement is above 60% recognition rate amidst noise and channel distortion. Results show superior performance with respect to speaker, language variation and environment distortion. Restriction is that it could work well in laboratory settings but extension is required to suit the real time environment.

[33] suggests a multiview supervised dictionary learning approach where two techniques are proposed and application to speech emotion recognition using two feature sets are investigated. Correlation coefficient as the evaluation measure in predicting the continuous dimensional affects, the proposed methodology achieves the highest performance among the other methods.

In [34] it is proposed that Fourier parameter features are effective in identifying the different emotions in speech with a focus in developing speaker independent emotion recognition. Salient features from emotional speech signals are extracted and validated and hence proved that the FP features are effective in characterizing and recognizing emotions in speech signals.

Audio features along with facial features and body gestures increased the recognition accuracy which is demonstrated by recent researches. Emotion recognition based on gesture and speech information has been proposed in [35], which uses decision level fusion to determine the emotion. The proposed method is able to be extended to use other modalities and helpful in automatic emotion recognition system.

Another bimodal approach is attempted by Kun LU and Yunde JIA using Boltzmann zippers which outperformed single modal models [36]. It is again one another person independent, induced prototypical emotional states recognition technique which proves that Boltzmann zippers outperforms coupled HMM because it takes into account the interior relationship between visual and audio modalities.

Table 3: Emotion Recognition Using Speech Signals

| Work Done | Approach | | | |
|----------------------|---|--|--|--------------------------|
| | Features used | Classification method | Database & Recognition rate | No of Classes Recognized |
| C.B.Marioor yad [31] | Acoustic features | I. Iterative Feature ormalization II. SVM | IEMOCAP Corpus, Improved accuracy when compared to global normalization | -- |
| Q.Mao [32] | I. LIF (Local Invariant Features) II. Affect-salient discriminative features | Convolutional neural network | I. SAVEE database, 73.56% II. Emo-DB,85.2% III. DES,79.9% IV. MES,78.3% (Speaker independent) | -- |
| M.J.Gangeh [33] | Sparse Coefficients | Multiview Supervised Dictionary Learning | AVEC 2012,-- | -- |
| K.Wang [34] | Fourier Parameters | SVM | I. EMOADB II. CASIA, >90% | 6 |
| K.A.Vu [35] | Gestures & Speech | Decision level fusion(Weight Criterion, Majority Vote) | Own Database, 76.28 & 85.39% | 4 |
| K.Lu [36] | Facial features & Speech (Prosodic, Spectral) | Boltzmann Zippers | Own database, >80% | 7 |

3.4 Physiological signals

Other than audiovisual signals there are other modes by which the inner emotional state of a person can be identified. Taking this into account, recent researches have explored the significance of such signals, especially physiological signals which are directly or indirectly related to the emotional state of a person. Different recent approaches have been mentioned in Table 4.

Study of emotion recognition based on physiological signals like heart rate variability parameters and respiration frequency is attempted and suggested that it could be combined with other parameters like facial expressions to detect emotions [37]. The accuracy obtained is not appreciable and it would increase if other important physiological parameters like galvanic skin conductance and electrocardiogram signals.

Table 4: Emotion Recognition Using Physiological signals

| Work Done by | Approach | | | |
|--------------------|---|--|--|--------------------------|
| | Features used | Classification method | Database & Recognition rate | No of Classes Recognized |
| J.Kortelainen [37] | Heart rate variability, Respiration frequency and Facial Features | KNN (K nearest neighbors) | CK+ for facial expressions, 54.5% with arousal, 38% with valence | 2 |
| V.Rozgic [38] | Segment level EEG features | I. NBNN (Naïve Bayesian Nearest Neighbor) II. NN-Voting III. RBF-SVM (Radial Basis Function-Support Vector Machines) | DEAP, Valence, Dominance, Arousal, Liking(>60%) | 4 |
| D.Lakens [39] | OXY, GSR & HR | Random Forests Classifier | Own database, 74% | 5 |
| Wanhui Wen [40] | Biochemical features of Breath Gas | Artificial Neural Network & SVM | Own database, 70% | 2 |
| J.Hernandez [41] | GSR, Skin Temperature | Mirror Interfacing | Own database,-- | -- |

EEG (Electro Encephalo Gram) signals are effectively used by scientists in different ways to identify the emotions of a person [38]. In order to overcome the limitation of loss of information when the features are highly non-stationary, there is a transformation of multiple segment-level feature vectors to a single response level feature vector in a manner that retains information relevant for classification task. The work suggests an effective modality to recognize the affect.

Selection of physiological signals which are immediately linked to the internal state of a person like Fingertip blood oxygen saturation (OXY), Galvanic Skin Response (GSR) and heart rate [39] are used by scientists in the field of AC. The detailed affective data recording of the subjects who are induced emotions by watching a film and the multi-variant correlation analysis in the proposed methodology have helped them to achieve better results when compared to previous works.

Exploration on emotion recognition using breath gas information is attempted by Kazuhiko Takahashi[40]. Sensing devices used for experimenting on the breath gas may be a overhead but it again gives an idea of another modality to the researchers of the field which could categorize the emotions as positive and negative.

[41] offers a novel sensor system to combine physiological data and visual context to capture and reflect on their daily activities. For instance, if a person can check his daily activities, he can increase his memory and gain more attentiveness of his time allotment. An user interface for daily logging is suggested in the paper which could open up different applications in the field.

3.5 Other Modalities

Many research works in AC involve individual or different combinations of modalities with an aim of producing an emotion detection system. Each modality is directly or indirectly related to the affect of an individual which in different ways can be used for the affect recognition.

The contribution of the basic and often attempted modalities and the outcomes of the systems have been already discussed. Other distinct modalities and their significance in designing an affect sensitive module are suggested by several researchers which are listed in Table 5.

Table 5: Emotion Recognition using other modalities

| Work Done By | Approach | | | |
|------------------|--|--|---|--------------------------|
| | Features used | Classification method | Database & Recognition rate | No of Classes Recognized |
| P.Ren [42] | Pupil Diameter Signal | K*, MLP (Multi layer Perceptron), Naïve Bayes classifier, Random Forests | Own Database, 83.16% | 2 |
| G.Caridakis [43] | Facial features, Body gestures and Speech parameters | Bayesian Classifier | Own database, >70% | 8 |
| K.Tang [44] | 2D,3D features, speech features | CCA (Canonical Correlation Analysis) | Data from Kinect Sensor, >70% | 7 |
| M.Soleymani [45] | EEG features, Pupillary response & gaze distance | SVM with RBF Kernel | Own database, 68.5% for Valence 76.4% for Arousal | 6 |
| W.Zheng [46] | EEG features, Pupillary response | SVM Feature Level Fusion Decision level Fusion | Own database, 73.59% 72.98% | 3 |
| J.Wagner [47] | Vocal Features, Facial Expressions and Gestures | Naïve Bayesian Classification | CALLAS Expressivity Corpus,-- | 4 |

The eye being a potent agent of expressing emotions, is utilized in detecting the affective state of a soul through the eye tracking information and variation in pupil diameter(PD) in [42]. Previously, detection of the

emotional state of a user using PD had not been completely investigated. This was successfully attempted in the work as there is a transition from “relaxation” to “stress” mood. It also included GSR as another modality in the evaluation but individually it is able to yield only 65.79% but is able to produce 83.16% when used along with PD.

Multimodal Emotion Recognition (MER) systems have comparatively outperformed other systems. Feature level and Decision level fusion of the multimodal features like facial expressions, gestures of the body and speech parameters are done in [43]. Even when some of the feature values of any particular modality is missing the other modalities could favor the system in such a multimodal scheme.

In [44] a novel MER system is proposed wherein there are two parts, a new database comprising of 2D facial images, 3D facial feature points and concurrent time based audio waves and recognition part including a new fusion level combination. In the new fusion methodology, decision level fusion and feature level fusion results are combined considering feature level fusion result as another feature which is a different thought and it yielded good results.

[45] introduces a method to recover affective responses to videos using EEG signals, pupillary response and gaze distance. It is an attempt to design a precise and user-independent classification model to identify emotions. The work also tries to evaluate and also proves to an extent that the non-verbal affective cues could possibly replace the affective self.

Another fusion model is built based on Eye tracking information and emotion relevant features from EEG signals to develop a multimode recognition system [46]. Different features contributing to the Pupil diameter and EEG signals are effectively used in bringing out good results. Handling all the features may lead to computational overhead and including other modalities like facial expressions may improve the accuracy further.

Research on the fusion of the multi modalities is also well carried out as in [47]. This paper concentrates on another dimension i.e., missing data in any modality. Taking the vocal, facial and gestural behavior into consideration the author evaluates in a broader way the different fusion methodologies and their suitability in finding the missing data. Discussion paves way for researchers to think in a different aspect.

4 Emotion recognition databases

As a result of wide-ranging research in emotion identification, multiple databases have emerged covering different modalities including image, video, audio and physiological data sets. Databases used for different works are cited in the tables presented above. In this section a brief outline on the various databases is provided.

4.1 Image and Video Databases

Various Image and video based databases are publicly available for the facial expression analysis and human gesture analysis and validation to detect emotions.

- **C-K (Cohn-Kanade) database [48]:**

Being a video based action unit coded face database which involves posed expressions, it includes 486 sequences from 97 posers. Beginning with a neutral expression, the sequence progresses to an extreme expression. The peak expression is fully FACS coded. Number of Subjects included is 100. Released in the year 2000, for the cause of advancing research into detecting individual facial expressions, It has become one of the most commonly used data sets. But the drawbacks found were that emotion labels were not validated and non-availability of common performance measure in order to compare with standard algorithms.

- **C-K + (Extended Cohn-Kanade) database [49]:**

To address the issues in CK database, extension has been done. It lets in both posed and spontaneous manifestations. For posed expressions, the number of sequences is increased from the initial database by 22% and the number of subjects is enhanced by 27%. The target expression for each sequence is again fully FACS

coded. Number of Subjects included is 123. It serves as a very valuable addition to the already existing data set.

- **JAFFE (Japanese Female Facial Expression) database [50]:**

The database contains 213 images of 7 facial expressions posed by 10 Japanese female models. Each image has been rated on 6 emotion classes on 60 Japanese subjects. As it involves multiple subjects exhibiting multiple emotions it has been extensively used in the field of research.

- **BU-3DFE (Binghamton University-3D Dynamic Facial Expression) database [51]:**

3D facial models have been extensively employed in research. This database includes 100 subjects with 2500 facial expression models. With the exception of the neutral expression, each of the six prototypic expressions namely happiness, disgust, fear, angry, surprise and sadness are included with four levels of intensity. It considered different cultural and race condition subjects in different ages allowing way for research in different dimensions.

- **BU-4DFE (BU-3DFE+time) [52]:**

A 3D dynamic facial expression database which is available with six model sequences showing six facial expressions of the BU-3DFE database. Each sequence of expression contains nearly about 100 frames. The database comprises 606 3D facial expression sequences captured from 101 subjects. Including dynamic characteristics in the ancestor database, it can replace other databases when involving application in real time.

- **MMI [53]:**

A database which consists of over 2900 videos and high-resolution still images of 75 subjects, partially coded for frame-level, indicating for each frame whether an AU is in either the neutral, onset, apex or offset phase. It also includes audio-visual laughers. In particular, it contains recordings of the sequential pattern of a facial expressions, from Neutral, through a progression of onset, apex, and offset phases and back again to a neutral face.

Other databases focusing on expressions of the six basic emotions, the MMI Facial Expression Database contains both these classical expressions and expressions with a single FACS Action Unit (AU) activated, for all existing AUs including naturalistic recordings also. It is readily available for research community.

- **NVIE (Natural Visible and Infrared facial Expressions) [54]:**

Dataset involves visible and infrared imaging of posed & spontaneous expressions of 215 students. The database includes basic expressions. It was the first of its kind involving both visible and infrared videos. The advantage of the database is mainly that posed, spontaneous expressions with and without glasses are included thus supporting researchers in further algorithm comparison and evaluation

- **Affective-MIT Facial Expression [55]:**

Dataset includes 242 facial videos with 168,359 frames recorded in real world conditions. Frame- by-frame labels are available for 10 symmetrical FACS action units, 4 asymmetric (unilateral) FACS action units, 2 head motions, smile, general expressiveness and gender. The videos in this dataset were recorded in real-world conditions. They exhibit non-uniform frame rate and non-uniform lighting. The camera position relative the viewer varies from video to video.

The videos contain viewers from a range of ages and ethnicities some with glasses and facial hair. It is particularly suggested for researchers working in the domains of facial expression recognition, affective computing, psychology and marketing as the responses from the video stimuli are recorded.

- **DISFA (Denver intensity of spontaneous facial action database) [56]:**

Video based database involving spontaneous expressions of 27 participants. Intensity of 12 action units is coded in 1,30,000 frames. Most previous databases for automated facial expression analysis are limited to posed facial expressions and holistic labels. Facial behavior is thoroughly annotated using anatomically based descriptions, FACS AUs.

The intensity of each AU is annotated on a six-point intensity scale. DISFA thus provides continuous annotation of graded changes in spontaneous facial expression of emotion. It will help researchers to develop and assess new methods for spontaneous facial expression recognition.

- **LIRIS-ACCEDE [57]:**

Contains 9,800 video excerpts which are selected from 160 feature films and short films. It is the largest video database currently in existence interpreted by an extensive population using induced emotional labels. In order to address the issue of lack of large databases, this database is developed with an aim to promote sharing and usage among research community.

- **FABO (Bimodal Face and Body Gesture) [58]:**

The FABO database contains videos of face and body expressions recorded by the face and body cameras, simultaneously. Posed visual data gathered from volunteers in a science lab setting by asking and directing the participants along the mandatory activities and operations. This database is the initiative to combine facial and body displays in a truly bimodal manner, thus enabling significant future advances in affective computing research.

- **Kinect FaceDB [59]:**

Dataset consisting of the multimodal facial images of 52 persons acquired by Kinect. The data is captured in different time periods involving 9 different facial expressions under different conditions. It is the first publicly available face database based on the Kinect sensor. The database consists of different data modalities and multiple facial variations. The aim is to create database of images of different facial expressions in different lighting and occlusion conditions to serve various research purposes.

Thus the image and video databases have been developed and freely accessible with much use by the research community.

4.2 Speech Databases

Research in Audio signals which lead to the decision in emotion recognition require standard databases. Some of the often used speech emotion databases are discussed in this section.

- **Belfast Naturalistic [60]:**

It consists of audiovisual clips from 125 speakers of both genders. Emotional clippings are events that help in interpreting a peak in the emotion display and to demonstrate how it develops over time. To cover a wide range of emotions. It aims to develop the first large audiovisual naturalistic database of emotion and also to cover a wide range of emotions.

- **Biwi 3D Audiovisual Corpus [61]:**

It contains dynamic 3D scans of faces recorded while uttering a set of English sentences. The information is read by tracking the frames using a basic face template, dividing the speech signal into phonemes, and evaluate the emotions by an online survey. Database stands out from all available datasets during its time, which are either completely posed or missing 3-D information.

Applications include audio-visual emotion recognition, emotion-independent lip reading, or view-independent facial expression recognition.

- **SAVEE (Surrey Audio-Visual Expressed Emotion) [62]:**

SAVEE database has been recorded as an important requirement for the development of an automatic emotion recognition system. The database consists of recordings from 4 male actors in 7 different emotions, 480 British English utterances in total. Database contains emotional speech utterances covering seven emotions including anger, disgust, fear, happiness, sadness, surprise, and neutral deliberately displayed by four English speakers.

- **IEMOCAP (Interactive Emotional Dyadic Motion Capture) Corpus: [63]**

To facilitate the investigations involving the combination of visual and audio features for emotion recognition the database was developed. Ten trained actors are engaged in five dyadic interactions using scripts and spontaneous actions that are selected to bring out happiness, anger, sadness, and frustration. Naturalistic non-acted human interactions could be observed in the database.

The motion capture information, the interactive setting to bring forth emotions, and the size of the database make this data set a valuable addition to the existing databases in modeling of multimodal human communication.

- **EMODB [64]:**

It comprises 10 sentences that cover seven classes of emotion from everyday communication, namely, anger, fear, happiness, sadness, disgust, boredom and neutral. They could be rendered in all emotional contexts without semantic inconsistency. Database serves as a basis of numerous studies in the field.

- **DES (Danish Emotional Speech database) [65]:**

Includes emotional speech utterances covering five emotions (i.e., anger, joy, surprise, sadness and neutral) displayed by four Danish actors. The work also insisted that it is difficult to identify real time emotions from Danish speech. The sample rate is 48 kHz.

- **MASC (Mandarin Affective Speech Corpus) [66]:**

The database contains recordings of 68 native speakers (23 female and 45 male) and five kinds of emotional states: neutral, anger, elation, panic and sadness. Each speaker pronounces 5 phrases, 10 sentences for three times for each emotional states and 2 paragraphs only for neutral. This corpus is built for prosodic and linguistic analysis and detection of emotion expression in Mandarin. It can also be used for recognition of affectively stressed speakers.

- **SEMAINE [67]:**

Being an audiovisual database, it serves as a component of an approach in building Sensitive Artificial Listener (SAL) agents that employ the persons in emotionally colored conversation. Recordings of the total 150 participants with 959 conversations are done. Multiple types of gesture are present—facial movements, head nods and shakes, and laughs. It provides a resource that computational research can use immediately. Beyond that, it is natural to add new types of labeling to the recordings.

- **AVEC (Audio Visual Emotion recognition Challenge) [68]:**

A database which is given as challenge for researchers who use audio and video features to recognize emotions. AVEC 2012 challenges to recognize four continuously valued affective dimensions such as arousal, expectancy, power and valence. It provides a benchmark test data set for multimodal data processing and to decide on the merits of the two approaches individually and on fusion basis.

4.3 Multimodal databases:

- **CALLAS Expressivity Corpus [69]:**

Constructed within the European Integrated Project CALLAS and contains affective behavior, incorporating vocal utterances, facial expressions, and gesture expressivity in three primary emotion classes. It was developed with the objective of the research teams to apply the data set in feature extraction, multimodal analysis and synchronization and fusion techniques and, hopefully, will provide reference point for future attempts within the affective computing community.

- **MAHNOB-HCI(Multimodal Analysis of Human Nonverbal Behaviour) [70]:**

Consists of two experiments. The responses including EEG, eye gaze, audio recording, and facial expressions of few people were registered. Two experiments, one with the videos extracted from movies and the other one

with the tag agreement where the videos with and without the tag were used. The large collection of modalities recorded (multi camera video of face, head, speech, eye gaze, pupil size, ECG, GSR, respiration amplitude, and skin temperature) and the high synchronization accuracy between them makes this database a valuable contribution to the research group of emotion recognition.

- **DEAP (Database for Emotion Analysis using Physiological signals [71]):**

It holds the highest number of participants in publicly available databases for analysis of spontaneous emotions from physiological signals. It uses music videos as emotional stimuli. The database contains physiological signals of 32 participants (and frontal face video of 22 participants), where each participant watched and rated their emotional response to 40 music videos along the scales of arousal, valence, and dominance as well as their liking of and familiarity with the videos.

5 Emotion recognition Applications

As the approaches and databases recently used in the field of AC have been discussed in the previous sections, the present section focuses on the areas of where to apply them for the benefit of a particular community. Research in emotion recognition leads to diverse applications which is substantiated by research publications.

- **Education:**

Emotion Integrated e-learning architecture model is developed and machine's ability to read emotions from physiological signals is discussed by Shen et al [72]. The work suggests an affective model to promote the learning experience in real time learning environment. Noteworthy conclusions from the research included that engagement and confusion are the most important and frequently occurring emotions in learning. One of the limitations found is that only biophysical signals are used to predict learner's emotional state.

Much work has been carried out to measure the affective dimension of engagement of students in learning to improve the learning environment [73] [74] [75]. [73] presents an automated facial recognition approach to analyze student facial movements during tutoring using the Computer Expression Recognition Toolbox (CERT), which tracks well-defined facial movements. The predictive models highlighted relationships between facial expression and aspects of engagement, frustration, and learning. The work has great potential for educational data mining.

In [74], an attempt for the development of real-time automated recognition of engagement from students' facial expressions is again explored. The remarkable intuition was that teachers constantly assess the level of their students' engagement, and facial expressions play a vital role in such assessments. The results suggest that machine learning methods could be used to develop a real-time automatic engagement detector with comparable accuracy to that of human observers.

In [75], it is identified that emotion indicators (valence and arousal) that combined with on- and off-task variables to represent positive and negative states associated with student learning, as well as physical behaviors linked to emotional states. In more detailed manner, correlations between low-level observations (i.e., chair movement) and higher-level observations (on-off task behavior) and then between these higher-level observations and student learning and attitudes were analyzed and used. Additionally there is an evaluation done on the impact of the presence of gendered characters on student motivation and achievement within learning environments. The motive of the work is to amalgamate examination of expressing factors like facial expression, empathy as they influence learning, human interaction and relationship development.

- **Aid for the Disabled:**

AC system can effectively work as a supporting aid for the people with certain disorders like syndrome and autism and also elderly people [76] [77]. [76] covers the latest research concerning the measurement of physiological signals of children with autism, particularly for the study of changing emotions in various environments. Answers to important questions regarding autistic children's physiological activity are examined, and it is observed that within a non-social environment, physiological responses are the same between children with and without autism but different in environments with social contexts. It is insisted in

the work that the physiological signals can be used as a trustworthy indicator of emotions of children with autism and the the latest developments in wearable sensor technologies available is also discussed.

The model in [77] draws motivation from "emotional-indexing" approach to teach emotions to children with autism. An automated mind reading system that infers complex mental states from facial expressions in real time video and a reaction advisor that recommend appropriate reactions. This paper presents results in the recognition of six classes of complex mental states of the children. Results though promising has to be improved.

- **HCI / Robotics :**

To improve HCI is the prime goal of an emotion recognition system which leads to applications in many areas including robotics. [78] focuses on the evolution of a real time emotion recognition system for a humanoid robot which conveys the information about emotion to the individual in interaction through speech signals. Intelligent neural network-based upper and lower facial action analyzers to recognize respectively six upper and 11 lower facial Action Units is proposed. Then another feed forward neural network and linear and nonlinear SVMs are employed to respectively build the emotional facial behavior recognizers. The research could also be further extended to recognize wider categories of AUs and facial emotional behaviors such as interest and stress.

[92] describe a human-robot interaction application which refers to mimic the facial expressions of a person perceived by a robot's camera. AIBO(Artificial Intelligence Robot) robot is used for the application. The input to the system is a video stream capturing the user's face. Temporal recognition method classifies a given image in the video to one of the emotion classes by modeling the dynamics. Such works will stimulate the researchers to use their methodology of emotion detection in the field of robotics so as to promote automated systems.

- **Safety aid:**

Human Centered Transportation system by detecting emotions of a driver of a vehicle through facial expressions is proposed in [79],[91] wherein the automated safety system acts in the case of difference in driver's normal mental state. Large number of road accidents take place due to drowsiness, fatigue or bad mood of the driver. The system proposed in this paper[79] takes into account both the Facial Gesture tracking and Emotion recognition so that if there is any sign of less alertness of the driver, the car will switch to automatic mode. A novel fuzzy system is created, whose rules are being defined through analysis of Facial Gesture variations. The idea behind this paper is to detect Facial Gesture by detecting the motion of eyes & lips along with classification of different facial expressions into one of the four basic human emotions, viz. happy, anger, sad, and surprise.

[91] proposes real-time computer vision system for monitoring drowsy driver. It uses one remotely located charge coupled device (CCD) camera to acquire video of the driver's face. From the video, various computer vision algorithms are employed and in real time recognize the facial behaviors that closely relate to the driver's level of vigilance. The facial behaviors include rigid head movement (characterized by 3D face pose), non rigid facial muscular movement (characterized by facial expressions), and eye gaze movement. The system being tested in a simulating environment with different subjects is found to be robust, reliable, and accurate in characterizing facial behaviors.

- **Entertainment:**

Applications involving entertainment also use an emotion detection system in maintaining their engagement with the entertaining agent. Physiological signals and facial expression features are often used in such type of functions. To adapt the gaming level to the difficulty the participants face while playing, [80] suggests an emotional assessment method using EEG signals and peripheral signals. Fusion of physiological information with other modalities such as facial expressions, speech, and vocal signals would certainly improve the accuracy.

User Interface applications integrate speech emotion recognition system to supply an improved understanding of the emotional state of the user and respond accordingly [81]. The purpose of this work is to propose a system framework based on smart logic that can be used to develop an emotional speech-

recognition and speech-synthesis engine suitable for integration into handheld embedded devices. Future development work aims to extend the capabilities of the recognition system to a larger set of emotions, so that continuous emotion status may be recognized.

Music is an important means of deciding the mood. There is a broad application potential for music players that support this ability to direct mood. Such devices need to include unobtrusive affect measurements and go for music selection accordingly to be able to direct mood to a target state[93]. The user study illustrated in this paper validated the concept of the affective music player and the affective music player takes in a desired mood as input and then selects songs that direct toward that required mood. Experiments were conducted in a real-world office setting.

Applications of AC techniques are still emerging and finding its place in different ventures and gives scope for other researchers for their contributions.

6 Conclusion and Future Directions

Directions for research in the field of emotion recognition are numerous and have distinct objectives in achieving accurate and effective models.

Modalities used for emotion recognition are mentioned in the previous sections. It is proved in the publications that multiple modal systems outperform single modal systems. Work on different combinations of modalities along with the strategies followed for fusing the decisions based on the modalities is available in the literatures [82], [83]. More contributions are expected by the AC scientists in this direction.

Spontaneous expressions are targeted by the research community in place of the posed ones, in order to develop a system that could adapt to the real time environment rather than conventional lab setup [84]. Befitting to the demands of the applications which require real time based interface, research trend has found a shift with this motif. Design and development of spontaneous standard data sets that act as a benchmark for the research perspectives serve as an example of this trend. AC literature has many publications on posed expressions and few on spontaneous expressions, which gives open opportunities for other researchers to conceive in that way.

Gender, Age, and Cultural differences between individuals influence the emotional state of a person and thus can act as significant parameters in the system to be developed. Taking them into consideration, publications have found their place in the literature [85] [86] [87] [88]. Promising results are expected when those parameters are likewise assumed into account in detecting an emotion which is one of the possible avenues in which the work can be conducted.

The context in which other emotions are experienced is another important deliberation required by scientists. Context is based on the actual emotional inner state of the person and the environment that induces the current emotion. Context sensitive approaches for emotion recognition are attempted by the researchers [89] [90]. More contributions can also be expected in this direction.

Diverse techniques in the Emotion recognition approaches, Databases designed and in the applications used are promising and looking forward for more research to design an accurate emotion recognition system.

References

- [1] Stuart Russell, Peter Norvig, *Artificial Intelligence A modern approach*. Third Edition, 2010.
- [2] Randolph R.Cornelius, "*Theoretical perspectives of emotion*", ITRW on Speech and Emotion, <http://www.isca-speech.org/archive>
- [3] R. W. Picard, *Affective Computing*. The MIT Press, 1997
- [4] Rosalind W. Picard, "Affective computing: challenges," *Int'l. J. Human-Computer Studies* 59, pp. 55-64, 2003. DOI: 10.1016/S1071-5819(03)00052-1
- [5] Paul Ekman, *Emotions Revealed*. TIMES BOOKS, 2003.

- [6] Z. Zeng, M. Pantic, G. I. Roisman and T. S. Huang, "A survey of affect recognition methods: audio, visual and spontaneous expressions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, pp. 39-58, Jan. 2009. DOI: 10.1109/TPAMI.2008.52
- [7] Rafael A. Galvo and Sidney D'Mello, "Affect detection: an interdisciplinary review of models, methods and their applications," *IEEE Trans. Affective Computing*, vol. 1, pp. 18-37, Jan. - Jun. 2010. DOI: 10.1109/T-AFFC.2010.1
- [8] Jimmy Or, *Affective Computing Focus on Emotion Expression, Synthesis and Recognition*. I-Tech Education and Publishing, 2008.
- [9] Didem, Yilidirim, *Affective Computing and Interaction Psychological, Cognitive and NeuroScientific Perspectives*. Information Science Reference, 2011.
- [10] Michael Lewis, Haviland-Jones and Lisa Feldman, *Handbook of Emotions*. The Guilford Press, 2008.
- [11] V. Koladyazhniy, S.D. Kreibig, J.J. Gross, W.T. Roth and F.H. Wilhelm, "An Affective computing approach to physiological emotion specificity: Toward subject-independent and stimulus independent classification of film induced emotions," *Psychophysiology*, vol. 48, pp. 908-922, 2011. DOI: 10.1111/j.1469-8986.2010.01170.x.
- [12] P. Ekman, W. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, 1978.
- [13] Mu-Chun Su, Yi-Jwu Hsieh and De-Yuan Huang, "A simple approach to facial expression recognition," *In Proc. of the WSEAS Int'l Conference on Computer Engineering and Applications*, pp. 456-461, Jan. 2007.
- [14] Jagdish Lal Raheja, Umesh Kumar, "Human facial expression detection from detected in captured image using back propagation neural network," *Int'l Journal of computer science & Information Technology*, vol.2, pp. 116-123, Feb. 2010.
- [15] Zisheng, Jun-Ichi, Masahide, "Facial-component-based bag of words and PHOG descriptor for facial expression recognition," *In Proc. of IEEE Int'l conference on Systems, Man and Cybernetics*, pp.1353-1358, Oct. 2009. DOI: 10.1109/ICSMC.2009.5346254
- [16] L. Zhang, D. Tjondronegoro, "Facial expression recognition using facial movement features," *IEEE Transactions on Affective Computing*, vol. 2, pp. 219-229, 2011. DOI: 10.1109/T-AFFC.2011.13
- [17] G. Zhao and M. Pietikainen "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.29, no.6, pp.915-928, June 2007. DOI: 10.1109/TPAMI.2007.1110
- [18] Ramchand Hablani, Narendra Chaudari, Sanjay Tanwani, " Recognition of facial expressions using local binary patterns of important facial parts," *Int'l Journal on Image Processing*, vol. 7, pp. 163-170, 2013.
- [19] F. Dornaika and F. Davoine. "Simultaneous facial action tracking and expression recognition in the presence of head motion". *Int'l Journal of Computer Vision*, vol.76, no. 3, pp. 257-281, March 2008. DOI: 10.1007/s11263-007-0059-7
- [20] F. Dornaika, A.Moujahid, B. Raducanu, "Facial expression recognition using tracked facial actions: classifier performance analysis," *Engineering Applications of Artificial Intelligence*, vol. 26, pp. 467-477, 2013. DOI: 10.1016/j.engappai.2012.09.002
- [21] M. Hayat, M. Bennamoun, "An automatic framework for textured 3D video-based facial expression recognition," *IEEE Transactions on Affective Computing*, vol. 5, pp. 301-313, 2014. DOI: 10.1109/T-AFFC.2014.2330580
- [22] Seyedehsamanah, Wei-Yun, Karthik, Jun Li, Eam Khwang, "Robust representation and recognition of facial emotions using extreme sparse learning," *IEEE Transactions on Image Processing*, vol. 24, pp. 2140-2152, 2015. DOI: 10.1109/TIP.2015.2416634
- [23] Andrea Kliensmith and Nadia, "Affective body expression perception and recognition: A Survey," *IEEE Transactions on Affective Computing*, vol. 4, pp. 15-33, 2013. DOI: 10.1109/T-AFFC.2012.16

- [24] G. Castellano, S.D. Villalba and A. Camuri, "Recognising human emotions from body movement and gesture dynamics," *In Proc. of 2nd Int'l Conference on Affective Computing and Intelligent Interaction*, pp. 71-82, 2007. DOI: 10.1007/978-3-540-74889-2_7
- [25] H. Gunes and M. Piccardi, "Affect recognition from face and body: early fusion vs. late fusion," *In Proc. of IEEE Int'l conference on Systems, Man and Cybernetics*, vol. 4, Oct. 2005. DOI: 10.1109/ICSMC.2005.1571679
- [26] Tadas Baltrušaitis et al., "Real-time inference of mental states from facial expressions and upper body gestures," *In Proc. of the IEEE Int'l Conference on Automatic Face and Gesture Recognition and Workshops*, pp. 909-914, 2011. DOI: 10.1109/FG.2011.5771372
- [27] H. Gunes and M. Piccardi, "Automatic temporal segment detection and affect recognition from face and body display," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 39, Feb 2009. DOI: 10.1109/TSMCB.2008.927269
- [28] S. Chen, Y.L. Tian, Q. Liu and D. N. Metaxas, "Recognizing expressions from face and body gesture by temporal normalized motion and appearance features," *Image and Vision Computing*, vol. 31, pp. 175-185, 2013. DOI: 10.1109/CVPRW.2011.5981880
- [29] Ayadi M. E., Kamel M. S. and Karray F., 'Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases', *Pattern Recognition*, vol. 44 (16), pp. 572-587, 2011. DOI: 10.1016/j.patcog.2010.09.020
- [30] I. Luengo and I. Hernaez, "Feature analysis and evaluation for automatic emotion identification in speech," *IEEE Transactions on Multimedia*, vol. 12, pp. 490-501, Oct 2010. DOI: 10.1109/TMM.2010.2051872
- [31] Carlos. B, Mariooryad. S, Metallinou. A and Narayanan. S, "Iterative feature normalization scheme for automatic emotion detection from speech," *IEEE Transactions on Affective Computing*, vol. 4, pp. 386-397, 2013. DOI: 10.1109/T-AFFC.2013.26
- [32] Q. Mao, M. Dong, Z. Huang and Y. Zhan, "Learning salient features for speech emotion recognition using convolutional neural networks," *IEEE Transactions on Multimedia*, vol. 16, pp. 2203-2213, Dec 2014. DOI: 10.1109/TMM.2014.2360798
- [33] M. J. Gangeh, P. Fewzee, G.A, Mohamed, S. Kamel and F. Karray, "Multiview supervised dictionary learning in speech emotion recognition," *IEEE Transactions on Audio, Video, and Language Processing*, vol. 22, pp. 1056-1068, June 2014. DOI: 10.1109/TASLP.2014.2319157
- [34] K. Wang, A. Ning, B.N. Li and Y. Zhang, "Speech emotion recognition using fourier parameters," *IEEE Transactions on Affective Computing*, vol. 6, pp. 69-75, 2015. DOI: 10.1109/TAFFC.2015.2392101
- [35] H. A. Vu, Y. Yamazaki, F. Dong and K. Hirota, "Emotion recognition based on human gesture and speech information using RT middleware," *In proc. of IEEE Int'l Conference on Fuzzy Systems*, pp. 787-791, 2011. DOI: 10.1109/FUZZY.2011.6007557
- [36] K. LU and Y. JIA, "Audio visual emotion recognition using boltzmann zippers", *In proc. of IEEE Int'l Conference on Image Processing*, pp. 2589-2592, 2012. DOI: 10.1109/ICIP.2012.6467428
- [37] J. Kortelainen et al., "Multimodal emotion recognition by combining physiological signals and facial expressions: a preliminary study", *In proc. of IEEE Int'l conference on Embedded Systems*, pp. 5238-5241, Sep. 2012. DOI: 10.1109/EMBC.2012.6347175
- [38] V. Rozgic, Shiv N. Vitaladevuni and R. Prasad, "Robust EEG emotion classification using segment level decision fusion", *in proc. Of IEEE Int'l. Conference on Acoustics, Speech and Signal Processing*, pp. 1286-1290, 2013. DOI: 10.1109/ICASSP.2013.6637858
- [39] D. Lakens, "Using a smartphone to measure heart rate changes during relived happiness and anger", *IEEE Transactions on Affective Computing*, vol. 4, pp. 238-241, 2013. DOI: 10.1109/T-AFFC.2013.3
- [40] Wanhui Wen et al., "Emotion recognition based on multi-variant correlation of physiological signals", *IEEE Transactions on Affective Computing*, vol. 5, pp. 126-140, 2014. DOI: 10.1109/TAFFC.2014.2327617
- [41] J. Hernandez, D. McDuff, R. Fletcher and R.W. Picard, "Inside-Out: reflecting on your inner state," *In proc. of IEEE Int'l Conference on Pervasive Computing and Communications Workshops*, pp. 324-327, 2013. DOI: 10.1109/PerComW.2013.6529507

- [42] P. Ren, A. Barreto, Y. Gao and M. Adjouadi, "Affective assessment by digital processing of the pupil diameter," *IEEE Transactions on Affective computing*, vol. 4, pp. 2-14, 2013. DOI: 10.1109/T-AFFC.2012.25
- [43] G. Caridakis et. Al., "Multimodal emotion recognition from expressive faces, body gestures and speech," *in proc. Int'l. Conf. Artificial Intelligence Applications and Innovations*, pp. 375-388, 2007. DOI: 10.1007/978-0-387-74161-1_41
- [44] K. Tang, Y. Tie, T. Yang and L. Guan, "Multimodal emotion recognition system," *in proc. of IEEE Canadian Conference on Electrical and Computer Engineering*, pp. 1-6, 2014. DOI: 10.1109/CCECE.2014.6900993
- [45] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE transactions on Affective Computing*, vol. 3, pp. 211-223, 2012. DOI: 10.1109/T-AFFC.2011.37
- [46] W. Zheng, B. Dong and B. Lu, "Multimodal emotion recognition using EEG and eye tracking data," *in proc. Intl. Conf. Engineering, Medicine and Biology Society*, pp. 5040-5043, 2014. DOI: 10.1109/EMBC.2014.6944757.
- [47] J. Wagner, F. Lingenfelter, E. Andre and J. Kim, "Exploring fusion methods for multimodal emotion recognition with missing data," *IEEE Transactions on Affective Computing*, vol. 2, pp. 206-218, 2011. DOI: 10.1109/T-AFFC.2011.12
- [48] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive Database for Facial Expression Analysis," *In proc. of IEEE Int'l Conference on Automatic Face and Gesture Recognition*, pp. 46-53, 2000. DOI: 10.1109/AFGR.2000.840611
- [49] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A Complete Dataset for Action Unit and Emotion-Specified Expression," *In Proc. of IEEE Conference in Computer Vision and Pattern Recognition Workshops*, pp. 94-101, June 2010. DOI: 10.1109/CVPRW.2010.5543262
- [50] M.J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding Facial Expressions with Gabor Wavelets," *In Proc. of Third IEEE Conference on Automatic Face and Gesture Recognition*, pp. 200-205, Apr, 1998. DOI: 10.1109/AFGR.1998.670949
- [51] L. Yin, X. Wei, Y. Sun, J. Wang, and M.J. Rosato, "A 3D Facial Expression Database for Facial Behavior Research," *In Proc. of Seventh IEEE Int'l Conference on Automatic Face and Gesture Recognition*, pp. 211-216, 2006. DOI: 10.1109/FGR.2006.6
- [52] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale, "A high-resolution 3d dynamic facial expression database," *in Proc. of 8th Int'l Conference on Automatic Face and Gesture Recognition (FGR08)*, pp. 17-19, Sep. 2008. DOI: 10.1109/AFGR.2008.4813324
- [53] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-Based Database for Facial Expression Analysis," *In Proc. of IEEE Int'l Conference on Multimedia and Expo*, vol. 5, pp. 6-8, July 2005. DOI: 10.1109/ICME.2005.1521424
- [54] W. Shangfei, L. Zhilei, L. Siliang, L. Yanpeng, W. Guobing, P. Peng, C. Fei, and W. Xufa, "A Natural Visible and Infrared Facial Expression Database for Expression Recognition and Emotion Inference," *IEEE Transactions on Multimedia*, vol. 12, no. 7, pp. 682-691, Nov.2010. DOI: 10.1109/TMM.2010.2060716
- [55] D. McDuff, R. El Kaliouby, T. Senechal and Amr. M, "Affectiva-MIT Facial Expression Dataset (AM-FED): Naturalistic and Spontaneous Facial Expressions Collected "In-the-Wild"," *In proc. of IEEE Int'l conference on Computer Vision and Pattern Recognition Workshops*, pp. 881-888, 2013. DOI: 10.1109/CVPRW.2013.130.
- [56] S.M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh and J. F. Cohn, "DISFA: A Spontaneous Facial Action Intensity Database," *IEEE Transactions on Affective Computing*, vol. 4, pp. 151-159, 2013. DOI: 10.1109/T-AFFC.2013.4
- [57] Y. Baveye, E. Dellandrea, C. Chamaret and L. Chen, "LIRIS-ACCEDE: A Video Database for Affective Content Analysis," *IEEE Transactions on Affective Computing*, vol. 6, pp. 43-55, 2015. DOI: 10.1109/TAFFC.2015.2396531

- [58] H. Gunes and M. Piccardi, "A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior", *In Proceedings of International Conference on Pattern Recognition*, pp. 1148-1153, 2006. DOI: 10.1109/ICPR.2006.39
- [59] R. Min, N. Kose and J.L. Dugelay, "KinectFaceDB: A Kinect Database for Face Recognition," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 44, pp. 1534-1548, 2014. DOI: 10.1109/TSMC.2014.2331215
- [60] E. Douglas-Cowie, N. Campbell and R.P. Cowie, "Emotional speech: Towards a new generation of databases", *Speech Communication* vol. 40, pp. 33-60, 2003. DOI: 10.1016/S0167-6393(02)00070-5
- [61] Gabriele Fanelli, Juergen Gall, Harald Romsdorfer, Thibaut Weise and Luc Van Gool "A 3-D Audio-Visual Corpus of Affective Communication," *IEEE Transactions on Multimedia*, vol. 12, No. 6, pp. 591 - 598, October 2010. DOI: 10.1109/TMM.2010.2052239
- [62] S. Haq, P. J. B. Jackson, and J. D. Edge, "Speaker-dependent audiovisual emotion recognition," *In Proc. of Int'l Conference on Auditory-Visual Speech Process.*, pp. 53-58, Sep. 2009.
- [63] C. Busso, M. Bulut, C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. Chang, S. Lee, and S. Narayanan, "IEMOCAP: Interactive Emotional Dyadic Motion Capture Database," *Journal on Language Resources and Evaluation*, vol. 42, no. 4, pp. 335-359, Dec. 2008.
- [64] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, and B. Weiss, "A database of german emotional speech," *in Proc. INTERSPEECH*, pp. 1517-1520, 2005.
- [65] Inger S. Engberg, A. Hansen, O. Andersen, and P. Dalsgaard, "Design, recording and verification of a Danish emotional speech database," *In Proc. of 5th Eur. Conference on Speech Commun. Technology*, pp. 1695-1698, Sep. 1997.
- [66] T. Wu, Y. Yang, Z. Wu and D. Li, "MASC: A Speech Corpus in Mandarin for Emotion Analysis and Affective Speaker Recognition," *IEEE Speaker and Language Recognition Workshop*, pp. 1-5, 2006. DOI: 10.1109/ODYSSEY.2006.248084
- [67] G. McKeown, M. Valstar, R. Cowie, M. Pantic and M. Schroder, "The Semaine Database: annotated multimodal records of emotionally colored conversations between a person and limited agent," *IEEE Transactions on Affective Computing*, vol. 3, pp. 5-17, 2012. DOI: 10.1109/T-AFFC.2011.20
- [68] B. Schuller, M. Valstar, F. Eyben, R. Cowie, and M. Pantic, "AVEC 2012—the continuous audio/visual emotion challenge," *In Proc. 14th ACM Int'l Conference on Multimodal Interact.*, pp. 449-456, 2012.
- [69] G. Caridakis, J. Wagner, A. Raouzaoui, Z. Curto, E. Andre', and K. Karpozis, "A multimodal corpus for gesture expressivity analysis multimodal corpora", *In Proc. Workshop Multimodal Corpora*, 2010.
- [70] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A Multi-Modal Database for Affect Recognition and Implicit Tagging," *IEEE Trans. Affective Computing*, vol. 3, no. 1, pp. 42-55, 2012. DOI: 10.1109/T-AFFC.2011.25
- [71] S. Koelstra et al., "DEAP: A database for emotion analysis using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, pp. 18-31, 2012. DOI: 10.1109/T-AFFC.2011.15
- [72] L. Shen, M. Wang, and R. Shen, "Affective e-Learning: Using "Emotional" Data to Improve Learning in Pervasive Learning Environment," *Educational Technology & Society*, pp. 176-189, 2009.
- [73] J.F. Grafsgaard, J.B. Wiggins, K.E. Boyer, E.N. Wiebe and J.C. Lester, "Automatically recognizing facial expression: predicting engagement and frustration," *In Proc. of 6th Int'l Conference on Educational Data Mining*, 2013.
- [74] J. Whitehill, Z. Serpell, Y. Lin, A. Foster and J. R. Movellan, "The faces of engagement: automatic recognition of student engagement from facial expressions," *IEEE Transactions on Affective Computing*, vol. 5, pp. 86-98, 2014. DOI: 10.1109/TAFFC.2014.2316163
- [75] B. Woolf et al., "Affect-aware tutors: recognising and responding to student affect," *In Proc. of Int'l J. Learning Technology*, vol. 4, pp. 129-164, 2009. DOI: 10.1504/IJLT.2009.028804
- [76] Karla Conn Welch, "Physiological signals of autistic children can be useful," *IEEE Instrumentation & Measurement Magazine*, pp. 28-32, 2012. DOI: 10.1109/MIM.2012.6145259
- [77] R. el Kaliouby and P. Robinson, "The emotional hearing aid: an assistive tool for children with asperger syndrome," *In proc. Univ Access Inf Soc*, pp. 121-134, 2005. DOI: 10.1007/s10209-005-0119-0

- [78] L.Zhang, A.Hossain, M.Jiang, "Intelligent Facial Action and Emotion Recognition for Humanoid Robots," *In Proc. Int'l Joint Conference on Neural Networks*, pp. 739-746, July 2014. DOI: 10.1109/IJCNN.2014.6889647
- [79] Urvashi Agrawal and Shubhangi Giripunje,"Emotion and Gesture Recognition with Soft Computing Tool for Drivers Assistance System in Human Centered Transportation," *In Proc. Int'l. Conference on Systems,Man and Cybernetics*, pp. 4612-4616, 2013. DOI: 10.1109/SMC.2013.785
- [80] G.Chanel, C. Rebetez, M. Betrancourt and T.Pun, "Emotion assessment from physiological signals for adaptation of game difficulty," *IEEE Transactions on Systems,Man and Cybernetics*, vol. 41, pp. 1052-1063,Nov.2011.DOI: 10.1109/TSMCA.2011.2116000
- [81] M.Malcangi, "Smart recognition and Synthesis of Emotional Speech for Embedded Systems with Natural User Interfaces," *In Proc. of Int'l. Joint Conference on Neural Networks*, pp. 867-871, 2011. DOI: 10.1109/IJCNN.2011.6033312
- [82] S.Mariooryad and C. Busso,"Exploring Cross-Modality affective reactions for audiovisual emotion recognition," *IEEE Transactions on affective computing*, vol. 4, pp. 183-196, 2013. DOI: 10.1109/T-AFFC.2013.11
- [83] A. Metallinou, A. Katsamanis, Y. Wang and S. Narayanan,"Tracking changes in continuous emotion states using body language and prosodic cues," *in proc. IEEE Int'l. Conference on Acoustics, Speech and Signal Processing*, pp. 2288-2291, 2011. DOI: 10.1109/ICASSP.2011.5946939
- [84] S. Wang et. al., "Analyses of a multimodal spontaneous facial expression database," *IEEE Transactions on Affective Computing*, vol. 4, pp. 34-46, 2013. DOI: 10.1109/T-AFFC.2012.32
- [85] G. Guo, R. Guo and X. Li,"Facial Expression Recognition influenced by human aging," *IEEE Transactions on Affective Computing*, vol. 4, pp. 291-298, 2013. DOI: 10.1109/T-AFFC.2013.13
- [86] G. Guo and X. Wang," A study on human age estimation under facial expression changes," *In proc. of IEEE Int'l Conference on Computer Vision and Pattern Recognition*, pp. 2547-2553, 2012. DOI: 10.1109/CVPR.2012.6247972
- [87] O. T-C. Chen, J.J.Gu, P-T Lu and J-Y. Ke,"Emotion-Inspired age and gender recognition systems,"*In Proc. of Int'l. Symp. Circuits and Systems*, pp. 662-665, 2012. DOI: 10.1109/MWSCAS.2012.6292107
- [88] M. Sidorov, S. Ultes and A. Schmitt,"Comparison of gender and speaker adaptive emotion recognition," *In proc. of Int'l Conference on Language Resources and Evaluation*, pp. 3476-3480, 2014.
- [89] A. Metallinou et. al., "Context-Sensitive Learning for Enhanced Audiovisual Emotion Classification," *IEEE Transactions on Affective Computing*, vol. 3, pp. 184-198, 2012. DOI: 10.1109/T-AFFC.2011.40
- [90] O. Rudovic, V. Pavlovic and M. Pantic,"Context-Sensitive dynamic ordinal regression for intensity estimation of facial action units," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, pp. 944-958, May 2015. DOI: 10.1109/TPAMI.2014.2356192
- [91] J.Chen and Q.Ji,"*Drowsy driver posture, facial, and eye monitoring methods*", *Handbook of Intelligent Vehicles*", pp. 913-940, 2012. DOI: 10.1007/978-0-85729-085-4_35
- [92] B. Raducanu and F. Dornaika. "Dynamic facial expression recognition using Laplacian Eigenmaps-based manifold learning". *IEEE International Conference on Robotics and Automation*, pp. 156-161, Alaska, USA, May 2010. DOI: 10.1109/ROBOT.2010.5509290
- [93] M.D. Van der zwaag, J.H. Janssen and J.H.D.M. Westerink," Directing Physiology and Mood through Music:Validation of an Affective Music Player", *IEEE Transactions On Affective Computing*, vol. 4, no. 1, pp. 57-68, 2013. DOI: 10.1109/T-AFFC.2012.28