International Journal of Current Research and Modern Education (IJCRME)
ISSN (Online):2455 - 5428
(www.rdmodernresearch.com) Volume I, Issue I, 2016

MULTIMODAL HUMAN-MACHINE EMOTIONAL INTELLIGENCE SYSTEM

K. Suresh*, Dr.C. Chellappan, Dr. A. Banumathi*****

* Research Scholar, Department of CSE, Anna University, G.K.M. College of Engg. & Tech., Research Centre, Chennai, Tamilnadu, India.

**Senior Professor & Principal,G.K.M. College of Engg. & Tech., Chennai, Tamilnadu, India.

*** Associate Professor, Department of ECE, Thiyagarajar College of Engg. & Tech., Madurai, Tamilnadu, India.

Abstract:

The lot of researcher have developed intelligent human behaviours system that can effectively perform human behaviours detection method but in order to react appropriately as like a human, the computer would need to have some perception of the emotional state of the human, so that this research is used to evaluate the potential for emotion recognition technology to improve the quality of human- computer interaction in real time. Emotions play an important role in human-to-human communication and interaction, allowing people to express them beyond the verbal domain. The ability to understand human emotions is desirable for the computer in several applications. We present the basic research in the field and the recent advances into the emotion recognition from facial, voice, text, body gesture, body posture and physiological signal, w here the different modalities are treated independently. Moreover, an increasing number of efforts are reported toward multimodal fusion technique for human affect analysis.

Keywords: Emotion detection, human machine-interaction, expression, physical movements.

I Introduction

To create a highly accurate emotion recognition system that is able to reflect a real emotion using computer technology.Even though facial expression represents the most significant factor in human emotional expression, we might not be able to accurately and fully identify individuals' real emotion so multimodal emotion recognition systems is necessary.Multimodal recognition systems involve combination of different types of input data.Multimodal recognition system involves multiple modalities based on different categories of features.

Research on multimodal face modality is a very challenging field that target methods to make effective human computer interaction. Facial expression carries crucial information about the mental, emotional and even physical states of the conversation. Recognition of facial expression in the input image needs two functions: locating a face in the image and recognizing its expression. When we watch two photos of a human face, we can answer which photo shows the facial expression more strongly. In human interaction, the articulation and perception of facial expressions form a communication channel, that is additional to voice and that carries crucial information about the mental, emotional and even physical states of the conversation. It detects face and ignores anything else, such as buildings, trees and bodies.

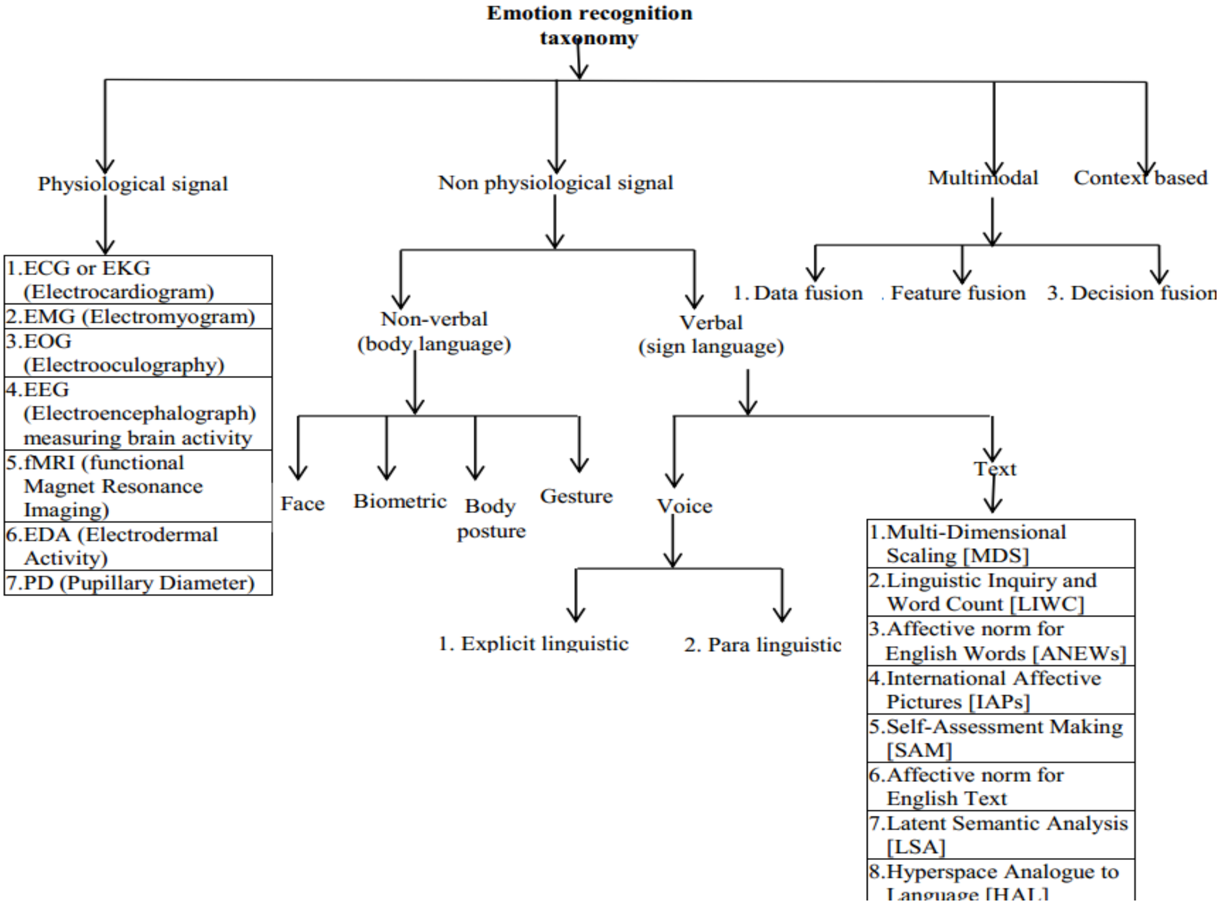
Emotion recognition taxonomy
Physiological signal
Non physiological signal
Multimodal
Context based
1.ECG or EKG (Electrocardiogram)
2.EMG (Electromyogram)
3.EOG (Electrooculography)
4.EEG (Electroencephalograph) measuring brain activity
5.fMRI (functional Magnet Resonance Imaging)
6.EDA (Electrodermal Activity)
7.PD (Pupillary Diameter)
Non-verbal (body language)
Verbal (sign language)
1. Data fusion
. Feature fusion
3. Decision fusion
Face
Biometric
Body posture
Gesture
Voice
Text
1. Explicit linguistic
2. Para linguistic
1.Multi-Dimensional Scaling [MDS]
2.Linguistic Inquiry and Word Count [LIWC]
3.Affective norm for English Words [ANEWs]
4.International Affective Pictures [IAPs]
5.Self-Assessment Making [SAM]
6.Affective norm for English Text
7.Latent Semantic Analysis [LSA]
8.Hyperspace Analogue to Language [HAL]

Fig. 1 Multimodal interaction taxonomy

Face detection [17] can be regarded as a more general case of face localization. In face localization, the task is to find the locations and sizes of a known number of faces (usually one). In face detection, face is processed and matched bitwise with the underlying face image in the database. When we seeing a photos of a human face, we can answer which photo shows the facial expression more strongly.The multisensory data are typically processed separately and only combined at the end. People display audio and visual communicative signals in a complementary and redundant manner. In order to accomplish a human-like multimodal analysis of multiple input signals acquired by different sensors, the signals cannot be considered mutually independent and cannot be combined in a context-free manner at the end of the intended analysis.Emotion recognition is performed by fusing information coming from physiological signals [16],visual and auditory modalities.

II Related Work

As per Albert Mehrabian'sin communication research [1], 7% of meaning in the words that are spoken, 38% of meaning is paralinguistic (the way that the words are said), 55% of communicating cues can be judge by facial expression, hence recognition of facial expression become a major modality. In 1872- Darwin's Charles demonstrated the universality of facial expression and their continuity in man and animals and claimed among other things [2]. In early 1970s Paul Ekman has performed extensive study of human facial expressions. 1971 American psychologist Ekman and Friesen defined six basic emotions: angry, disgust, fear, happiness, sadness, and surprise [3]. The approach on Facial Action Coding System (FACS) which separates the expression into upper and lower face action[8]. In 1978 FACS was developed by Ekman [7] for facial expression description. 1978-suwea, et.al, presented a preliminary investigation on automatic facial expression analysis from an image sequence [4]. In 1991-Mase and Pentland used 8 directions of optical flow changes to detect the movement of FACS. Before the year of 2005 the most facial expression recognition systems was based on 2-D static images only. In 2005, the scholars of the university of science and technology put forward a 3-D facial model which is based on the facial expression recognition method [5]. In 2008 Cao and Tong proposed a new method based on embedded Hidden markov model and Local binary pattern [11].

III Architecture

Fig. 2 shows the entire system working model of multimodal emotion recognition systems with each level. In the first level each modality acquired the input data from associated devices. In level two from acquired data each modality concentrating on relevant information other unnecessary information are ignored. In level three each modality performing preprocessing on relevant data to speed up the performance. In level four each modality extracts the dominant key features from the preprocessed data, in level five all processed information’s are fused using multimodal fusion algorithms and finally in output stage classifier algorithm classify the recognized in to the particular state of emotions.

IV Challenges

Factors that potentially affect the performance of emotion recognition

S.no	Challenges
1	A low resolution camera may produce inferior results compared to a high resolution camera
2	The quality of voice recording in a noisy environment may be lower than that in a quiet place.
3	An unsecured database can result in possible hacking to the database itself.
4	An inefficient algorithm may take longer to execute and/or lead to lower recognition accuracy.
5	For instance, a person crying with a mumbling voice may smile at the same time. Determining if a person is experiencing greater happiness or sadness requires the consideration of multiple factors.
6	The lighting condition used in the recording environment.
7	The angle of the user’s face.
8	To maintain a certain level of the quality, we are faced with two critical issues, i.e. the device selection and the recording procedure.
9	Changes in the shape of mouth caused due to speech.
10	The distance between facial landmarks vary from person to person, thereby making the person independent expression recognition system less reliable.
11	Resolution required for the extraction of facial feature is much larger than the one for body movement detection or hand gesture tracking.

Table I: Various challenges at facial emotion recognition

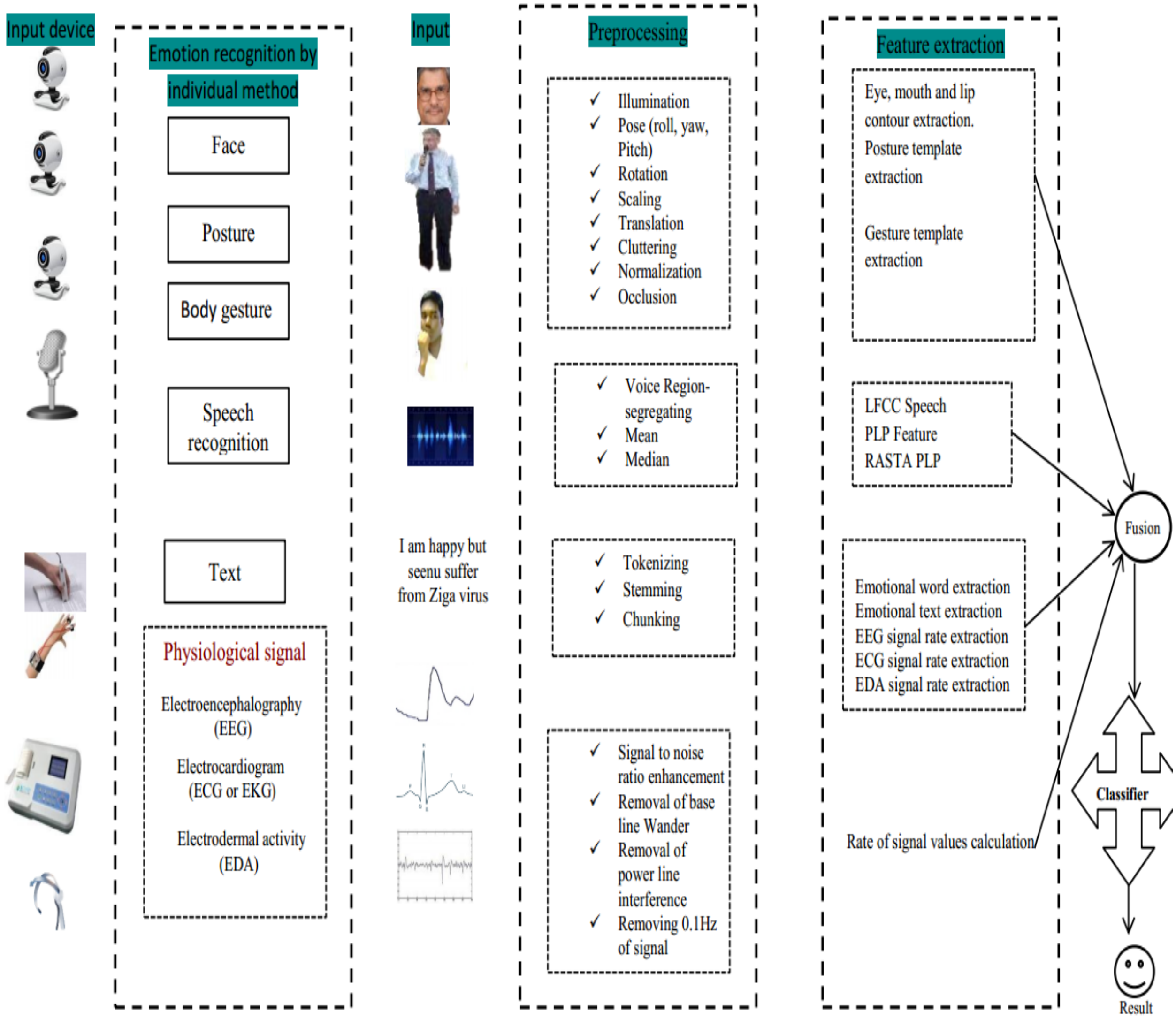
Input device
Emotion recognition by individual method
Face
Posture
Body gesture
Speech recognition
Text
Physiological signal
Electroencephalography (EEG)
Electrocardiogram (ECG or EKG)
Electrodermal activity (EDA)
Input
I am happy but seenu suffer from Ziga virus
Preprocessing
✓ Illumination
✓ Pose (roll, yaw, Pitch)
✓ Rotation
✓ Scaling
✓ Translation
✓ Cluttering
✓ Normalization
✓ Occlusion
✓ Voice Region-segregating
✓ Mean
✓ Median
✓ Tokenizing
✓ Stemming
✓ Chunking
✓ Signal to noise ratio enhancement
✓ Removal of base line Wander
✓ Removal of power line interference
✓ Removing 0.1Hz of signal
Feature extraction
Eye, mouth and lip contour extraction.
Posture template extraction
Gesture template extraction
LFCC Speech
PLP Feature
RASTA PLP
Emotional word extraction
Emotional text extraction
EEG signal rate extraction
ECG signal rate extraction
EDA signal rate extraction
Rate of signal values calculation
Fusion
Classifier
Result

Fig. 2 Multimodal human-machine interaction architecture.

V Applications

Multimodal emotion recognition system most widely used in the following application areas

- Authentication
- Artificial intelligent based control
- Automated access control
- Advance driver assistance systems
- Autism spectrum disorders support
- Behavior understanding& prediction
- Business environment
- Customer satisfactions studies for broadcast and web services cognitive load
- Patient healthcare monitoring[20] and so on.

VI Modules

Every multimodal human machine recognition system must perform a few steps before classifying the recognized input into a particular emotion. For facial methods first it needs to find the face of the subject from the image or video feed data.

A. Data acquisition:

Data acquisition is the first and foremost primary work in every human computer interaction systems. In facial movement systems without image acquisition the processing on the image is not possible so the first step in facial expression recognition system is capturing an image. Image acquisition procedure itself includes several issues like an image properties, number of devices connected (camera, digitizer), size of the face image, total image dimensions, ambient lighting [6] and etc. Images acquired in low light or coarse resolution can provide less information about facial features. The methods that work well in studio lighting may perform poorly in more natural lighting when the angle of lighting changes across an image sequences. Most researchers use single-camera setups, usage of single camera may difficult to standardize when input is in out of plane rotation.

In audio data the spatial arrangement of the selected deviceis used to record the raw data, we first place the KINECT device on the table, The table is 0.95 meter high and 1.1 meter away from the chair that the camera faces the room size does not affect the recording quality, but the direct distance between the KINECT and the chair will affect the result. This ensures that our basic space coordinates between the device and the users are proper.

B. Input level

Similar to cropping a face (region of interest) from acquiesced photo, from other modality region of interest area is segmented i.e. segmenting the shape of body posture and gesture from input data, segmenting the rate of emotional peak and tolerance signal from recorded speech data, scan the text using text scanner device and segment the motional text from it.Electroencephalography, electrocardiogram and electrodermal activity devices are used to measures the different functioning of our body. From this data rate of physiological signal are calculated.

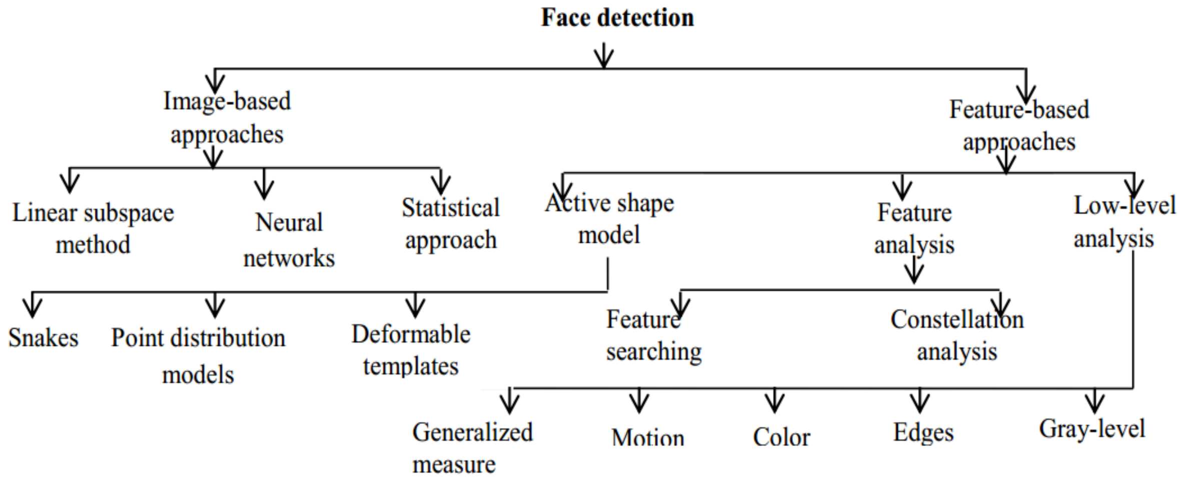

Fig. 3 Face detection approaches

C.Preprocessing

Image pre-processing techniques make the image easier to process the data and increase the chances of getting correct matches. Better chances of success with change in illumination, pose, and picture quality can decrease the processing time and increase the feature detection performance, compared with a non-preprocessed image. Common pre-processing methods in facial images are resampling, edge detection and face alignment is an essential step and is usually carried out by detection and horizontal positioning of eyes. Changing in lighting conditions provide considerable decrease in recognition performances. To overcome these disadvantage better preprocessing methods should be used before the extraction stage. Preprocessing methods used for illumination normalization in face images are the gamma intensity correction method, the logarithm transform method, discrete cosine transform method and histogram equalization method [18].The pre-processing of the audio datato cut the leading and lagging edge which have no relationship to the emotions, this become part of the normalization on audio data, as well as the noise reduction.

Preprocessing in speech modality include segregating voice region, tolerating mean and median. In text modality preprocessing consider the stemming, chunking and tokenizing of emotional text. Preprocessing in physiological signal include signal to noise ration enhancement, removal of base line wander, removal of power line interference, removing 0.1Hz of signal.

C. Feature extraction

Feature extraction is the process of extracting relevant information from input data for further processing.Feature extraction is an important part in multimodal human machine real time interactions. Feature extractions methods heavily depends on the choices of databases.Features extraction and representation are critical in multimodal human machine interaction system. For each model selection of feature extraction algorithm affects the classification accuracy. Extraction of features from facial images is a key role. People constantly change their positions when subjects are expressing emotions. In emotion recognition systems feature extraction responsible to identify facial features points[8]the prominent features of the face such as eyebrows, eyes, pupil diameter, nose, mouth, and chin.The three major features in audio-based emotional recognitions are pitch, energy, and frequencies [7] [17]. Canny edge detection [21] algorithm is applied to extract the important facial emotional features of mouth, eyes and eyebrows.

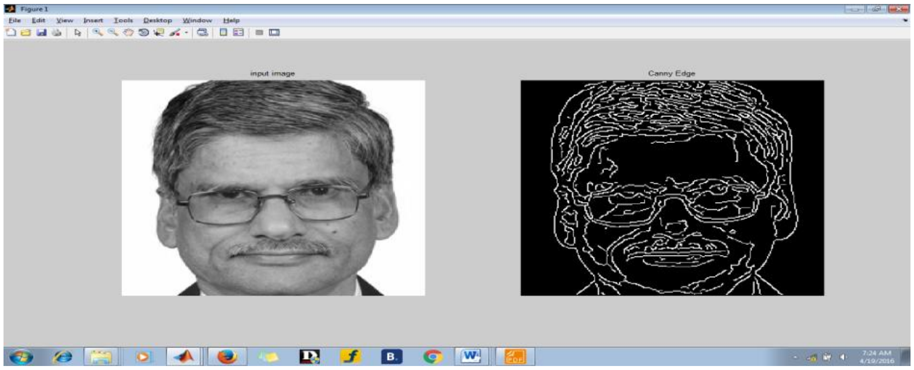

Fig. 4 Emotional feature cropping by canny edge detection

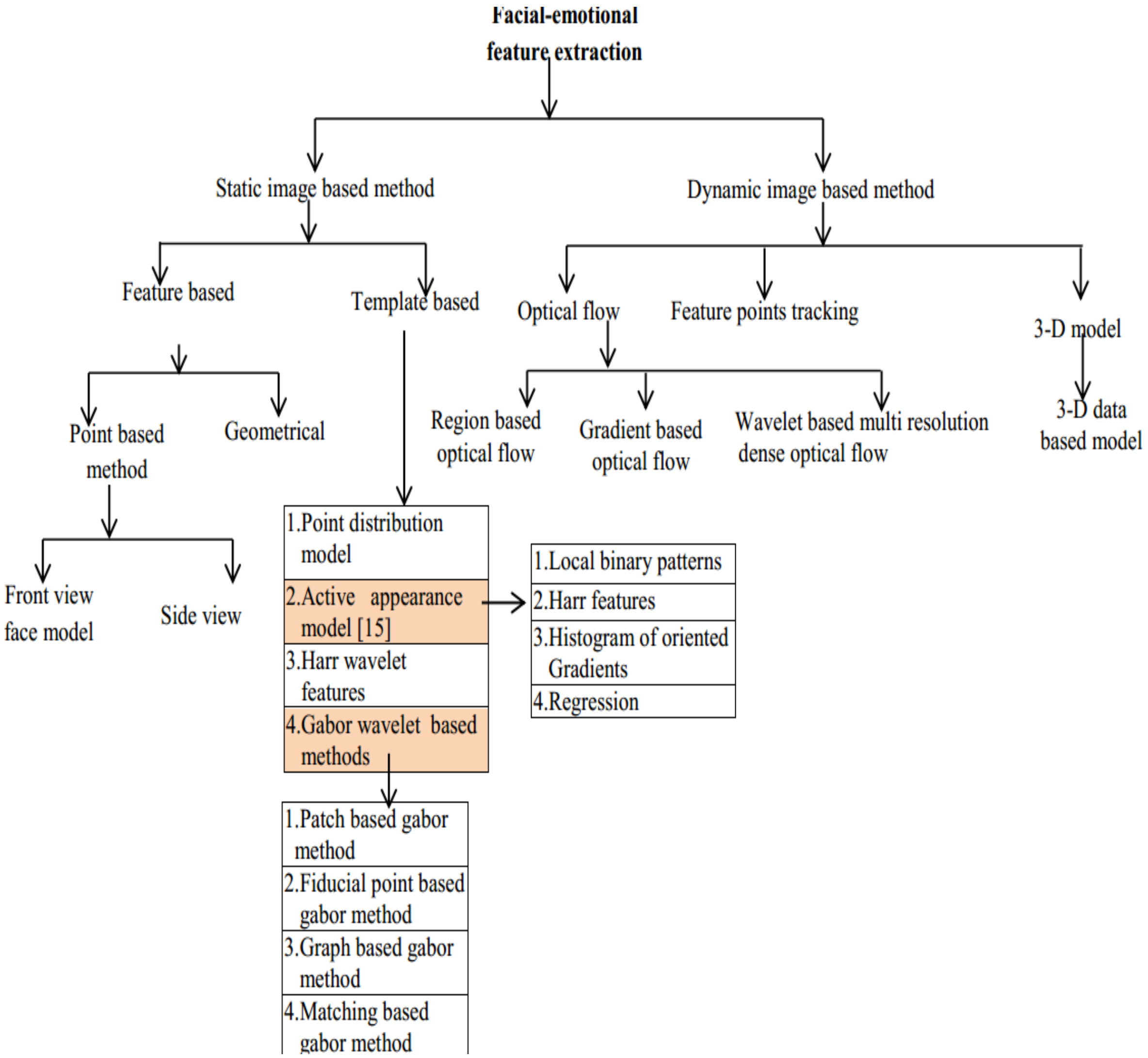
Facial-emotional feature extraction
Static image based method
Dynamic image based method
Feature based
Template based
Optical flow
Feature points tracking
3-D model
Point based method
Geometrical
Region based optical flow
Gradient based optical flow
Wavelet based multi resolution dense optical flow
3-D data based model
Front view face model
Side view
1.Point distribution model
2.Active appearance model [15]
3.Harr wavelet features
4.Gabor wavelet based methods
1.Local binary patterns
2.Harr features
3.Histogram of oriented Gradients
4.Regression
1.Patch based gabor method
2.Fiducial point based gabor method
3.Graph based gabor method
4.Matching based gabor method

Fig. 5 Taxonomy of facial feature extraction

Most widely used speech [12] feature extraction techniques are linear predictive coding, linear predictive cepstral coefficients, perceptual linear predictive coefficients, power spectral analysis, mel-frequency cepstral coefficients, relative spectra filtering of log domain coefficients, wavelet features and auditory features. Like facial, dominant features of speech include linear frequency cepstral coefficients, Perceptual Linear Predictive (PLP) and RASTA PLP [22].

D. Fusion methods

The fusion methods are divided into the following three categories: rule based methods, classification based methods, and estimation-based methods. This categorization is based on the

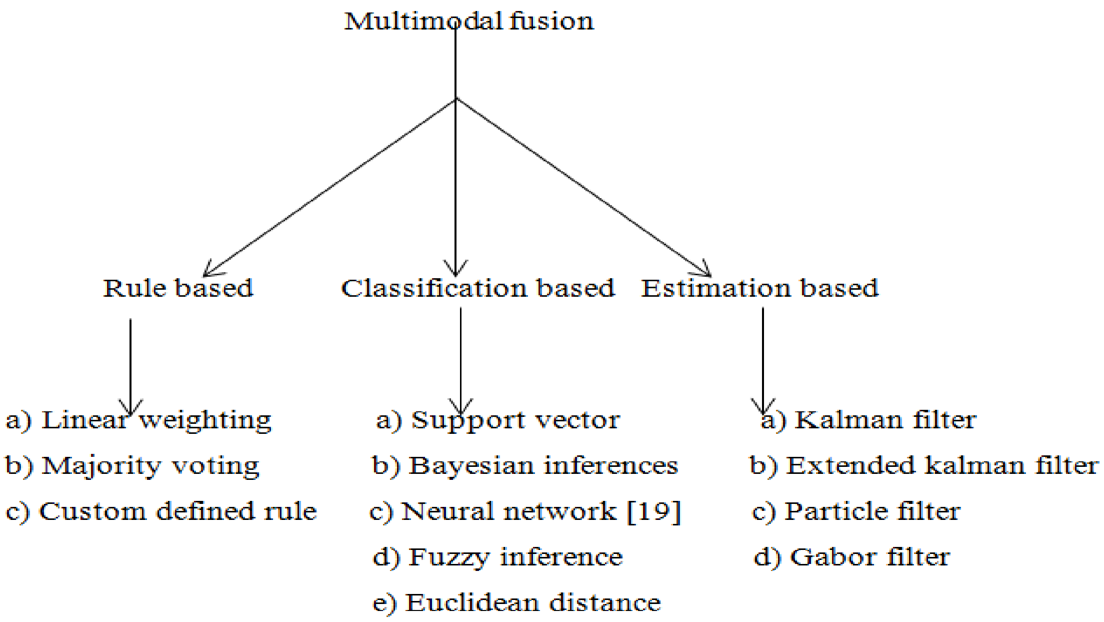

Fig. 6Taxonomy of multimodal fusion

basic nature of these methods and it inherently means the classification of the problem space, such as, a problem of estimating parameters is solved by estimation based methods. Similarly the problem of obtaining a decision based on certain observation can be solved by classificationbased or rule based methods. However, if the observation is obtained from different modalities, the method would require fusion of the observation scores before estimation or making a classification decision.

a. Rule-based fusion

The rule-based fusion method includes a variety of basic rules of combining multimodal information. These include statistical rule-based methods such as linear weighted fusion (sum and product), MAX, MIN, AND, OR, majority voting. There are custom-defined rules that are constructed for the specific application perspective. The rule-based schemes generally perform well if the quality of temporal alignment between different modalities is good.

b. Classification-based fusion

This category of methods includes a range of classification techniques that have been used to classify the multimodal observation into one of the pre-defined classes. The methods in this category are the support vector machine, bayesianinference, dempster–shafer theory, dynamic bayesiannetworks, neural networks [18] andmaximum entropy model. Note that we can further classify these methods as generative and discriminative models from the machine learning perspective. For example, bayesian inference and dynamic bayesian networks are generative models, while support vector machine and neural networks are discriminative models.

c. Estimation-based fusion

The estimation category includes the kalman filter, extended kalman filter and particle filter fusion methods. These methods have been primarily used to better estimate the state of a moving object based on multimodal data. For example, for the task of object tracking, multiple modalities such as audio and video are fused to estimate the position of the object.

E. Classifier

Classification is a process of making a judgment. A classification aims at mapping emotional features in to one of several emotion classes. Expression classification [13] is a method of teaching a computer (machine learning) [14]to make and improve predictions or behaviours based on some data. The classifier module used to classify the recognized input in to the one of the universal state six basic emotions (angry, disgust, fear, happiness, sadness and surprise). Based on the application the classifiers is chosen most widely used classifiers are support vector machine, radial basis function network, multi-layer neural networks, adaboost classifier, k-nearest neighbor and so on.

IV Conclusion

In order to improve the quality of multimodal human machine interactionsthis paper demonstrate the performance of an interparticipant emotion recognition tagging approach using facial movements, body postures and gestures, speech recognition [23], text recognition, physiological signal [10] like, electroencephalography, electrocardiogram and electrodermal activity for each modality individually input information acquired, acquired data preprocessed to increase the recognition rate, features are extracted from preprocessed data and finally improve the performance using multimodal fusion techniques. Even though recognition rate is improved it consuming more time, so in future enhancement founding the single algorithm to fit all modality may reduce the time consuming factor.

V References

1. Mehrabian A., “Communication without words”, Psychology Today, Issue 2, Volume (9), pp. 52-55, 1968.
2. Darwin C., “The Expression of the Emotions in Man and Animals”, John Murray, 1872.
3. Ekman P, Friesen W V., “Facial Action Coding System: A Technique for the Measurement of Facial Movement”, Palo Alto: Consulting Psychologists Press, 1978.
4. M. Suwa, N. Sugie, and K. Fujinmora, “A Preliminary Note on Pattern Recognition of Human Emotional Expression”, 4th International Conference on Pattern Recognition, pp. 408-410, 1978.
5. K.Mase and A. Pentland, “Recognition of Facial Expression from Optical Flow”, IEICE Transactions, E 74(10):3474-3483, October 1991.
6. Suresh K., Chellappan C., “Overview and Comparative Analysis of Human Face Detection”, International Journal of Research in Engineering and Advanced Technology, Vol. 3, Issue 2, April-May, 2015.
7. Y. Wang and L. Guan, “An investigation of speech-based human emotion recognition”, IEEE 6th Workshop on Multimedia Signal Processing, 2004.
8. Robert Brunelli, TomasoPoggio, “Face Recognition: Features Versus Templates”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, No. 10, October 1993.

9. M. F. Valstar and M. Pantic, "Fully automatic recognition of the temporal phases of facial actions", IEEE Transaction. Systems. Vol 42, no.1, pp. 28-43, 2012.
10. L. Leon, G. Clarke, F. Sepulveda, and V. Callaghan, "Optimized attribute selection for emotion classification using physiological signals", in Proc. IEEE International Conference Automatic Face Gesture Recognition, pp. 184-187, Mar. 2005
11. Z. Guoying and M. Pietikainen, "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29, no. 6, pp. 915-928, June 2007.
12. Z. Zeng, M. Pantic, G. I. Roisman and T. S. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.31, no.1, pp.39,58, Jan. 2009.
13. AshimSaha, Anurag De, M .C. Pal and NirmalyaKar, "Different Techniques of Automatic Facial Expression Recognition: A Survey", second international conferences on advances in computing, communication and information technology, 2014.
14. Jason Brownlee, "A Tour of Machine Learning Algorithms", 2013, link: http://machine learningmastery.com/a-tour-of-machine-learning-algorithms.
15. Mostafa K. Abd El Meguid and Martin D. Levine, "Fully Automated Recognition of SpontaneousFacial Expressions in Videos Using Random Forest Classifiers", IEEE Transactions on affective computing, Vol. 5, No. 2, 2014.
16. Yuan GU and et.al, "Analysis of Physiological Responses from Multiple Subjects for Emotion Recognition", IEEE 14th International Conference on e-Health Networking, Applications and Services, pp 178-183, 2012.
17. Ming-HsuanYang and David J. Kriegman, "Detecting Faces in Images: A Survey", IEEE transactions on pattern analysis and machine intelligence, vol. 24, no.1, 2002.
18. H. A. Rowley, S. Baluja, and T. Kanade,"Neural network based face detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 1, pp. 23–38, Jan. 1998.
19. S. K. Oh, S.-H. Yoo, and W. Pedrycz, "Design of face recognition algorithm using PCA-LDA combined for hybrid data pre-processing and polynomial-based RBF Neural networks", The Transactions of the Korean Institute of Electrical Engineers, Expert syst. Appl., Vol.6, no.5, pp. 744-752, 2012.
20. ManidaSwangnetr and David B. Kaber, "Emotional State Classification in Patient–Robot Interaction Using Wavelet Analysis and Statistics-Based Feature Selection", IEEE Transaction on human-machine systems, Vol. 43, No.1, January 2013.
21. Xiaoming Chen and Wushan Cheng, "Facial expression recognition based on edge detection" , International journal of computer science and engineering survey, Vol. 6, No. 2, April 2015.
22. HynekHermansky and Nelson Morgan, "RASTA processing of speech", IEEE Transactions on speech and audio processing Vol. 2, No. 4, October 1994.
23. Moataz El Ayadi, et.al. , "Survey on speech emotion recognition: Features, classification schemes, and databases", ELSEVIER, pattern recognition 44, pp 572-587, 2011.