This chapter provides a self-review of the thesis and outlines the future direction of possible research in the broad spectrum of emotion recognition. It re-examines the justification of the proposed techniques and their possible extension in the context of emotion recognition from unimodal and multi-modal perspectives. A list of possible applications of the proposed research is appended at the end of the chapter.
5.1. CONCLUSIONS

The thesis introduced three distinct approaches to emotion recognition using both single and multiple modalities. The first approach, covered in Chapter 2, deals with facial expression analysis for emotion recognition of an unknown subject using type-2 fuzzy sets. The second approach introduced in Chapter 3 offers a novel approach to recognizing emotion from lip contour of the subject. The third approach discussed in Chapter 4 presents a technique for multi-modal emotion recognition. The multi-modal approach is concerned with emotion recognition using facial and voice features. It is apparent that unimodal approaches presented in Chapter 2 and 3 requires relatively small time for feature extraction, and thus is computationally efficient. The multi-modal approach on the other hand relies greatly on the features of individual modality, and thus has a high computational overhead.

The thesis is unique for its coverage on type-2 fuzzy logic based emotion recognition. Two distinct approaches of type-2 fuzzy reasoning using Interval Type-2 and General Type-2 Fuzzy Sets (IT2FS and GT2FS) are prevalent in the literature. Both the approaches employ a set of type-2 classification rules, where the antecedent clause of each rule includes a set of fuzzy propositions, describing the qualitative attribute of individual features. The consequent of each rule indicates the emotion class. The IT2FS based reasoning mechanism computes the aggregation of all fuzzified features using two membership curves, referred to as UMF and LMF. The average of the aggregated fuzzy memberships induced by the UMF and LMF provides the degree of support of the features to a given emotion class. The principle of computing the degree of support of a given emotion class to a given set of measurements is now repeated for each emotion, and the one with the highest degree of support is inferred as the emotion in favor of the given measurements. The GT2FS based recognition is similar with IT2FS based recognition, but before the recognition process is initiated, we here transform the primary membership at each measurement by the product of primary and secondary measurements at that point. Such transformation improves the qualitative performance in emotion classification.

The IT2FS based recognition is highly time-efficient for its small computational overhead. However, GT2FS based recognition requires extensive computational complexity, particularly to compute the secondary memberships corresponding to all primary membership functions for each facial expression of the known subjects covering all the five emotions. The main computational overhead of GT2FS lies in the minimization of an objective function. This cost is repeated several times to compute a large number of secondary membership grades, stated above. Of course, GT2FS gives very good classification accuracy at the cost of high offline computational overhead. One way to bridge the qualitative gap between GT2Fs and IT2Fs based recognition is to delete unwanted noisy data points by a statistical technique, called the Interval Approach (IA). The IA requires small computations to delete unwanted data points, consequently IA-IT2FS undertaken in Chapter 2 gives better classification accuracy than IT2FS only.

The emotion recognition by lip contour analysis also added a new flavour to the thesis. Most of the current research on emotion recognition often ignores the lip contour and emphasize more on the facial action points on the cheek, eye and eyebrow regions. The research results obtained from lip contour analysis, however, establishes its significance in emotion analysis. A parametric model has been used for lip contour analysis, and a Support Vector Machine (SVM) is used to classify emotion from the lip parameters extracted from the model. Further, an evolutionary approach is used to tune the lip contour with a given subject’s actual lip contour. The novelty of the present research thus lies in lip parametric model design, parameter extraction by lip contour matching using Evolutionary algorithm, and lastly emotion classification by SVM.
The third interesting contribution of the thesis is multimodal emotion recognition using facial expressions and voice. The study includes individual and bi-modal recognition of emotion and comparison of their relative performance. The most important aspect of the work is to consider the effect of noise on the individual and bimodal approach to emotion classification. The results obtained confirm that bimodal recognition outperforms individual modality from the points of view of classification accuracy.

The thesis thus to the best of the author’s knowledge and belief reaches its target to propose and design a novel scheme for emotion recognition, and validate the performance of the end design results by testing and evaluating the performance of the emotion recognition system. The three techniques proposed for the emotion recognition problem are novel and original and have been thoroughly examined by the author for its fruitful utilization in real world system for application in emotion recognition.

5.2. Future Research Directions

Physiological parameters have a great influence on the arousal of emotion. Current research on emotion recognition indicates that electroencephalogram (EEG), electro-cardiograph (ECG), blood pressure volume, respiration rate, galvanic skin response (GSR), pulse rate, respiration amplitude, and skin temperature have a good correlation with the arousal of emotion [1]. In our future research, we would include one or more of the above modalities along with facial features and/or voice for emotion recognition. Unlike facial expression or voice, where the users may voluntarily modulate the manifestations, the above psychological parameters can hardly be modulated by the subject. This justifies the importance of bio-potential signals like EEG, ECG and other physiological parameters referred to above.

Gesture provides more natural human response to stimulus, and thus has a great role in the emotion recognition problem. In [11], the authors consider gesture as the basic modality for emotion recognition. The current research reveals that hand, head and shoulder movements are primarily considered for emotion recognition. Gesture not only provides the emotion of the subject, but it also offers a more spontaneous reaction of the subject to a stimulus and thus helps users to easily visualize the subject’s emotion. While only gesture may sometimes mislead the recognition process, gestural features with psychological parameters together provide a strong measure of the degree of emotion that a subject experiences during arousal of emotion.

Facial expression analysis undertaken here considers only static features of the face and ignores the dynamic features in face video because of movements of the muscles. These movements can be regarded as the collective response of several facial action points, and thus the dynamic features could improve the classification accuracy by a great margin.

In our research, we considered recognition of emotion from 2D facial expressions. Sometimes, 2D information is not adequate to extract facial features. One approach to overcome the problem is to recognize emotion from 3D facial expression of the subjects. The 3D facial expression is expected to provide more reliable and accurate features, thereby reducing the chances of false classification.

Further, to keep the recognition system reason like humans, the IT2FS based model introduced in Chapter 2 could be extended with Computing with Word [13] model. The computing with word model requires a word description of the facial features as input, which later is to be transformed to membership functions to appropriately use IT2FS based reasoning mechanism to infer the emotion of the subject.

In typical IT2FS based system, the footprint of uncertainty (FOU) is determined by the upper and lower membership functions. In a recent chapter, we have shown that the FOU in IT2FS can be reduced by
redefining upper and lower membership functions as the maximum and minimum of the average and modal distributions of all possible embedded fuzzy sets. Because of reduced degree of uncertainty, the uncertainty in classification accuracy of emotion is also reduced. Several extensions to the basic framework of type-2 uncertainty reduction can be taken up to improve the classification accuracy in the emotion recognition further.

Lastly, the thesis is concerned with recognition of one of six basic emotions of the subject. However, on occasions, the subject experiences mixed emotions. Unfortunately, there are fewer techniques [14] for mixed emotion recognition, and there remains much scope take initiative in this regard.

5.3. POSSIBLE APPLICATIONS OF THE PROPOSED RESEARCH

The thesis aimed at designing novel schemes for emotion recognition with an aim to supplement the current Human-Computer Interface (HCI) systems. Most of the current generation computers are equipped with cameras to capture facial images of the subjects. So, facial expression analysis by current generation HCI requires extension at software level to classify an acquired facial expression into one of several emotion classes. With additional audio attachment, such as microphone and audio card including Analog-to-Digital Converters, we can recognize emotion from voice data of the subjects. Lastly, any bio-potential sources including EEG and ECG can be interfaced with a computer to recognize emotion from bio-potential signals.

Besides HCI, another interesting application of emotion recognition lies in psychotherapy. Researchers have attempted to study the influence of music and video to control emotion of subjects [3]. In one recent paper, the authors examined control of emotion using ultrasonic sources [15]. There exists ample scope of research to examine the role of music and video on emotion control.

In [1], the authors considered a new approach to match parameters of emotional dynamics of two subjects to identify their team spirit. They also considered the parameter matching scheme in automatic matrimonial counseling. For example, certain parameters like onset time (representing the time required for arousal of emotion after submission of the stimulus), and offset time (representing the time duration the emotion sustains even after removal of the stimulus) have a great impact in determining emotionally matched marriage partners. For example, during arousal of anger of a couple, a large onset time of at least one of the partners and a small offset time for at least one of them is preferred. On the other hand, for happiness and relaxation, a small onset and large offset time for the couple is preferred. Currently, one of our contemporary research groups at Jadavpur University is taking keen interest to accurately measure onset and offset time of emotional dynamics using EEG analysis. We hope that once these measurements are accurate, we would be able to use them for matrimonial and other sociological applications, where team spirit among members is a necessity.

References


