Emotion recognition : from physiological signals to affective states

Gu, Yuan

2014

Gu, Y. (2014). Emotion recognition : from physiological signals to affective states. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/61879>

<https://doi.org/10.32657/10356/61879>

Emotion Recognition: From Physiological Signals to Affective States

Gu Yuan

School of Computer Engineering

A thesis submitted to Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy

October 2014

Acknowledgements

First, I would like to express my sincere thanks to all people who have helped me and inspired me during my whole PhD study.

Especially, I am deeply grateful to my supervisors Dr. TAN Su-Lim and Dr. WONG Kai-Juan, for their continuous support, endless encouragement, invaluable comments through my entire Ph.D study and research. Besides my supervisors, I would also like to thank Dr. QU Li and Dr. Ringo HO for their valued guidance with respect to the physiological experiments. This dissertation is simply impossible without them.

I also wish to express my warm and sincere thanks to all my lab buddies and technicians at the Centre for High Performance Embedded Systems (CHiPES) for providing great environment and excellent facilities to pursue my graduate study. They also make it a convivial place for to study and work. My special thanks goes to the lab mates, CUI Jin, CHEN Yupeng, and LIN Lin, for their kind support and assistance during my Ph.D study .

My deepest gratitude goes to my parents for their unflagging love, support and encouragement throughout my life and the whole course of my studies; and to my husband, WANG Zhimin for his unconditional love, understandings and diligent help. I would not have been able to complete my Ph.D study without many years of support from him.

Summary

In recent years, the rapid growth of human computer interaction research has accelerated the improving research interest in emotion recognition field. Emotion recognition in humans is one of the most important research areas of human interaction. This technology is useful in many disciplines, such as neuropsychology, artificial intelligence, human computer interaction, signal processing, affective computing, ethology, military technological R&D, and image processing.

There are many crucial steps in emotion recognition using physiological responses. These include data collection, signal preprocessing, feature extraction/selection, classification strategy, and classification result analysis. Among these issues, the feature selection and classification steps are especially important. On the other hand, the importance of data collection and signal preprocessing steps should not be underestimated. Well designed and robust data collection method will help to preserve the data accuracy and consistency, so that the real world events or objects are correctly described by these collected data. Good signal preprocessing methods will further enhance this relationship. This thesis focuses on the development of accurate, robust and generalized emotion recognition systems which take into account all the above mentioned issues.

One of the key contributions of this thesis is the development of the Biometric Signature Based (BSB) System. The proposed system uses a novel framework to eliminate the influence from the “individual varieties” occurred at the conventional subject-independent procedures. The additional subject classification process will transform a subject-independent emotion recognition problem into several subject-dependent tasks, which will help to improve the recognition rate by identifying each

individual prior to the recognition phase. There are two phases in this method. In the 1st stage, data from each subject is trained into separated statistical subject models. A new incoming sample will be classified to one subject model that best suits its inner structure. In other words, this data sample shows the most similar characteristics to the assigned statistical model. Then in the 2nd stage, a general feature selection plus classification procedure is applied to achieve the task of emotion recognition. The experimental results showed that, the BSB system performs more effective than using conventional methods. In addition, through a mutual test experiment, the system shows a general usage in that it is not required for the users to be trained by the system, as long as there are enough representative subject models stored beforehand.

We also extend the concept of BSB system in a more robust way. A new robust version of BSB system has been introduced and derived by using the multivariate t-Distribution model. Compared to the original Guassian mixture model (GMM) version, the new BSB system has advantages in providing better recognition rates and is able to simplify the model complexities of the original one.

We found out that it is not required for the system to prepare statistical models for each individual subject, as long as there are enough representative models known (trained) by the system. This leads to another key contribution of this thesis, an improved version of the BSB system, called “The Self-adaptive BSB system (SaBSB)”. Instead of directly creating several models based on individual subjects, the self adaptive procedure first assumes the whole data pool as one single statistical model, and then successively splits the model into two new ones, until the model number reaches a pre-defined value. Hence, the total number of statistical models in the first stage is reduced, also the models are generated in a more adaptive and flexible manner. By comparing SaBSB with conventional BSB and robust BSB, the results show that SaBSB achieves relatively comparable

results but requires less constraint in the number of subject models to be built.

Besides the efforts on the development of adaptive, robust and generalized emotion recognition systems, we also spend lots of time and efforts on experiment design, data collection, and raw data preprocessing. They are the important and indispensable parts of this research work.

Table of Contents

Acknowledgements	i
Summary	ii
List of Figures	x
List of Tables	xiii
Author's Publications	xv
1 Introduction	1
1.1 Motivations	3
1.2 Objectives	5
1.3 Contributions	5
1.4 Organization of the thesis	8
2 Literature Review	11
2.1 Introduction	11

2.2	Theoretical background in psychological studies	12
2.2.1	Basic Emotions	12
2.2.2	Continuous Dimensions	14
2.3	Relevant knowledge of emotion elicitation	15
2.3.1	Selected IAPS images in this study	17
2.4	Related Studies	17
2.5	Conclusion	23
3	Experimental Design for Data Collection	24
3.1	Introduction	24
3.2	Experimental Settings	25
3.2.1	Experimental elicitation	25
3.2.2	Hardware acquisition tools	26
3.2.3	Software Materials and Questionnaires	28
3.3	Experimental Procedure	30
3.4	Conclusion	31
4	Signal Preprocessing and Feature Extraction	33
4.1	Introduction	33
4.2	QRS detection of the ECG signals	34
4.3	Feature Evaluation	35

4.4	Conclusion	38
5	Conventional methods for emotion classification	41
5.1	Introduction	41
5.2	Preliminaries	42
5.2.1	Feature reduction and selection algorithms	42
5.2.2	Selected classification algorithms	49
5.3	Performance comparison of selected feature reduction/selection methods	51
5.3.1	Results comparison of 2ES classification	54
5.3.2	Results comparison of 4ES classification	63
5.3.3	Summary	64
5.4	Recognition results with/without SFFS & GA	65
5.4.1	Results without feature selection	65
5.4.2	Results with feature selection	68
5.4.3	Summary	72
5.5	Discussion on robust features	73
5.6	Conclusion	78
6	The Biometric Signature Based (BSB) System	80
6.1	Introduction	80

6.2	Conventional methods for emotion recognition	82
6.3	Methodology of the BSB system	84
6.3.1	Stage 1: subject classification process	86
6.3.2	Stage 2: emotion recognition process	89
6.4	Experimental results	89
6.4.1	Effectiveness and robustness	91
6.4.2	General usage of the system	97
6.5	Robust BSB system based on multivariate t-Distribution	104
6.5.1	Introduction	104
6.5.2	Stage 1: robust subject classification processing using multivariate t-Distribution	105
6.5.3	Testing phase and emotion recognition process	107
6.5.4	Experimental results	108
6.6	Conclusion	109
7	The Self-adaptive BSB System	113
7.1	Introduction	113
7.2	The self-adaptive BSB (SaBSB) system	114
7.2.1	Bayesian information criterion (BIC)	114
7.2.2	Self-adaptive procedure	115

7.3	Experimental Results	116
7.4	Conclusion	123
8	Conclusions and Recommendations for Future Research	125
8.1	Conclusions	125
8.2	Recommendations for Future Research	130
	Appendix A	132
	References	139

List of Figures

2.1	Thayer's valence-arousal emotion plane	15
2.2	Rating distribution (valence-arousal) of the IAPS pictures used in the experiment	19
3.1	Example pictures from the IAPS	25
3.2	System setup for the experiment	27
3.3	Experimental procedure of the actual test session	32
4.1	The QRS complex within the ECG signal	34
4.2	Illustration of the QRS detection	36
4.3	Process of feature extraction	39
5.1	Schematic illustration of crossover and mutation	48
5.2	Categories of affective states on 2D emotion model	53
5.3	Results comparison of Exp I with subject-independent approach using combinational methods	55

5.4	Results comparison of Exp II with subject-independent approach using combinational methods	56
5.5	Recognition rates of valence and arousal against feature dimensionality based on SFFS: Exp I	57
5.6	Recognition rates of valence and arousal against feature dimensionality based on SFFS: Exp II	58
5.7	Results comparison of Exp I with subject-dependent approach using combinational methods	61
5.8	Results comparison of Exp II with subject-dependent approach using combinational methods	62
5.9	Comparison of valence and arousal recognition rates without dimensional reduction using kNN, dw-kNN and fkNN	66
5.10	Comparison of valence and arousal recognition rates using kNN, SFFS-kNN, fkNN, and SFFS-fkNN	69
5.11	Comparison of results with dimensional reduction	70
5.12	Results of GA-kNN and GA-LDF with various parameter settings .	71
5.13	Recognition rates of valence and arousal against feature dimensionality based on SFFS	74
5.14	Summary of each feature chosen in SFFS-kNN	76
5.15	Summary of each feature chosen in SFFS-LDF	77
5.16	Block diagram of the general emotion recognition system	78
6.1	Block diagram of the conventional emotion recognition system . . .	83

6.2	Architecture of the biometric signature based emotion recognition system	84
6.3	Stage 1 of the BSB system	85
6.4	Categories of affective states on 2D emotion model	90
6.5	Result of experiment 1	94
6.6	Results of experiment 2	96
6.7	Results of the mutual tests: exp 2 test models of exp 1	102
6.8	Results of the mutual tests: exp 1 test models of exp 2	103
6.9	Comparison of results using three kinds of implementations for the 4EQs recognition	110
6.10	Results of subject classification using data of exp 1 and 2 for 4EQ recognition	111
7.1	Demonstration of the self-adaptive process	119
7.2	Comparison of results using four kinds of implementations for the 4ES recognition	120
7.3	Comparison of results with different number of clusters generated by the SaBSB	121
8.1	Block diagram of the general emotion recognition system	125
8.2	Block diagram of the improved emotion recognition system	127

List of Tables

2.1	IAPS: Ratings of selected pictures	18
4.1	Data distribution of the 1st experiment	40
4.2	Data distribution of the 2nd experiment	40
5.1	Results of subject-independent vs. subject-dependent using SFFS-kNN	63
5.2	Comparison of results from Exp I and II for 4ES recognition	64
5.3	Recognition rates of valence and arousal without dimensional reduction	67
5.4	Recognition rates with dimensional reduction using SFFS	68
5.5	Recognition rates using GA-based dimensional reduction	72
6.1	Data distribution of the 1st experiment	92
6.2	Data distribution of the 2nd experiment	92
6.3	Results comparison of BSB and SFFS-kNN: experiment 1	98
6.4	Results comparison of BSB and SFFS-kNN: experiment 2	99

6.5	Results comparison of BSB and SFFS-kNN: experiment 1 and 2 . . .	100
6.6	Comparison of Results	104
7.1	SaBSB: Number of clusters for each round of validation	119
8.1	Primary parameters in an emotion recognition study	126
A.1	Results of valence recognition without dimensional reduction using kNN, dw-kNN and fkNN	133
A.2	Results of arousal recognition without dimensional reduction using kNN, dw-kNN and fkNN	134
A.3	Summary of each feature chosen in SFFS-kNN for valence recognition	135
A.4	Results of valence and arousal recognition using SFFS-kNN and SFFS-fkNN	136
A.5	Summary of each feature chosen in SFFS-kNN for arousal recognition	137
A.6	Summary of each feature chosen in SFFS-LDF	138

Author's Publications

- P1. Yuan Gu, Su-Lim Tan, “Biometric Signature Based Emotion Recognition”, submitted to International Journal of Biomedical Engineering and Technology, 2014.
- P2. Yuan Gu, Su-Lim Tan, Moon-Ho Ringo Ho, Li Qu, and Kai-Juan Wong, “Emotion Awareness For Consumer Electronics Through Physiological Signals”, submitted to Advances in Human-Computer Interaction, 2014.
- P3. Yuan Gu, Kai-Juan Wong, Su-Lim Tan, “Analysis of Physiological Responses from Multiple Subjects for Emotion Recognition”, e-Health Networking, Applications and Services (Healthcom), 2012 IEEE 14th International Conference on, pp. 178-183, Oct 2012
- P4. Yuan Gu, Su-Lim Tan, and Kai-Juan Wong, “Self-Adaptive Biometric Signature Based Emotion Recognition System”, the 4th International Conference on Applied Human Factors and Ergonomics, July 2012.
- P5. Yuan Gu, Su-Lim Tan, Moon-Ho Ringo Ho, Li Qu, and Kai-Juan Wong, “A Biometric Signature based System for Improved Emotion Recognition using Physiological Responses from Multiple Subjects”, in 8th IEEE International Conference on Industrial Informatics, pp. 61-66, Jul 2010.
- P6. Yuan Gu, Su-Lim Tan, Moon-Ho Ringo Ho, Li Qu, and Kai-Juan Wong, “A GMM based 2-stage architecture for multi-subject emotion recognition using physiological responses”, in Proceedings of the 1st Augmented Human International Conference, pp. 3:1–3:6, 2010.

- P7. Yuan Gu, Su Lim Tan, Kai Juan Wong, Moon-Ho Ringo Ho and Li Qu, “Using GA-based feature selection for emotion recognition from physiological signals”, International Symposium on Intelligent Signal Processing and Communication Systems, pp. 1-4, Feb 2008.
- P8. Yuan Gu, Su Lim Tan, Kai Juan Wong, Moon-Ho Ringo Ho and Li Qu, “Emotion-aware technologies for consumer electronics”, IEEE International Symposium on Consumer Electronics, pp. 1-4, April 2008.

Chapter 1

Introduction

Emotion plays a critical role in people's daily lives. Various findings from neuroscience, psychology, and cognitive science suggest that emotion can influence an individual's learning, decision making, behavioral control, and social interaction [1, 2]. As computer and computer-based applications are becoming more sophisticated and common nowadays, equipping these devices with some form of emotional characteristics will facilitate more intelligent interaction with humans.

In the mid 1990s, Dr. Picard proposed the initial concept of affective computing, which was defined as “computing which relates to, arises from, or deliberately influences emotions” [3]. In a later publication [1], she emphasized the needs of adding emotions as new features to human-computer interactions in order to achieve more natural and intelligent communications. Making computer aware of human's affective state becomes a key prerequisite to the development of affective computing system.

Though emotion is an individual's personal feeling, it also includes physical and cognitive components [1]. Therefore, it is possible to recognize emotional states based on certain physical “languages” such as facial expressions [4, 5], speech [6–9], and physiological signals [1, 2]. Behavioral changes in eye gazing, facial expressions,

gestures, speech are relatively easier to recognize. Physiological changes such as variations of heart beat, breathing or sweating are relatively difficult to analyze mainly because it is challenging to annotate physiological data as a universal data set in terms of emotion states. Compared to other behavioral changes, physiological patterns are not that straightforward when labeled onto specific emotion [10].

However, there are two unique features that make the approach of physiological signals more advantageous compared to visual-audio ways [2, 10]. Firstly, physiological signals are controlled by the human autonomous nervous system. The reactions are more spontaneous and robust compared to facial expressions or voices that can be easily disguised on purpose. There exists a phenomenon of “display rules” that limits the range of acceptable facial expressions. For example, in business or in some social settings, usually it is quite inappropriate for a business man to show extreme disgust or anger during a negotiation process. In some countries like Japan, subordinate usually appears reluctant to display his true feelings on face when meeting with superiors [1, 11]. Secondly, certain emotional states are constraint by means of expression such as interest or depression, which can be hard to present in audiovisual ways. Psychologists deal with a lot of depression cases from teenagers to adults. In most cases, those people are unwilling to share their thoughts or even have difficulties expressing their feelings. Physiological pattern recognition can potentially aid in assessing and quantifying those kinds of emotions without unnecessary irritation or frustration. Furthermore, with the rapid advancement of wearable sensing devices, measuring physiological signals become rather sophisticated and convenient nowadays. Imagine an “affective detector” which is a wearable device (e.g. watch) capable of detecting and evaluating the severity of depression. This portable equipment could continuously monitor the emotional status of the patient and log the moments when negative emotions are generated. So the psychological clinician in charge is able to master the whole picture and help the treatment. Or thinking about an affective play system, which

can adjust the conditions of the game, such as lowering down the difficulty of the level or switch the scene to a less stressful one based on the emotional states of the player. These potential applications show the merits of using physiological signals as the channel for affective understanding.

1.1 Motivations

For years, research groups have put their efforts in developing a generally applicable emotion recognition system based on the use of physiological signals. Basically, the complete system can be divided into three major steps: (1) data collection from the designed experiment, (2) signal preprocessing and feature extraction, and (3) feature selection/transformation and classification. Each technical step plays a different role in the emotion recognition system, and all of them are important steps in the whole emotion recognition process. In our specific experiment, the whole process of an emotion recognition system will usually consist of the following three steps:

1. Data collection: With the supervision and advice of professionals from our psychology department, two experiments of using 6 channels of bio-sensors and 28 pictures from International Affective Picture System (IAPS) are set up. The raw data entries from the 6 channels of bio-sensors are collected and stored using hard disk.
2. Signal preprocessing and feature extraction: Raw data entries from 6 channels of bio-sensors including ECG, Blood volume pulse (BVP), Skin-conductance (SC), EMG (one for zygomatic EMG (EMGz) and the other for corrugator EMG (EMGc)), and Respiration sensor (Rsp), from the data collection step are collected. A signal preprocessing process is performed in this step. First, the heart rate (HR) information from the ECG data is computed. After this,

6 statistical time-domain features from each signal channel are extracted, i.e. HR, BVP, SC, EMGz, EMGc and Rsp, and form a total of 36 statistical time-domain features vector for each raw data entry.

3. Feature selection/transformation and classification: The extracted 36 statistical time-domain features could contain irrelevant and redundant features, and may affect the efficiency of the learning process. In order to overcome this problem, a feature selection/reduction procedure will be performed in this step to increase the prediction accuracy of the applied classifier. The last stage of this step is the classification and recognition result evaluation.

Among the many technical issues involved in the emotion recognition system, the feature selection and classification step is an especially important one. A good feature selection and classification method should be able to reveal the inherent structure of the input data and generate meaningful classification results for further analysis and evaluation. It also determines the performance and reliability of a emotion recognition system. On the other hand, the importance of data collection and signal preprocessing steps should not be underestimated. Well designed and robust data collection method, as well as good signal preprocessing procedures will help to preserve the data accuracy and consistency.

To this end, we can say that accuracy, consistency, universality, and robustness are the important factors when one evaluates a new emotion recognition system. However, there is no universal recognition technique that is always applicable for recognizing the variety of emotional structures present in the data sets. Therefore, the development of accurate, robust and generalized emotion recognition systems is the most important motivation of this thesis.

1.2 Objectives

The main objective of this thesis is to design and develop robust emotion recognition systems for physiological applications. We focus on several areas in this research, such as experimental design, comprehensive analysis on conventional methods for physiological signals based emotion recognition, and development of novel emotion recognition systems. In experimental design, the aim is to design multi-subject experiments for data acquisition. This design should be more useful for real applications. After the experimental design, comprehensive comparison and analysis of the performances of several commonly used feature reduction/selection methods, as well as investigating the diversity of results between subject-dependent and subject-independent approaches are conducted. In the development of novel emotion recognition systems, adaptive, generalized and robust methods are investigated and developed. The key objectives, can therefore be summarized as:

- (i) To design multi-subject experiments for data acquisition that is more useful than the single-subject design.
- (ii) To conduct comprehensive comparison and analysis on the feature reduction/selection methods in the literature.
- (iii) To develop and propose an adaptive, robust and generalized emotion recognition processes that is able to provide more accurate classification results.

1.3 Contributions

In this research work, two rounds of psychological experiments for effective data collection are designed and conducted. The data gathered are used to compare and analyze the performances of several feature reduction/selection methods, as well as classification algorithms in the literatures. In addition, the robust features

based on our experiments are summarized.

In order to resolve the problem of “individual varieties” when multiple subjects are involved, a novel framework called the biometric signatures based (BSB) emotion recognition system and a robust version are developed. Furthermore, Self-adaptive BSB (SaBSB) is proposed for improvement of the BSB to make the system more flexible in usage.

The major contributions of this thesis are listed as below.

1. **Experimental design for effective data collection**

In order to gather a fair amount of good affective data for future analysis, it is essential to design and conduct an appropriate experiment. With the supervision of professionals from the psychology department, two rounds of experiments are conducted, one of 50 subjects, and the other of 43 subjects. Each subject wears 6 channels of bio-sensors including ECG, BVP, SC, EMGz, EMGc, and Rsp. 28 pictures from the IAPS are selected for emotion elicitation. In all, we manage to gather 784 and 868 data entries respectively for experiment 1 and 2.

All the collected raw data are preprocessed before the feature extraction process. Instead of directly using the ECG signals, we extract the heart rate (HR) information from them. Along with the rest of five signals, BVP, SC, EMGz, EMGc and Rsp, 36 statistical time-domain features are calculated, with 6 features from each signal channel.

2. **Analysis on conventional methods for physiological signals based emotion recognition**

Data from two rounds of our experiments are used to compare and analyze the performances of several commonly used feature reduction/selection methods including Fisher projection (FP), Sequential Floating Forward Search

(SFFS) and hybrid SFFS with FP, in combination of some classification algorithms such as k-Nearest-Neighbor (kNN) Rule, Support Vector Machine (SVM), Linear and Quadratic Discriminant Function (LDF and QDF). The diversity of results between subject-dependent (single subject) and subject-independent (all signals mixed together for different subjects) approaches are also investigated. In addition, a series of robust features are statistically calculated based on our results.

3. The biometric signature based (BSB) system

A novel framework called the biometric signature based (BSB) emotion recognition system is proposed for solving the problem of “individual varieties” when multiple subjects are involved. This novel emotion recognition system employs a two stage processing framework for the task of emotion recognition. Firstly, the system classifies the whole training data set into several Gaussian mixture models (GMM), each of which represents one subject model. A new incoming sample is classified as one subject model using the Maximum Posterior Probability (MAP) rule to capture the intrinsic characteristics of the input data. By doing so, the original subject-independent problem is able to transform into several subject-dependent cases. So the second stage follows a conventional emotion recognition procedure. Compared with the usual structure, the BSB system eliminates the influence of the “individual varieties” therefore enhances the recognition performance of the whole system. In order to add robustness into our BSB system, another study is performed on the BSB system that utilizes the multivariate t-Distribution model resulting in a new robust BSB system. Compared to the GMM version, the new BSB system has advantages in providing better recognition rates and is able to simplify the model complexities of the original one.

4. The Self-adaptive BSB (SaBSB) system

A finding from our study shows that it is not required for the system to prepare statistical models for each individual subject, as long as there are enough representative models known (trained) by the system. Hence, the first stage of the BSB system is improved by using a self-adaptive model generating method and a modified BSB system is proposed, called the Self-adaptive BSB (SaBSB) system. The SaBSB system is different from the original BSB system in that, instead of directly creating several models based on individual subjects, the self adaptive procedure first assumes the whole data pool as one single statistical model, and successively divides the model into two new ones based on their statistical structures, until the model number reaches a pre-defined value. By doing so, the models in the first stage are built in a more flexible way rather than being limited to individual subject. With this new method, the total number of statistical models in the first stage is reduced, also the models are generated in a more adaptive and flexible manner. Compared to the BSB system, the SaBSB requires less constraints in the number of subject models to be built.

1.4 Organization of the thesis

The rest of the thesis is organized as follows:

Chapter 2 gives a detailed literature review on defining emotions, experimental methods for emotion elicitation, as well as several related studies.

Chapter 3 introduces our experimental design. Details of the experimental architecture, the experimental settings, and our rating system will be presented in this chapter and the procedures of raw data collection will also be demonstrated.

Chapter 4 presents our signal preprocessing and feature extraction steps. In this chapter, the process of raw signal preprocessing and feature extraction are reviewed. With two rounds of experiment, 784 and 868 data entries respectively are gathered and all the following analysis in this thesis are based on these two data sets to ensure the fairness of comparison.

Chapter 5 reviews the conventional methods of emotion classification. Comprehensive comparisons on the feature search and classification methods are presented here. In this chapter, implementation and comparison of various combinations of the classification algorithms with or without using feature selection/reduction methods are conducted. Investigation of the diversity of results between subject-dependent and subject-independent approaches is also shown.

Chapter 6 presents an improved classification structure, the biometric signature based (BSB) system, which aims to solve the problem of “individual varieties” when multiple subjects are involved. The basic idea is to transform a traditional subject-independent problem into several subject-dependent cases (separate the mixed data samples into subject groups based on their biometric characteristics), then proceed to the regular emotion recognition process. In order to add robustness into our BSB system, another study is also performed on the BSB system that utilizes the multivariate t-Distribution model to derive the new robust BSB system.

Chapter 7 proposes another improved structure using a self-adaptive model generating method. It will be different from the original BSB system in that, instead of directly creating several models based on individual subject, the self adaptive procedure first assumes the whole data pool as one single statistical model, and successively divides the model into two new ones based on their statistical structures, until the model number reaches a pre-defined

value.

Chapter 8 concludes this thesis with a summary and some discussions on future work and research directions.

Chapter 2

Literature Review

2.1 Introduction

The study of human emotions can be quite profound. It can involve psychology, philosophy, social and behavioral sciences, biology and neuroscience [1, 12–15]. Dr. TenHouten [13] pointed out a three-level analysis for deriving a general theory of emotions, which encompasses (a) the biology and the neuroscience, (b) the mental and the psychological, and (c) the social and the cultural. He referred such an approach as neurocognitive sociology of the emotions. For us, we may just need some background knowledge of the first two levels.

This chapter is organized as below. First, we explain two most acknowledged systems for defining emotions in psychological studies: basic emotions and continuous dimensions. Based on these two models, we introduce two sets of emotion elicitation standards: the film clips of Gross and Levenson [16], also the International Affective Picture System (IAPS) [17]. After that, we give an overview of several related studies including how they conducted their experiments, the techniques they used for classification, also the corresponding results.

2.2 Theoretical background in psychological studies

Darwin, father of the theory of evolution, realized that evolution not only applied to anatomy and morphology but also to an animal's mind and expressive behavior [18]. Although his focus was on facial expression, he observed an evolutionary connection between animal and human behavior. For example, monkeys flush face with anger and human act the similar way [13]. Ever since Darwin, evolution-oriented theorists of emotions have believed that emotions are employed as a result of long environmental-driven adaptation to better serve the problems posed by the environment and thereby increase the chances of survival and reproduction [19]. Human ethologist Eibl-Eibesfeldt assumed that the universality of several emotional expressions is due to genetically-inherited "fixed action patterns" that vary across individuals and contexts [20]. Some other theorists such as Tomkins [14, 15], Izard [21, 22] and Plutchik [23] also brought forward similar conceptualizations in their emotions theories. In all, the evolution-oriented theory of emotion proposed the idea that emotions are universal, which provided a basic theoretical foundation for researchers to investigate a general methodology that is able to systematically categorize emotions. Two acknowledged systems for defining emotions are introduced below.

2.2.1 Basic Emotions

Theorists have had long discussion on the existence of a small set of privileged emotions usually called basic, primary or fundamental emotions [24–26]. Although many researchers have shared the views that some emotions are basic, few of them agreed on how many the basic emotions are, which ones and why they are basic [24]. Between the 3rd and 11th centuries, Hindu philosophers stated eight basic or nat-

ural emotions [27] translated as (a) sexual passion, love or delight; (b) amusement, laughter or humour; (c) sorrow; (d) anger; (e) fear or terror; (f) perseverance; (g) disgust; and (h) amazement. After that, many theorists gradually produced their own set of basic emotions, e.g. Descartes and Darwin [18]. In 1962, Tomkins brought up the idea of eight basic emotions, anger, interest, contempt, disgust, distress, fear, joy, shame and surprise [14, 15]. Later in 1980, Plutchik produced a set of eight basic emotions but with different components - fear, anger, sorrow, joy, disgust, surprise, acceptance and anticipation [28]. Ortony and others have collected summaries of lists regarding identifications of basic emotions [24, 29]. From these lists, the most common four emotions are fear, anger, sadness and joy (combining near synonyms like joy and happiness), and the next two are disgust and surprise [1].

Basic emotions may be defined in many ways due to the diversity of how to express emotions, e.g. using facial expressions or body movements. Perhaps the most thorough and well-accepted definition was proposed by Paul Ekman, who linked basic emotions to distinctive universal facial expressions [4, 5]. Under these criteria, the six basic emotions are identified as fear, anger, sadness, happiness, disgust and surprise. Ortony and Turner refused to accept the notion of “basic” or “primary” in defining emotions in that different scholars proposed various criteria for identifying basic emotions and listed quite different sets of basic emotions (refer to the diverse lists of basic emotions mentioned before). However they still agreed that it is reasonable to develop a research strategy to classify emotions in certain ways, for example based on distinctive facial expressions [24], as they believed that some emotions are universal while some of them are influenced by cultural elements.

2.2.2 Continuous Dimensions

Some scholars alternatively preferred to use continuous dimensions of emotions instead of basic emotions or discrete emotion categories in representing emotions [30–33]. Lang’s report showed that in terms of cross subjects, it is more reliable of using dimensions than discrete emotion categories [34].

A structure of three dimensions is commonly stressed, consisting of arousal, valence and dominance [35, 36]. However, the names of the first two are widely accepted, which are “valence” referring to positive vs. negative and “arousal” referring to calm vs. excited [1]. Lang has assembled an international archive of imagery called the International Affective Picture System (IAPS) which is rated in terms of arousal and valence [17, 37]. Several researchers developed variant versions of the continuous dimensional structure. Russell proposed a circumplex model which is a two-dimensional system coding with pleasure to displeasure and arousal to sleepiness [38]. Rolls’ theory of emotion is a similar two-dimensional model in neuroscience approach. The vertical axis presents emotions elicited by the presentation of a reward or punishment, and the horizontal axis shows emotions produced by the termination or omission of a reward or punishment [39, 40].

Mapping between continuous dimensions of emotions and basic emotion categories can be in several different ways [1, 41]. In her book “Affective Computing” [1], Picard listed several “cognitive appraisal” models that are able to partition continuous emotional space into some discrete emotions, such as the Ortony Clore Collins (OCC) Cognitive Model [29], the Em system for emotion generation of animated characters [42] and Roseman’s cognitive appraisal model [43]. A research group in the US conducted experiments to identify images from the IAPS that is able to elicit one discrete emotion more than others [41]. Figure 2.1 illustrates the emotion classes divided into the four quadrants in Thayer’s valence-arousal emotion plane [44].

Figure 2.1: Thayer's valence-arousal emotion plane (adapted from [44])

2.3 Relevant knowledge of emotion elicitation

To vary subjects' emotional states in a psychological process, experimental emotion induction is one of the known procedures. It induces emotional states experimentally by means of special emotion induction stimuli such as presentation of emotion-inducing material like picture, film, music clips and others [45]. Because of the possibility of standardizing the observations, emotion induction procedure has become the most rigorous method of changing emotional states in the laboratory or in the field. We introduced below two sets of generally defined emotion induction materials that are ever employed in previous emotion recognition studies. Notice that one set of the materials follows the model of basic emotions, and the other is generated based on continuous dimensions.

Gross and Levenson [16] in the psychology department in the University of California, Berkeley, investigated a series of film clips that reliably elicit eight emotional states (amusement, anger, contentment, disgust, fear, neutral, sadness and surprise). From a corpus of over 250 commercial films, 78 were selected for further evaluation by an ethnically diverse group of 494 English-speaking subjects. The

criteria of selection follows three rules: (1) length, has to be relatively short; (2) intelligibility, needs to be understood without explanation, and (3) discreteness, needs to elicit a single emotion from the eight emotional states being investigated. Out of the 78 films, two films for each of the eight target emotions resulting in the highest subjects' responses of intensity and discreteness were presented. Eventually, 16 films are chosen for eliciting amusement, anger, contentment, disgust, sadness, surprise, a relatively neutral state and to a lesser extent, fear.

The other set of material, the IAPS [17], is an archive of imagery as affective stimuli for psychological research. Developed and distributed by the NIMH Center for Emotion and Attention (CSEA) at the University of Florida, the IAPS aims to provide standardized materials that are available to researchers in the study of emotion and attention. It covers a broad set of emotion-eliciting color pictures depicting mutilations, snakes, insects, attack scenes, accidents, contamination, illness, population, landscape scenes and many others. These images are collected from different sources such as newspapers, books and magazines. By conducting a series of experiments with groups of participants, these pictures are rated according to a dimensional model of emotion (valence, arousal and dominance) [46]. Though the IAPS is originally proposed under the three dimensional structure, the first two coordinates (valence and arousal) are widely accepted by related studies [1].

It is advisable to use these kinds of standardized emotional stimuli to help eliciting subjects' emotional states, especially for experiments conducted on multiple subjects. Since the materials themselves are generated from groups of subjects following professional (psychological defined) process, we can assume that they have reached a unified level of inducing certain degrees of emotional reactions. Some studies however, choose to use subject-specific materials for emotion elicitation, e.g. the guided-imagery used by Picard et al. [2]. It is still scientifically reasonable as long as the experiment is designed for this particular single subject.

2.3.1 Selected IAPS images in this study

This study selected 28 pictures from the International Affective Picture System (IAPS) [17] based on the criteria that the distribution of ratings along pleasant/unpleasant (valence) and excited/calm (arousal) should be relatively balanced.

Since it is still unclear whether people from different cultural background would respond to the same emotional stimuli similarly, instead of following the original IAPS ratings from Lang's experiments, participants are required to rate pleasantness and emotional intensity in a nine-point scale. Table 2.1 displays the original IAPS ratings for the selected 28 pictures, which are plotted in affective space with arousal-valence coordinates in Figure 2.2 marked with small circle. For comparison, the same figure also plots the rating results from our experiments, which are marked with asterisk. It shows that the arousal ratings suffer a distinct variance between the original ratings and the experimental ratings while valence ratings stay in the same pleasant/unpleasant areas. This phenomenon could be due to culture differences between western people and Asians. Professionals from psychology department who have joined our study will analyze this further. Upon such case, the next step of data analysis is based on the experimental ratings rather than the original ones to ensure that the affective labels are consistent with the feelings of the subjects.

2.4 Related Studies

Dr. Picard and colleagues [1–3] conducted their experiment following a single-subject multiple-day manner, meaning to ask one single subject to repeatedly go through the same experimental procedure over many weeks of time to reach a relatively large amount of data. A computer controlled prompting system called the “Sentograph” developed by Clynes [47], and a set of personally-significant imagery

Table 2.1: IAPS: Ratings of selected pictures

	Slide No.	Valence Mean	Valence SD	Arousal Mean	Arousal SD
Snakes	1111	3.25	1.64	5.2	2.25
Snake	1120	3.79	1.93	6.93	1.68
Spider	1205	3.65	1.76	5.79	2.18
PolarBears	1441	7.97	1.28	3.94	2.38
Puppies	1710	8.34	1.12	5.41	2.34
Shark	1932	3.85	2.11	6.47	2.2
Baby	2053	2.47	1.87	5.25	2.46
Hospital	2205	1.95	1.58	4.53	2.23
Fingerprint	2206	4.06	1.4	3.71	2.03
Children	2216	7.57	1.31	5.83	2.2
Butcher	2235	5.64	1.27	3.36	1.92
Woman	2375.1	2.2	1.31	4.88	2.21
Man	2500	6.16	1.54	3.61	1.91
Couple	2530	7.8	1.55	3.99	2.11
CryingBoy	2900	2.45	1.42	5.09	2.15
Mountains	5600	7.57	1.48	5.19	2.7
Leaves	5800	6.36	1.7	2.51	2.01
Attack	6350	1.9	1.29	7.29	1.87
Umbrella	7150	4.72	1	2.61	1.76
Lamp	7175	4.87	1	1.72	1.26
IceCream	7330	7.69	1.84	5.14	2.58
MeatSlicer	7361	3.1	1.73	5.09	2.48
FerrisWheel	7508	7.02	1.46	5.09	2.11
Skier	8030	7.33	1.76	7.35	2.02
Rollercoaster	8499	7.63	1.41	6.07	2.31
SportCar	8510	7.32	1.72	4.93	2.56
Exhaust	9090	3.56	1.5	3.97	2.12
Exhaust	9090	3.69	1.92	4.8	2.23
Dog	9570	1.68	1.23	6.14	2.31

Figure 2.2: Rating distribution (valence-arousal) of the IAPS pictures used in the experiment: Original ratings are marked with “o” while experimental ratings are marked with “*”. Arrow starts from the point indicating the original ratings to the point indicating the experimental ratings in each picture, whose number is labeled at the start of each arrow.

were used to help elicit eight affective states including no emotion, anger, hate, grief, platonic love, romantic love, joy and reverence. Four types of bio-sensors from the Thought Technologies ProComp unit, a triode electromyogram (EMG), a photoplethysmograph measuring blood volume pressure, a skin conductance sensor (SC) and a Hall effect respiration sensor (RSP) were employed. Since the ProComp automatically computed the heart rate (HR) as a function, in all there were five physiological signals collected from the experiment. The overall recognition accuracy was 81 percent for eight emotions (no emotion, anger, hate, grief, platonic love, romantic love, joy and reverence) using hybrid linear discriminant classification.

Haag et al. [48] conducted a similar study also with one single subject for data collection. Alternatively, they chose to use the IAPS for emotion elicitation with ratings of valence and arousal for labels of emotional states. Six kinds of bio-sensors, the electrocardiogram (ECG), blood volume pulse (BVP), SC, EMG, RSP and temperature (TMP) sensors were used. The valence and arousal dimensions were classified separately using a neural network classifier. The recognition accuracy rates reached 89.9 percent and 96.6 percent respectively.

Instead of experimenting with one single subject, Nasoz et al. [49] recruited several participants for their data collection. First, 14 students were involved with a pilot panel study to examine the effectiveness of some selected movie clip for emotion elicitation of the subsequent experiment. With 7 movies reported from Gross and Levenson's work and additional 14 movies suggested by themselves, finally they managed to determine five movie clips with the highest level of agreements. Each clips targeted for inducing one of the following emotions: sadness, anger, amusement, fear, and surprise. In the emotional data collecting session thereafter, 29 subjects were enrolled. Each of them was attached with a non-invasive SenseWear armband to collect signals from GSR (galvanic skin response), HR and TMP. They

managed to get a maximum recognition rate of 83 percent for all the 6 emotions using Marquardt Backpropagation algorithm (MBP).

More recently, Wagner et al. [50] reported their attempts on one single subject using personal selected music songs in respect of four targeted emotion classes, joy, anger, sadness and pleasure. Four-channel bio-sensors are used to record EMG, ECG, SC and RSP. A later publication from the same group, which was written by Kim and André [10], reported some following work by conducting the same experiment on three subjects. 110 expanded features were employed and a pseudo LDA (pLDA) with Sequential Backward Search (SBS) was applied. The final result for each subject exceeded 95 percent but only 70 percent across all the subjects. Besides, the paper obtained a set of best emotion-relevant features extracted from the four-channel physiological signals.

Van den Broek and others reviewed 20 studies in their publication [51]. In addition, they provided their own research with EMG and SC recordings of 21 people. By testing a range of techniques including ANOVA, PCA for feature selection, as well as kNN, SVM and Artificial Neural Networks (ANN) for classification, eventually they received 61.31% correctness for four emotion classes (neutral, positive, mixed and negative).

From the reviewed studies, several of them [2, 48, 50] only focused on collecting data from one single subject. Picard et al. [2] requested their participant to repeatedly experience eight affective states at roughly the same time of day for many weeks. Haag et al. [48] gathered data from one individual on different days and different times of the day. By designing their experiments in the above manner, they managed to obtain enough data for further analysis although the scope was limited to one subject.

The main reason for conducting experiments with one subject is to avoid problems

caused by variances across multiple subjects (so called “individual varieties”). Here with respect to the study of emotion recognition using physiological signals, individual differences could behave in two ways. First, an acknowledged research discovery made by Ekman and his colleagues says that even simply labeled emotions like “joy” and “anger” can be interpreted differently across individuals within the same culture [2]. Therefore it complicates the signal processing and classification procedures when the goal is to examine whether subjects elicit similar physiological patterns for the same emotion. Secondly, to say the least, supposing multiple subjects can be expected to interpret and react psychologically consistently, physiological signals gathered from individuals could vary significantly [2]. By limiting experimental subject to one single person, these variability could potentially be removed.

Single-subject (subject-dependent) approach might be considered in the early stages of affective recognition since it is valuable to develop person-dependent methods [2]. However, it is foreseen that the precise features and recognition results obtained from one single person may not be the same for other subjects. In other words, the classification algorithms which could be efficient with one particular data set might not be that useful on others. Hence, the weakness of single-subject approach is that the developed recognition methods are not generally applicable.

Therefore in our study, we decide to recruit multiple subjects in participating our experiments. By doing so, we can have the opportunity to examine the results in either subject-dependent or subject-independent (mixed data from multiple-subject) ways.

2.5 Conclusion

In this chapter, we introduce some relevant background knowledge from models of defining emotions to the methods of emotion elicitation.

A proper conducted experiment for data collection is very important for the succeeding analysis. For initial study of emotion, it is rather necessary to avoid complicate interpretation of a measured affective response. Therefore compared to those kinds of affective stimuli such as films, stories, etc. which contain movements, narrative scenes or other dynamic changes, static pictures are more favorable in that sense. Hence, we choose to use standardized emotion elicitation method, pictures from the IAPS database for our experiments. And the experiments are designed with multiple subjects for flexible analysis.

Chapter 3

Experimental Design for Data Collection

3.1 Introduction

In order to gather a fair amount of good affective data for future analysis, it is essential to design and conduct an appropriate experiment. As reviewed in Chapter 2, several issues have to be decided at the beginning of the experimental design. First of all, we have determined to experiment on multiple subjects so that the data can be used for both single-subject and multiple-subject approaches. Therefore, we need a consistent environmental setting employing the same data acquisition instruments to ensure the consistency on all collected data. Also, the emotional stimulus must be generally applicable and standardized by professionals.

Taking into account of all the considerations above, we conducted two rounds of experiments for data collection. Both experiments were supervised by professionals from the psychology division, School of Humanities and Social Sciences, Nanyang Technological University, to ensure that the process of emotion elicitation was ethically and properly conducted. In total, 93 (50 and 43 for two experiments) healthy subjects older than 18 years of age participated in the experiments so as to

(a) IAPS No.1441

(b) IAPS No.8030

Figure 3.1: Example pictures from the IAPS

fulfill part of their course requirement. Each subject wore 5 types of bio-sensors for about 40 minutes to complete the process of the experiment. All data collection was carried out in the same environment using the same sets of equipment. The details are explained as below.

3.2 Experimental Settings

3.2.1 Experimental elicitation

For emotion elicitation for both rounds of experiments, 28 pictures were selected from the International Affective Picture System (IAPS) [17] based on the criteria that the distribution of ratings along pleasant/unpleasant (valence) and excited/calm (arousal) should be relatively balanced. Figure 3.1 shows two example images used in our experiments.

3.2.2 Hardware acquisition tools

Figure 3.2 illustrates the experimental setup for collecting the physiological signals during the experiment. The equipment includes:

1. Two desktop computers: one was used for presenting the stimuli to the subject and the other was attached to the data acquisition device (the video camera only used in the first round of experiment).
2. One set of ProComp Infiniti system from Thought Technology with five types of bio-sensors was used for bio-signal acquisition. This system provides 8 optically isolated channels for recording with two high speed channels at 2048 Hz and six lower speed channels at 256 Hz.

The bio-sensors used are electrocardiogram (ECG), blood volume pulse (BVP), skin conductivity (SC), electromyography (EMG) and respiration (Rsp) sensors.

1. Electrocardiogram (ECG): The ECG signal is used to measure the contractile activity of the heart. In this experiment, it was collected using a 3-electrode sensor attached to the chest. There are several useful features that can be extracted from this signal, for example, heart rate variability (HRV). Past study had shown that low HRV can indicate a state of relaxation where as a high HRV can indicate a potential state of mental stress or frustration [52].
2. Blood volume pulse (BVP): BVP measures the instantaneous amount of blood running through the blood vessels under the skin. The sensor is held against the surface of a fingertip with an elastic strap. BVP signal can also provide heart-related signals like heart rate, HRV or inter-beat intervals (IBI).

Figure 3.2: System setup for the experiment

3. Skin conductivity (SC): SC is a measure of the skin's ability to conduct electricity. It is recorded by strapping two electrodes onto two different fingers on one hand. SC is a good indicator for emotional states that may cause sweating such as stress, fear or anger [2, 48].
4. Electromyography (EMG): EMG measures muscle activity or frequency of muscle tension. This study focused on corrugator (EMGc) and zygomatic EMG (EMGz) which indicated facial activities when the subjects were viewing the pictures. Two EMG sensors are placed on the left side of the subject's face, with one over the eyebrow and the other one on the cheek.
5. Respiration sensors (Rsp): The respiration signal is a relative measure of chest expansion and contraction during respiration. The sensor is strapped around a subject's chest with a rubber band. The speed and extent of breathing could be affected by emotions such as anger or fear [2, 48].

3.2.3 Software Materials and Questionnaires

In the 1st round of experiment, two application softwares, MediaLab and DirectRT from Empirisoft were used to design and run the experimental procedures. MediaLab shows great flexibility and ease of use with respect to questionnaires, while DirectRT is excellent in timing precision. We used another psychological software called the Eprime as the substitute for MediaLab and DirectRT in the 2nd round of experiment.

Besides the computer-based emotional induction test, participants were also requested to fill in some questionnaires.

- *The Positive and Negative Affective Scales (PANAS)* [53] is a widely used questionnaire used to assess emotional states and traits. It consists of 20 ad-

jectives that described different feelings and emotions, such as happy, upset, excited, etc. Subjects were asked to indicate to what extent they felt at that moment with respect to these emotional descriptors on a 1 to 5 scale (1-not at all, 5-very much). In our experiment, PANAS was designed to appear before and after the emotional induction task.

- *A basic information sheet* includes questions about health, language ability and culture orientation.
- *Cultural Identification Questionnaire* [54] is an 18 items questionnaire through which individual identifies his/her culture.
- *Big Five Inventory* [55] is a vastly used personality questionnaire including 44 items, 8-item scales for Neuroticism and Extraversion, 9-item scales for Agreeableness and Conscientiousness, and a 10-item measure of Openness. Participants respond to each item on a 5-point scale ranging from *strongly disagree* to *strongly agree*. This experiment only adopted the first section, as neuroticism and extraversion have been associated with different baseline arousal level and preferred arousal levels for optimal task performance [56, 57].
- *The Morningness-Eveningness Preferences (MEP) Scale* [58, 59] is a widely used questionnaire including 19 questions about personal time preferences in the day. Previous study suggested that people with different types of circadian rhythms may respond to emotional stimuli differently [57]. For example, evening type tends to respond to the unpleasant emotional stimuli more intensively compared to the morning type.

These questionnaires were assessed by psychological professionals for related research purposes.

3.3 Experimental Procedure

On arrival at the lab, participant is greeted by the experimenter, after which a brief description of the procedures of the study is given. He/She is shown the equipment for physiological data collection and told that he/she could terminate participation at any time without penalty. Then, he/she is asked to sign an informed consent form and indicate whether he/she agreed to be video recorded for the entire testing process.

After completing the consent form, participant will be given a copy of instruction booklet. The experimenter will explain the experimental procedures step by step and confirm that the participant has no doubts about the whole process. Subsequently, the participant will be asked to sit in front of the stimuli-presenting computer and the recording equipment will be put on him/her.

The computer-based experimental procedure consists of three sessions: the pre-test, the actual test and the post-test.

In the pre-test session, participant will practice performing the actual task in order to get familiar with the entire process. When he/she is confident to proceed, the actual test session will begin (Figure 3.3).

First, participant is shown a pre-test PANAS questionnaire on the computer screen. He/she will be asked to press the appropriate number key that best described his/her present mood state. Following that is a 3-minute recording session of the pre-test physiological levels. Participant is advised to remain in a comfortable and relaxed position during this period.

After the pre-test recording, the emotional induction phase begins. In total, there are 28 trials. Each trial consists of displaying a fixation point as “+” for 6 secs, a picture (randomly chosen from the 28 pre-selected IPAS pictures) for 6 secs,

and a black screen for 6 secs. Participant is asked to rate the viewed picture in terms of its arousal level and valence on a scale from 1-9 and verbally speak out a single word of emotion that best described his/her feelings after viewing the picture. Then, the participant chooses from a list of emotional descriptors extracted from Tomkins's concept of eight basic emotions: "anger", "interest", "contempt", "disgust", "distress", "fear", "joy", "shame", "surprise" [14,15], and "nothing" that best described his/her feelings after viewing the picture. Each trial is concluded with solving 5 simple mathematic problems so as to "wash out" the effect of the picture viewed on the subject before the next trial is administrated.

During the post-test session, another 3-minute recording of post-test physiological levels was conducted. Following that participant was required to fill in a post-test PANAS questionnaire again. The computer-based session took about 40 minutes.

Thereafter, the recording equipment would be removed from the participant. At the end, participant was requested to fill out a set of paper-based questionnaires including *information sheet*, *Cultural Identification Questionnaire*, *Big Five Inventory* and *The Morningness-Eveningness Preferences (MEP) Scale*. In total, the entire study took about 60 minutes.

3.4 Conclusion

In summary, we designed and conducted two rounds of experiments following a multi-sub approach using IAPS pictures as stimulus for emotional changes. Since the IAPS pictures were rated in continuous dimensions of valence and arousal, the subsequent data analysis was also classified based on these two dimensions. In addition, instead of using the original IAPS ratings, we required the participants to report their own ratings for consistency reasons, we also attempted to let them rate the pictures with discrete emotions. By doing so, not only do we hope to build

Figure 3.3: Experimental procedure of the actual test session

and improve the architecture of emotion recognition system based on valence and arousal, but may also seek a connection between the models of the continuous dimensions and the discrete emotions.

Though the data gathered from the two rounds of experiments were under almost the same settings, we choose to separately analyze the data sets rather than grouping them together. The main reason lies in the fact that whatever results we get from them can be mutually verified.

Collected raw signals have to be preprocessed and extracted as features, before classifying (for emotion recognition). Therefore the next chapter will elaborate the process of signal preprocessing and feature extraction, which is another essential step for preparing good data sets for analysis.

Chapter 4

Signal Preprocessing and Feature Extraction

4.1 Introduction

Collected raw signals need to be preprocessed before feature extraction. First, they were trimmed according to the targeted time period (e.g. 2 minutes in [50], 2 seconds in [48], and 12 seconds in our study from the beginning of displaying the picture stimulus until after showing the black screen.) Thereafter, they were passed through a series of filters: an 8th order Butterworth low pass filter with a cut off frequency (f_c) at $90H_z$, a very sharp Butterworth high pass filter with f_c at $0.1H_z$, and a Notch filter with f_c at $60H_z$ to remove power line noise.

Instead of directly using the ECG signal, the heart beat information is extracted for calculating the heart rate (HR), along with the five remaining signals (BVP, SC, EMGz, EMGc and Rsp) for subsequent feature extraction. This chapter will first elaborate the process of HR extraction, followed by the details of feature evaluation.

Figure 4.1: The QRS complex within the ECG signal

4.2 QRS detection of the ECG signals

Within an ECG signal, the QRS complex is the most striking waveform, as shown in Fig. 4.1. Because of its characteristic shape, it provides the fundamentals for heart rate calculation and other ECG related analysis [60]. Instead of directly using the ECG signals, most studies [2, 48, 50] chose to use the HR (heart rate) information, which was deduced from the intervals between successive QRS complexes. The derivative-based algorithm for extracting the QRS complexes [61] as described below:

$$y_0(n) = |x(n) - x(n - 2)| \quad (4.2.1)$$

$$y_1(n) = |x(n) - 2x(n - 2) + x(n - 4)| \quad (4.2.2)$$

$$y_2(n) = 1.3y_0(n) + 1.1y_1(n) \quad (4.2.3)$$

In Equation (4.2.1) - (4.2.3),

$x(n)$: The pre-processed version of the raw ECG signal.

$y_0(n)$: The approximate result of the smoothed three-point first derivative of the given signal $x(n)$.

$y_1(n)$: The approximate result of the second derivative of the given signal $x(n)$.

$y_2(n)$: Output of $y_0(n)$ and $y_1(n)$ weighted by 1.3 and 1.1 respectively.

The final result $y(n)$ is obtained by further smoothing $y_2(n)$ through a moving-average (MA) filter over M points, where $M = 16$ as given in Equation (4.2.4). Fig. 4.2 illustrates an example procedure of the QRS detection after each steps.

$$y(n) = \frac{1}{M} \sum_{i=0}^{M-1} y_2(n-i) \quad (4.2.4)$$

Next, a peak-searching algorithm was employed in our study to locate the heart beat.

1. Determine a threshold value y_{thr} . Scan the portions of $y(n)$ that satisfy $\{y(n)|y(n) > y_{thr}\}$ and label those samples as $g(n)$.
2. Predefine a value W and create a window with size equals to $2W + 1$. Shift this window along $g(n)$, select those samples whose value is the maximum within the window. In this study, $y_{thr} = 0.3$ and $W = 3$.

4.3 Feature Evaluation

In general, a feature is an individual scalar component of a feature vector that is used by the data learning algorithm. Feature vector is also known as a pattern

Figure 4.2: Illustration of the QRS detection

or an observation. Usually, the feature is represented as x_i and the feature vector is represented in boldface as $\mathbf{x} = (x_1, \dots, x_g)^T$, with g as the dimensionality of the feature vector. The input patterns can be presented by a capital boldface $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_N)$, where N is the number of input patterns.

The diversity of applied signals resulted in various types of features. So far there is no conclusive outcome on which feature extracted from which signal best annotates certain emotions. Thus, studies tend to investigate a wide range of features from different signals to explore the robustness towards targeted emotional states [2, 10, 50]. For example, Kim and André calculated a total of 110 features from four physiological signals: ECG, Rsp, SC and EMG. The investigated analysis domains vary from time series, frequency domain, geometric analysis, multiscale sample entropy and sub-band spectra, etc. to the amplitude variances and occurrences of transient changes for non-periodic signals such as EMG and SC [10].

Conventionally, the basic feature set for emotion recognition is statistically extracted from the time domain of all the physiological signals. Originally proposed by Picard et al. [2], these features have also appeared in Mera and Ichimura [62], Haag et al. [48], Wagner et al. [50], Kim and André [10] and Gu et al.. The main merit of using statistical time domain features lies in the fact that these features are computationally efficient, and therefore are more favorable for industrial applications. Suppose $x(n)$ represents the raw signal after preprocessing.

1. The mean of $x(n)$:

$$\mu = \frac{1}{n} \sum_{n=1}^N x(n) \quad (4.3.5)$$

2. The standard derivation of $x(n)$:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (x(n) - \mu)^2} \quad (4.3.6)$$

3. The mean of the absolute values of the first differences of $x(n)$:

$$\delta = \frac{1}{N-1} \sum_1^{N-1} |x(n+1) - x(n)| \quad (4.3.7)$$

4. The mean of the absolute values of the first differences of normalized $x(n)$:

$$\tilde{\delta} = \frac{1}{N-1} \sum_1^{N-1} |\tilde{x}(n+1) - \tilde{x}(n)| \quad (4.3.8)$$

5. The mean of the absolute values of the second differences of $x(n)$:

$$\gamma = \frac{1}{N-2} \sum_1^{N-2} |x(n+2) - x(n)| \quad (4.3.9)$$

6. The mean of the absolute values of the second differences of the normalized $x(n)$:

$$\tilde{\gamma} = \frac{1}{N-2} \sum_1^{N-2} |\tilde{x}(n+2) - \tilde{x}(n)| \quad (4.3.10)$$

where

$$\tilde{x}(n) = \frac{x(n) - \mu}{\sigma} \quad (4.3.11)$$

refers to the normalized signal of $x(n)$.

In our study, there were six physiological channels: HR, BVP, SC, EMGz, EMGc and Rsp, as shown in Figure 4.3. So eventually, each data entry has 36 statistical time-domain features extracted for further analysis.

4.4 Conclusion

In this chapter, we reviewed the process of raw signal preprocessing and feature extraction. Instead of directly using the ECG signals, we located the most striking

Figure 4.3: Process of feature extraction

waveform, the QRS complex, and used them as the heart beat information for HR calculation. Along with the five remaining signals, BVP, SC, EMGz, EMGc and Rsp, we extracted 36 statistical time-domain features, with 6 features from each signal channel.

From the two rounds of experiment, we managed to gather 784 and 868 data entries respectively (details are listed below in Table 4.1 and 4.2). Notice that not all of the participants' data were recorded into the data set for further analysis. One of the main reasons could be that after some frequent tests, the electrodes were corroded to some extent due to sweat or other contacts, which eventually caused inaccurate readings. In order to ensure the correctness of the subsequent analysis, we have to remove some corrupted data. All the following analysis in this thesis will be based on these two data sets to ensure the fairness of comparison.

Table 4.1: Data distribution of the 1st experiment

		Arousal		
		Rating 1-5	Rating 6-9	Total
Valence	Rating 1-5	187	294	481
	Rating 6-9	93	210	303
	Total	280	504	784

Table 4.2: Data distribution of the 2nd experiment

		Arousal		
		Rating 1-5	Rating 6-9	Total
Valence	Rating 1-5	283	261	544
	Rating 6-9	134	190	324
	Total	417	451	868

Chapter 5

Conventional methods for emotion classification

5.1 Introduction

As mentioned before, since there is no conclusive outcome on which feature extracted from which signal best annotates certain emotions, some studies tend to examine a wide range of features. As a result, the original feature space could contain irrelevant and redundant features that may affect the efficiency of the learning process, which finally degrade the learning performance and cause poor predictive accuracy. Therefore, most studies seek to reduce the amount of features (e.g. select a subset of features) before recognizing the emotion states, so as to increase the prediction accuracy of the applied classifier.

Roughly categorized, there are three kinds of feature selection/reduction procedures ever applied in previous studies. First of all, a pure dimensionality-reduction procedure, functions by evaluating the intrinsic characteristics of the original data set, e.g. Fisher projection (FP) [63], analysis of variance (ANOVA) [64] and principal component analysis (PCA) [65], which do not include any “search” actions on individual features. The second kind however, works mainly under a proper

strategy of searching, e.g. sequential forward search (SFS) [66], sequential backward search (SBS) [66], sequential floating forward search (SFFS) [67] and Genetic algorithm (GA) [68]. This type of selection usually introduces a pre-determined classification algorithm and uses its classification result of each generated feature subset as the evaluation criteria. Since the targeted classification algorithm is combined into the search process, the aim of improving the classification performance is more specific than the former type of feature reduction method. The third type is a hybrid procedure by embedding the reduction method into the search procedure, e.g. SFFS-FP [2].

The rest of the chapter is structured as below. First, by using data from two rounds of our experiments (Exp I & II), we compared the performances of applying FP, SFFS and SFFS-FP as the representatives of feature reduction and feature selection methods, with several classification algorithms such as support vector machine (SVM) [63] and k-nearest neighbors (kNN) [69] etc. Next, we especially investigated two types of feature selection techniques, the SFFS and GA, with a series of classifiers. Also, we provided a summary of robust features statistically calculated based on the results of Exp I.

5.2 Preliminaries

Before presenting the results of our experiments, this section briefly introduces some selected feature reduction/selection as well as classification tools that we used. Most of these methods had been applied in previous studies.

5.2.1 Feature reduction and selection algorithms

1. Fisher projection (FP)

Fisher projection is a well-known technique developed for reducing the dimensionality of the feature space by seeking an optimal orientation that makes all the projected samples are well separated [2, 63]. According to the nature of the projection, the original feature space is transformed into a new one with fewer dimensions. This process is briefly described below.

Let $\mathbf{x} = \{x_1, \dots, x_i, \dots, x_n\}$ denote a set of n d -dimensional sample vectors belonging to “ c ” classes.

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in c_i} \mathbf{x}, \quad (5.2.1)$$

$$\bar{\mathbf{m}} = \frac{1}{n} \sum \mathbf{x}, \quad (5.2.2)$$

where n_i is the number of vectors in the i th class, $\mathbf{m}_i$ stands for the mean of sample vectors within class c_i and $\bar{\mathbf{m}}$ is the mean of all vectors.

So the between-class scatter matrix is written as

$$S_B = \sum_{i=1}^c N_i (\mathbf{m}_i - \bar{\mathbf{m}})(\mathbf{m}_i - \bar{\mathbf{m}})^t. \quad (5.2.3)$$

where N_i represents the number of samples in class i . And the within-class scatter matrix reads as

$$S_W = \sum_{i=1}^c \sum_{x_j \in c_i} (x_k - \mathbf{m}_i)(x_k - \mathbf{m}_i)^t = \sum_{i=1}^c V_i, \quad (5.2.4)$$

where V_i is the covariance matrix of class c_i .

For a c -class problem, the natural generation of Fisher projection involves $c - 1$ discriminant functions. Thus, it actually projects data from a d -dimensional space onto a $(c - 1)$ -dimensional space. Let W be the $d - by - (c-1)$ projection matrix. The transformation of $\mathbf{x}$ can be written as

$$\mathbf{y} = W^t \mathbf{x}. \quad (5.2.5)$$

Thus the corresponding scatter matrices for projected samples $\mathbf{y}$ become

$$\tilde{S}_B = W^t \tilde{S}_B W \quad (5.2.6)$$

and

$$\tilde{S}_W = W^t \tilde{S}_W W. \quad (5.2.7)$$

The optimal transformation matrix W_F maximizes the ratio of the determinant of $\tilde{S}_B$ to the determinant of $\tilde{S}_W$, such that the separation between classes is maximized while the variance within a class is minimized (Fisher's criterion).

$$W_F = \arg \max_W \frac{|\tilde{S}_B|}{|\tilde{S}_W|} = \arg \max_W \frac{|W^t S_B W|}{|W^t S_W W|}, \quad (5.2.8)$$

where the columns of W_F are the generalized eigenvectors that correspond to the $c - 1$ largest eigenvalues in

$$S_B \mathbf{w}_i = \lambda_i S_W \mathbf{w}_i. \quad (5.2.9)$$

If S_W is non-singular, W_F can be obtained by applying the eigenvalue decomposition to the matrix $S_W^{-1} S_B$.

2. Sequential Floating Forward Search (SFFS)

Traditional feature set search algorithms like SFS and SBS suffer from an issue called “nesting effect”. It means that during the “top down” process of SBS, once discarded features cannot be reselected. While in the case of “bottom up” search of SFS, once selected features cannot be later discarded. This issue leads to the fact that by using SFS or SBS, the selecting results are only suboptimal [66]. SFFS is an improved search method of SFS by following a “bottom up” procedure but introducing a “floating” characteristic. The resulting dimensionality in each respective stages of the searching progress is not changing monotonically but is actually “floating” up and down. A summarized implementation procedures of SFFS algorithm proposed by Pudil et al. in [67] is described below.

Let $X_k = \{X_i : 1 \leq i \leq k, x_i \in Y\}$ be the set of k features from the set $Y = \{y_i : 1 \leq i \leq D\}$ of D available features. $J(X_k)$ is the feature selection criterion function.

1. Feature x_j from the set X_k is the *worst* feature in the set X_k if

$$J(X_k - x_j) = \max_{1 \leq i \leq k} J(X_k - x_i) \quad (5.2.10)$$

2. Feature y_j from the set $Y - X_k$ is the *best* feature with respect to the set X_k if

$$J(X_k + y_j) = \max_{1 \leq i \leq D-k} J(X_k + y_i) \quad (5.2.11)$$

Suppose at present, k features have already been selected from the complete set of Y to form set X_k with corresponding criterion function values $J(X_k)$. $J(X_i)$ for all preceding subsets of size $i = 1, 2, \dots, k-1$ are known and stored.

- *Step 1 (Inclusion):*

Using basic SFS method, select the best feature X_{k+1} with respect to the set X_k from $Y - X_k$ to form feature set X_{k+1} . Therefore, $X_{k+1} = X_k + x_{k+1}$.

- *Step 2 (Conditional exclusion):*

Find the worst feature x_r in the set X_{k+1} .

if ($x_r == x_{k+1}$)

Set $k = k + 1$ and return to Step 1.

else

Exclude x_r from X_{k+1} to form a new feature set X'_k .

if ($k == 2$)

Set $X_k = X'_k$, $J(X_k) = J(X'_k)$ and return to Step 1.

else

Go to Step 3.

- *Step 3 (Continuation of conditional exclusion):*

Find the worst feature x_s in the set X'_k .

if $(J(X'_k - x_s) \leq J(X_{k-1}))$

Set $X_k = X'_k$, $J(X_k) = J(X'_k)$ and return to Step 1.

else

Exclude x_s from X'_k to form a new feature set X'_{k-1} ,

Set $k = k - 1$.

if $(k == 2)$

Set $X_k = X'_k$, $J(X_k) = J(X'_k)$ and return to Step 1.

else

Repeat Step 3.

The program is initialized by setting $k = 0$ and $X_0 = 0$. Step 1 is repeated twice to obtain a feature set of cardinality 2 before continuing the algorithm from Step 1.

3. Hybrid SFFS with Fisher projection

The hybrid SFFS-Fisher method embeds Fisher projection into the progress of SFFS by feeding the Fisher algorithm with the subsets proposed by SFFS as the input. Brief description of the process is given below. In this review, we use k-Nearest Neighbor (kNN) rule ($k = 7$) as the classifier.

- Generate a subset of features following the rule of SFFS.
- Use the selected features for both training and testing the data sets. Calculate the Fisher projection matrix from only the training set, then project both training and testing set down to the d dimensions found by Fisher.
- Apply kNN rule on the projected training and testing sets to obtain the correct classification ratio (CCR).

- Take the CCR obtained as the criteria function for SFFS to adjust the next round of feature subset selection.

4. Genetic Algorithm (GA)

GA is a stochastic optimization procedure. It maintains a set of solutions called a *population*, to be analyzed and improved through successive iterations called *generations*. Each solution, a *chromosome*, is a string represented by a finite sequence of 0's and 1's. Suppose $a = \{\alpha_i\}$, $i = 1, \dots, d$, stands for a chromosome with d characters. In the application of feature selection, $\alpha_i = 1$ means the i^{th} feature has been selected, whereas $\alpha_i = 0$ denotes exclusion. The chromosomes are allowed to *mate*, *crossover* and *mutate* simulating the process of natural evolution. Only the “best” chromosomes of one generation will survive to the next cycle of reproduction.

When implementing GA to solve a particular problem, a fitness function must be specified to evaluate the quality of each solution. Siedlecki and Sklansky [70] introduced the ideas of penalty function and score with respect to the error rate of the kNN classifier to form a proper fitness function. As their goal is to seek the smallest or the least costly subset of features that does not deteriorate the classification result below a certain level, the size of selected subset was also considered in the function design. However in this study, the aim was simply to find a subset of features that best improves the performance of the combined classification method. Only the predicted error rate was required for GA to minimize. According to the mapping rule of objective functions to fitness form described in [68], the employed fitness function became like:

$$f(x) = c - e(x) \quad (5.2.12)$$

In Equation (5.2.12), $e(x)$ is the classification error rate of classification, c is a positive integer such that $f(x)$ is non-negative and monotonic decreasing.

Figure 5.1: Schematic illustration of crossover and mutation

The main process of the GA-based feature selection method is described below [68].

1. Set the GA operating conditions: *size_of_population* (m), *number_of_generations* (g), *percentage_of_selection* (p_s), *crossover rate* (p_c) and *mutation rate* (p_m).
2. Generate an initial population set, $pop = \{a_i\}, i = 1, \dots, m$.
3. For generation from 1 to g , repeat the following steps.
 - (a) Feed pop into the fitness function to calculate fits.
 - (b) For $n = p_s \cdot m$, select $npop = \{a_i\}, i = 1, \dots, n$ with the top most fits.
 - (c) Choose a pair a_j and a_l from pop and do $crossover(a_j, a_l)$ with probability p_c . Combine the results with $npop$.
 - (d) For $npop = \{a_i\}, i = 1, \dots, m$, do $mutation(a_i)$ with probability p_m .
 - (e) Set $pop = npop$.

Function *crossover* implements information exchange among two chromosomes (Figure 5.1). Suppose $a_j = \{\alpha_i\}, i = 1, \dots, d$ and $a_l = \{\beta_i\}, i = 1, \dots, d$.

1. Randomly choose an integer position $s \in [1, d]$ called the *crossover point*.

2. Swap $\{\alpha_1, \dots, \alpha_{s-1}\}$ with $\{\beta_1, \dots, \beta_{s-1}\}$ to form a new $a_j = \{\beta_1, \dots, \beta_{s-1}, \alpha_s, \dots, \alpha_d\}$ and a new $a_l = \{\alpha_s, \dots, \alpha_{s-1}, \beta_s, \dots, \beta_d\}$.

Function *mutation* is the occasional (with small probability) random alteration of a single bit in a chromosome (Figure 5.1). Suppose $a_j = \{\alpha_i\}, i = 1, \dots, d$.

1. Randomly choose an integer position $t \in [1, d]$.
2. If $\alpha_t = 1$, set $\alpha_t = 0$; Else, set $\alpha_t = 1$.

The *crossover rate* (p_c) is the probability of allowing a qualified pair of chromosomes for crossover, and the *mutation rate* (p_m) is the probability of switching bits in the chromosomes [70].

5.2.2 Selected classification algorithms

1. The k-Nearest-Neighbor (kNN) Rule:

This rule classifies x by assigning it with a label of the most frequently represented among the k nearest samples. In other words, a decision is made by examining the labels on the k nearest neighbors (the Euclidean distance) and taking a vote. The kNN rule is a suboptimal procedure. However, it has been shown that in the infinite sample situation, the error rate for the 1NN rule is bounded above by no more than twice the optimal Bayes error rate. Also as k increases, this error rate approaches the optimal rate asymptotically [63, 69].

2. The Distance-Weighted k-Nearest-Neighbor (dw-kNN) Rule:

Instead of using the Euclidean distance as a measure for locating the nearest neighbors, dw-kNN rule employs a weighting function that varies with the distance between the sample and the considered neighbor in such a manner that the value

decreases with increasing sample-to-neighbor distance. Three types of weighting functions applied in this study are depicted below [71].

dw-kNN(i):

$$weight_i = \frac{1}{i} \quad (5.2.13)$$

dw-kNN(ii):

$$weight_i = \frac{k - i + 1}{k} \quad (5.2.14)$$

dw-knn(iii):

$$weight_i = \frac{1}{d_i} \quad (5.2.15)$$

3. The Fuzzy k-Nearest-Neighbor (fkNN) Rule:

The fkNN algorithm introduces a concept named class membership to achieve “soft” assignment to a particular class, so there is no arbitrary assignments. Membership is a function of the sample’s distance from its k nearest neighbors and those neighbors’ memberships in the possible classes. After calculating the memberships of the test sample, it is assigned to the class with highest membership. Following the idea of Keller et al. [72], the membership function is depicted as below.

$$u_i(x) = \frac{\sum_{j=1}^k u_{ij} \left(\frac{1}{\|x-x_j\|^{\frac{2}{m-1}}} \right)}{\sum_{j=1}^k \left(\frac{1}{\|x-x_j\|^{\frac{2}{m-1}}} \right)} \quad (5.2.16)$$

In Equation (5.2.16), $u_i(x)$ is the assigned membership of the vector x (to be computed), u_{ij} is the membership in the i th vector of the labeled sample set, and the variable m is the scaling factor for fuzzy weights. In our implementation, $m = 2$ means the contribution of each neighboring point is weighted by the reciprocal of its distance from the point being classified.

4. Linear and Quadratic Discriminant Function (LDF and QDF):

LDF and QDF are two types of Discriminant Function Analysis (DFA) [63] with “linear” and “quadratic” kernels. DFA is a statistical method that builds a probability model for each class. For a new instance, the class whose model fits the best is chosen. In general, it is a very useful tool to seek variables that best separates the data.

5. Support Vector Machine (SVM):

SVM is a computational machine learning system which constructs a separating hyperplane in feature space in such a way that maximizes the margin between classes [73,74]. This study tested four types of kernel functions in SVM which are “linear”, “polynomial”, “radial bases functions (RBF)” and “Sigmoid”.

5.3 Performance comparison of selected feature reduction/selection methods

In this section, we used the data from two rounds of our experiments as a platform to test and compare several classification and feature reduction/selection methods used in the other studies. There were three major comparisons depicted in this section. Firstly, to review the performances of emotional states classification with/without using feature reduction/selection methods, also the effectiveness for valence and arousal recognition (2ES classification). Secondly, to investigate the differential results between subject-independent and subject-dependent approaches. Thirdly, to examine the consistency of results by comparing the outcomes using data of Exp II. We also conducted a 4ES classification as what Kim and André did in [10]. The examined classification methods included LDF, kNN ($k = 7$ in the tests) and SVM, as three representatives from other procedures ever used in the previous studies. These methods shared the property that they

required lower parameter changes during the classification process, hence they can be easily limited in a fair range. And the selected feature selection methods were FP, SFFS, and hybrid SFFS-FP.

Data entries of both Exp I and II were initiated with 36 feature dimensions. Considering the relatively large size of data samples and computational requirement, a 10-fold cross-validation was determined to evaluate the prediction performance of the classification methods. The whole data set was divided into 10 subsets (no overlap upon the validation subsets). For each round, one out of the ten subsets was acted as the test set while the other nine subsets were combined together as the training set. The final prediction accuracy was calculated by averaging the results from 10 rounds of tests.

Before going into the details of the results section, there are a few special terms that need to be introduced first.

1. **Subject-dependent approach:** Classification on individual subject (single subject approach).
2. **Subject-independent approach:** Classification on multiple subjects (mixed subject approach).
3. **2ES classification: 2 emotional states (ES) classification:** Differentiate valence (valence rating ≥ 6) vs. non-valence (valence rating ≤ 5), and arousal (arousal rating ≥ 6) vs. non-arousal (arousal rating ≤ 5).
4. **4ES classification: 4 emotional states (ES) classification:** Differentiate positive & high arousal (EQ1), negative & high arousal (EQ2), negative & low arousal (EQ3), and positive & low arousal (EQ4), as illustrated in Figure 5.2.

Figure 5.2: Categories of affective states on 2D emotion model: EQ1 = $val_rating > 5$ and $aro_rating > 5$, EQ2 = $val_rating \leq 5$ and $aro_rating > 5$, EQ3 = $val_rating \leq 5$ and $aro_rating \leq 5$, and EQ4 = $val_rating > 5$ and $aro_rating \leq 5$.

5.3.1 Results comparison of 2ES classification

1. Subject-independent approach

The recognition accuracies of valence and arousal with subject-independent approaches were separately displayed in Figure 5.3 and 5.4 for Exp I and II. The four bar groups respectively presented the results of LDF, kNN and SVM with non-feature reduction, FP, SFFS and SFFS-FP.

In Figure 5.3, apparently group 3 triumphs the others with the highest rate of 0.78 (valence) and 0.79 (arousal) using SFFS-kNN. Unexpectedly, instead of increasing the performances of the corresponding classification algorithms, the aid of FP brought negative impact on the results that were obtained when no feature selection method is used. However by combining FP with SFFS, the outcomes grew way up, not as good as group 3 though, but still achieving the highest rate of 0.77 (valence) and 0.76 (arousal) using SFFS-FP-kNN. Compared Figure 5.3 (a) with (b), it was not clear to conclude between valence and arousal, which one was more easy to be recognized. There is merely a difference between valence and arousal recognition when SFFS-kNN is applied. Using SFFS-LDF resulted in 0.77 for valence and 0.74 for arousal, while SFFS-SVM yielded 0.77 for valence and 0.78 for arousal.

As in Figure 5.4 where results from Exp II were shown, almost the same trend as that of Exp I is observed. When no feature selection method is used, SVM outperformed the other two in both valence and arousal. Adding FP results in worse outcomes, e.g. for SVM, drop rate of 5% (valence) and 20% (arousal). This time, SFFS-SVM achieved the best results with 0.76 (valence) and 0.81 (arousal), a bit higher than 0.76 (valence) and 0.79 (arousal) using SFFS-kNN. Compared to valence, arousal received roughly 20% more accuracy while SFFS was employed with each classifier.

(a) Valence

(b) Arousal

L: LDF, K: kNN, S: SVM;
F-L: FP-LDF, F-K: FP-kNN, F-S: FP-SVM;
S-L: SFFS-LDF, S-K: SFFS-kNN, S-S: SFFS-SVM;
SF-L: SFFS-FP-LDF, SF-K: SFFS-FP-kNN, SF-S: SFFS-FP-SVM

Figure 5.3: Results comparison of Exp I with subject-independent approach using combinational methods

(a) Valence

(b) Arousal

L: LDF, K: kNN, S: SVM;
F-L: FP-LDF, F-K: FP-kNN, F-S: FP-SVM;
S-L: SFFS-LDF, S-K: SFFS-kNN, S-S: SFFS-SVM;
SF-L: SFFS-FP-LDF, SF-K: SFFS-FP-kNN, SF-S: SFFS-FP-SVM

Figure 5.4: Results comparison of Exp II with subject-independent approach using combinational methods

(a) Exp I: Valence

(b) Exp I: Arousal

Figure 5.5: Recognition rates of valence and arousal against feature dimensionality based on SFFS: Exp I

(a) Exp II: Valence

(b) Exp II: Arousal

Figure 5.6: Recognition rates of valence and arousal against feature dimensionality based on SFFS: Exp II

The above results showed that, applying a proper feature selection method is essential for the classification process. By illustrating the recognition rates against feature dimensionality, it is possible to discover a rough dimension of subsets that could effectively serve the performances of the classifiers. In Figure 5.5 and 5.6, recognition rates using the three classifiers, LDF, kNN ($k = 5$) and SVM, were compared and contrasted against feature dimensionality. In the case of valence (Figure 5.5(a) and 5.6(a)), performance started to rise up to top around a dimensionality of 10 features and falls after 20 features. For arousal, prediction rates started to increase at around 5 and drop after 22 features. These phenomenon indicated that there do exist significant feature redundancy in the original 36-dimension features, which could explain why the bench marking results (without feature selection) are not so promising in the first place.

The redundancy could also be the main reason that causes fisher projection to be ineffective, since fisher projection works on the ground of the provided data structure and very much influenced by that. Whereas the “wrapper” kind of feature selection methods such as the SFFS aim in finding the optimal feature set that best suits the performance of the “embedded” classifier. The objective of “improvement” is more straightforward on the inputs. Thus, search algorithm such as the SFFS tries to search a subset of most representative features so that the influence of feature redundancy can be reduce.

Thus, applying SFFS or SFFS-FP as the feature selection procedure can help improve the performance dramatically by searching the close matching feature sets for the corresponding classifier used. With reference to the results from both valence and arousal based on data from Exp I and II, SFFS-KNN would be the most worthy of consideration.

2. Subject-dependent approach

Figure 5.7 and 5.8 presented the results of valence and arousal using the subject-dependent approach. Since there were 28 subjects in Exp I and 31 subjects in Exp II, the listed accuracies were the mean values obtained by averaging the individual results by 28 (Exp I) and 31 (Exp II). It showed some consistencies comparing the results of subject-dependent and the subject-independent approach. Firstly, SVM still triumphed the other two methods in both valence and arousal recognition when no feature selection method is applying. Results reached 80% (valence and arousal) in Exp I, 76% (valence) and 81% (arousal) in Exp II. Secondly, an additional step of Fisher projection hurt the performance of all the classifiers. E.g. for SVM, the results dropped roughly 18% (valence and arousal) in Exp I, 14% (valence and arousal) in Exp II. Thirdly, SFFS as usual, dramatically improved the accuracy to over 90% in both valence and arousal. The two interesting inconsistencies lie in the fact that firstly, SFFS-FP outperforms pure SFFS in the subject-dependent approach and secondly, instead of SFFS-kNN, SFFS-FP-LDF becomes the more effective selection for both valence and arousal.

Table 5.1 listed the results with subject-independent and subject-dependent approach using SFFS-kNN in Exp I and II. Apparently, subject-dependent approach vastly improved the recognition rates (average 18% for valence and 14.5% for arousal) in both experiments. This finding was consistent with Kim and André's study in [10], where roughly 14% increase for valence and 10% for arousal were seen. Note that, though in [10], the results of arousal were 12% better than valence, and in Haag et al.'s study [48], arousal is 7% better than valence recognition, our experiments obtained relatively close results for valence and arousal.

(a) Valence

(b) Arousal

L: LDF, K: kNN, S: SVM;
F-L: FP-LDF, F-K: FP-kNN, F-S: FP-SVM;
S-L: SFFS-LDF, S-K: SFFS-kNN, S-S: SFFS-SVM;
SF-L: SFFS-FP-LDF, SF-K: SFFS-FP-kNN, SF-S: SFFS-FP-SVM

Figure 5.7: Results comparison of Exp I with subject-dependent approach using combinational methods

(a) Valence

(b) Arousal

L: LDF, K: kNN, S: SVM;
F-L: FP-LDF, F-K: FP-kNN, F-S: FP-SVM;
S-L: SFFS-LDF, S-K: SFFS-kNN, S-S: SFFS-SVM;
SF-L: SFFS-FP-LDF, SF-K: SFFS-FP-kNN, SF-S: SFFS-FP-SVM

Figure 5.8: Results comparison of Exp II with subject-dependent approach using combinational methods

Table 5.1: Results of subject-independent vs. subject-dependent using SFFS-kNN

	subject-independent		subject-dependent	
	Valence	Arousal	Valence	Arousal
ExpI	0.78	0.79	0.95	0.94
ExpII	0.76	0.79	0.95	0.93

5.3.2 Results comparison of 4ES classification

Besides differentiating valence and arousal as two separate labels, we also conducted tests with 4ES pattern, which was formed by quartering the valence-arousal plane. Similar procedure was conducted by Kim and André in [10], where they developed a emotion-specific multilevel dichotomous (EMDC) scheme with pLDA to exploit the property of dichotomous categorization. The 4ES classification is treated as an advance step towards mapping the continuous emotional plane to discrete emotion states. Since kNN is capable of simultaneously classifying multiple classes (LDF and SVM only suit for two-class problem), it is employed with combinations of SFFS and SFFS-FP.

The results shown in Table 5.2 revealed that even with the help of SFFS, the recognition rates still are not promising with roughly 53% with subject-independent approach. The main reason lies in the fact that four-class classification model is more complex than the two-class one. Thus, some studies choose to transform a multiple-class problem into several two-class processes just to eliminate the impact. For example, in Kim and André’s study [10], they managed to obtain an accuracy of 70% following the EMDC scheme. However, when applying to subject-dependent case, the SFFS-FP-kNN was able to achieve average 95.5% accuracy for a simultaneous 4ES classification, compared to 95% using EMDC in [10].

Table 5.2: Comparison of results from Exp I and II for 4ES recognition

	subject-independent		subject-dependent	
	Exp I	Exp II	ExpI	ExpII
SFFS-KNN	0.53	0.55	0.87	0.84
SFFS-FP-KNN	0.53	0.5	0.95	0.96

5.3.3 Summary

Below is a short summary of this section. Firstly, for 2ES recognition of valence and arousal, using the 36 statistical time-domain features from ECG, BVP, SC, EMGz, EMGc and Rsp is more than enough to obtain rough 77% (valence) and 79% (arousal) accuracies following the subject-independent approach. Though it is acknowledged that the original 36-dimension feature space does exist some extent of redundancy. Thus, a feature set searching algorithm is strongly recommended to effectively find the optimal feature subset that is best suited for the corresponding classifier. Secondly, it is difficult for simultaneous 4ES recognition as the classification process is greatly influenced by the inner structure of each class. However, using subject-dependent approach can dramatically improve the performance either of 2ES or 4ES recognition algorithm. The intricate variety of individual differences can not be neglected in the study of emotion recognition. Therefore, techniques or algorithms on how to improve the existing emotion recognition system for general usage (multiple subjects) with the challenge of individual diversities should be one of the major challenges in the future. Last but not least, though previous studies reported that recognizing arousal tends to achieve higher accuracy than valence, the results from our experiment show relatively equal performance for both. It is an expected outcome, considering the different experimental settings in every

study, such as the number of the subjects, the stimuli for emotion elicitation, the target time period that is used for data analysis, etc.

5.4 Recognition results with/without SFFS & GA

This section mainly focuses on two feature selection methods, SFFS and GA, with data from Exp I. The basic purpose of combining SFFS and GA with several selected classifiers is to investigate the improvements of the classification accuracy by reducing the dimensionality of the feature space. The byproduct of the investigation is to explore the most robust and relevant features during the classification process. Thus the following contents are presented as follows. The first section shows a group of classification results without using feature selection methods for bench mark. Followed by the second section, where the results are compared separately using SFFS and GA with several classifiers. And the last section summarizes the effectiveness of selected features during the classification process.

All the tests (2ES classifications) are conducted with the same settings that are described in Section 5.3, following a subject-independent way on data from Exp I only. Examined classification techniques include kNN, dw-kNN, fkNN, LDF, QDF and SVM. The parameter k used in kNN, dw-kNN and fkNN is set from 1 to 20.

5.4.1 Results without feature selection

Figure 5.9 illustrated the results with respect to the value of k (details of the results can be referred to Appendix A.1 and A.2). As k increases, it revealed a slight uptrend (not monotonic though) on the classification performance through all the kNN related methods. The best prediction rates showed 64.29% for valence using dw-kNN(ii), $k = 19$ and 63.89% for arousal using dw-kNN(iii), $k = 19$.

(a) Valence

(b) Arousal

Figure 5.9: Comparison of valence and arousal recognition rates without dimensional reduction using kNN, dw-kNN and fkNN ($1 \leq k \leq 20$)

Table 5.3: Recognition rates of valence and arousal without dimensional reduction

		Valence	Arousal
kNN		0.6429	0.6339
dw-kNN(i)		0.6072	0.6148
dw-kNN(ii)		0.6429	0.6377
dw-kNN(iii)		0.6327	0.6389
fkNN		0.6352	0.6198
LDF		0.6224	0.5778
QDF		0.63	0.5408
SVM	Linear	0.6568	0.6428
	Polynomial	0.6735	0.6593
	RBF	0.6747	0.6568
	Sigmoid	0.685	0.6683

Table 5.3 summarized the maximum recognition rates for valence and arousal using kNN, dw-kNN and fkNN, as well as LDF, QDF and SVM. It clearly showed that using SVM, especially with “sigmoid” kernel presents the best performances in both valence (68.5%) and arousal (66.83%) recognition. LDF and QDF, particularly the latter, revealed significant drop in rate from the results of valence to arousal. Among the various kNN approaches, kNN and dw-kNN(ii) shared equal results in valence recognition (64.29%), whereas dw-kNN(iii) showed slightly better than these two classifiers for arousal (63.89%).

In general, there are no significant changes among kNN, three types of dw-kNN and fkNN. Simple variation of the distance measurements among feature points still fails to solve the issue of nicely matching the structure of 36-dimension feature space. Apparently without any dimensional reduction, SVM outperforms other tested classifiers. However, it is still unable to exceed a recognition rate beyond 70%.

Table 5.4: Recognition rates with dimensional reduction using SFFS

	SFFS-kNN	SFFS-fkNN	SFFS-LDF	SFFS-QDF
Valence	0.7795	0.7809	0.7718	0.7916
Arousal	0.788	0.7838	0.7384	0.8095

5.4.2 Results with feature selection

1. Hybrid with SFFS

Figure 5.10 showed the results of kNN and fkNN along with hybrid SFFS with kNN (SFFS-kNN) and fkNN (SFFS-fkNN). The detailed outcomes of using SFFS-kNN and SFFS-fkNN were given in Appendix A.4. The top performance for valence recognition appeared 77.95% with SFFS-kNN ($k = 7$) and 78.09% with SFFS-fkNN ($k = 6$). For arousal, they were 78.8% with SFFS-kNN ($k = 10$) and 78.38% with SFFS-fkNN ($k = 16$). There was significant improvement in the recognition rates of both valence and arousal. For valence, average increments were 14.65% (SFFS-kNN over kNN) and 16.18% (SFFS-fkNN over fkNN). For arousal, the improvement was 15.95% (SFFS-kNN over kNN) and 17.08% (SFFS-fkNN over fkNN) respectively. SFFS-fkNN showed slight advantage in valence recognition with increase in the value k , but remained continuously inferior compared to kNN in arousal recognition when k is chosen between 4 to 14. Generally, using SFFS-kNN and SFFS-fkNN have almost the same performance in classifying both affective states.

Along with LDF and QDF, Figure 5.11 compared the maximum recognition rates of valence and arousal before and after using the SFFS (detail numbers can be referred to Table 5.4). The average increase reached around 15% for valence and 18% for arousal. It shows great potential for using the SFFS in improving the classification performance by reducing the redundancy of the original feature space.

(a) Valence

(b) Arousal

Figure 5.10: Comparison of valence and arousal recognition rates using kNN, SFFS-kNN, fkNN, and SFFS-fkNN ($1 \leq k \leq 20$)

Figure 5.11: Comparison of results with dimensional reduction

Figure 5.12: Results of GA-kNN and GA-LDF with various parameter settings

We will further discuss this issue in Section 5.5.

2. Hybrid with GA

GA-based feature selection is implemented according to what was described in Section 5.2.1. Since the parameter settings will influence the performance of the algorithm, a range of values for *number_of_generations* (g), *size_of_population* (m), *percentage_of_selection* (p_s), *crossover rate* (p_c) and *mutation rate* (p_m) are experimented (initial values are obtained from [68, 70]). Each GA test consists of 5 runs of genetic search with the same parameter settings but for different initial populations. The outcome is the mean value of these 5 outputs to suppress the dependency of the results on the initial population. In this study, GA will be used together with kNN ($k = 5$) and LDF for evaluation.

Results shown in Table 5.5 revealed the strength of GA optimizing feature subset for improving the recognition rates for both valence and arousal. GA-kNN

Table 5.5: Recognition rates using GA-based dimensional reduction

	g	m	p_s	p_c	p_m	Valence	Arousal
GA-kNN ($k = 5$)	500	20	0.2	1	0.1	0.7269	0.7263
		20			0.15	0.7361	0.7299
		20			0.2	0.7325	0.7355
		50			0.1	0.7381	0.738
		50			0.15	0.7402	0.7487
		50			0.2	0.7417	0.7502
GA-LDF	500	20	0.2	1	0.1	0.7487	0.7141
		20			0.15	0.7543	0.7421
		20			0.2	0.7649	0.7152
		50			0.1	0.767	0.7421
		50			0.15	0.7665	0.7466
		50			0.2	0.7695	0.7446

achieved 9.88% (valence) and 11.63% (arousal) improvements over kNN. While greater improvements appeared for GA-LDF over LDF which are 14.71% (valence) and 16.88% (arousal). Figure 5.12 illustrated the influences of parameter settings (sequence refers to Table 5.5) for GA-kNN and GA-LDF in both valence and arousal cases. By increasing the values of *population* and *mutation*, performance of both GA-kNN and GA-LDF were gradually (though not monotonically) improved for valence. On the other hand, contrary to the monotonic growth of arousal recognition using GA-kNN, GA-LDF leads to slightly abnormal results with the first three settings and the best recognition rate appears at the second to last position. On average, among all the tested parameter settings, $g = 500$, $p = 50$, $p_s = 0.2$, $p_c = 1$ and $p_m = 0.2$ generated the most efficient one.

5.4.3 Summary

A short summary of using SFFS and GA is given below. Both algorithms show good potential for improving the recognition accuracies by reducing the irrelevance

and redundancy of the original feature space. Comparing the tested results of SFFS and GA, the former performs slightly better than the latter. Also, the SFFS does not request parameter settings as that are necessary for GA. Therefore it reduces the complexity of implementing the classification program.

5.5 Discussion on robust features

In Figure 5.13, recognition rates using four classifiers: kNN ($k = 5$), fkNN ($k = 5$), LDF and QDF, were compared and contrasted against the dimensionality of features. In the case of valence (Figure 5.13(a)), performance started to rise up to top around a dimensionality of 10 features and falls after 22 features. For arousal, prediction rates started to increase at around 5 and drops after 22 features. These phenomena indicated that there do exist significant feature redundancy in the original 36-dimension features, which could explain why the bench marking results (without feature selection) are not so promising in the first place.

Detailed results of selected features of the SFFS-kNN approaches were presented in Appendix A.3 and A.5. The features listed in the left-most column were grouped by signals for easier physiological interpretation: HR(H), BVP(B), SC(S), EMGc(E_c), EMGz(E_z) and Resp(R). The “Frq” column indicated the frequency of each feature being selected (summation along each row divided by test rounds(#rd)). In the SFFS-kNN case, since 10-cross validation was applied and the k value for kNN was selected from 1 to 20, in all there were 200 rounds of tests for each feature. Higher value of “Frq” suggested that the feature may be a robust one for the kNN classifier regardless of k . Appendix A.6 presented the selection results using SFFS-LDF (test round for each feature is 10). For comparison, Figure 5.14 and 5.15 respectively illustrated the “Frq” values in Appendix A.3, A.5 and A.6. Notice that for each round of validation, since the pair of train/test data sets

(a) Valence

(b) Arousal

Figure 5.13: Recognition rates of valence and arousal against feature dimensionality based on SFFS (5NN: kNN with $k = 5$, f5NN: fkNN with $k = 5$)

are different, the selected best subset of features may not be exactly the same or composed of a consistent dimension of features through all the 10 runs. Thus the analyzed feature subset actually means an accumulation of 10 optimal subsets.

There are a few interesting observations from Figure 5.14. Among all the 36 features, δ_B , $\tilde{\delta}_B$, γ_B and $\tilde{\gamma}_B$ from BVP, δ_S , $\tilde{\delta}_S$, γ_S and $\tilde{\gamma}_S$ from SC, as well as δ_R , $\tilde{\delta}_R$, γ_R and $\tilde{\gamma}_R$ from Resp reveal significantly lower chosen frequency in both valence and arousal recognition. In other words, they are rarely selected through multiple rounds of kNN tests. Other 16 features from BVP, SC and Resp however, present quite essential contribution to kNN classification. For example, μ_B is one of the most used features for valence and σ_S is among the top two features for arousal. Features from EMGz and EMGc share a generally consistent contribution between valence and arousal cases. The only exception is the performance of μ_{E_c} and σ_{E_c} , which show over 40% use in valence recognition. While in the case of HR, μ_H reaches a utilization rate of over 25% in valence case but drops to less than 10% for arousal.

The intention of putting Figure 5.15 here with Figure 5.14 was to compare the similarities and differences in feature selection using two classifiers. Surprisingly, $\tilde{\delta}_S$, which was seldom used by SFFS-kNN, revealed to be one of the most frequently selected features with SFFS-LDF for valence recognition. On the contrary, σ_S appeared not so active as in SFFS-kNN case.

In summary, regardless of the value of k , there do exist certain types of features showing distinctive strength when applying the kNN classifier. However, they are not the optimal subset of features for running the LDF. It is quite understandable since the quality of recognition accuracy is determined by the intrinsic relationship between the features and the applied classifier. Also, these robust features can only be treated as the most effective ones for our data set (Exp I).

Figure 5.14: Summary of each feature chosen in SFFS-kNN (legend $\delta(norm)$ and $\gamma(norm)$ stand for $\tilde{\delta}$ and $\tilde{\gamma}$ respectively)

Figure 5.15: Summary of each feature chosen in SFFS-LDF (legend $\delta(norm)$ and $\gamma(norm)$ stand for $\tilde{\delta}$ and $\tilde{\gamma}$ respectively)

Figure 5.16: Block diagram of the general emotion recognition system

5.6 Conclusion

In this chapter, we implement and compare various combinations of the classification algorithms with and without using feature selection/reduction methods. Also, we investigate the diversity of results between subject-dependent and subject-independent approaches.

As a conclusion from all the outcomes, we presented a block diagram as shown in Figure 5.16 that constructs the most effective procedure for the physiological signals based emotion recognition using a feature set searching algorithm e.g. the SFFS. As previously explained, the SFFS performs in a way that specifically finds the optimal feature subset that best suits the corresponding classification algorithm. Thus on one hand, it leads to good classification performance. However on the other hand, it could cost a relatively high computational expense, which mainly depends on the dimensionality of the original feature space. In our study, we apply 36 statistical time-domain features from 6 physiological signals to seek a balance.

Notice that results of the subject-independent approach tend to be poorer than the ones with subject-dependent approach. In fact, this issue has been recognized and described as “the intricate variety of nonemotional individual contexts among the subjects, rather than an individual ANS (autonomic nervous system) specificity in emotion” [10]. On one hand, the problem shows the concern of different inter-

pretation of emotions across individuals within the same culture [2]. On the other hand, the problem can be explained as the possibility of significant physiological signal variation across individuals, which would be the most probable cause of bad recognition performance.

The next chapter presents an improved classification structure, the biometric signature based (BSB) system, which aims to solve the problem of “individual varieties” when multiple subjects are involved. The basic idea is to transform a traditional subject-independent problem into several subject-dependent cases (separate the mixed data samples into subject groups based on their biometric characteristics), then proceed to the regular emotion recognition process. Further discussion of the BSB system will be presented in the next chapter.

Chapter 6

The Biometric Signature Based (BSB) System

6.1 Introduction

During the development of a general emotion recognition system, the “individual varieties” among multiple subjects is always an issue. Kim and André had briefly described the problem as “the intricate variety of nonemotional individual contexts among the subjects, rather than an individual ANS (autonomic nervous system) specificity in emotion” in [10].

On one hand, the problem shows the concern of different interpretation of emotions across individuals within the same culture [2]. Therefore it may complicate the signal processing and classification procedures when the goal is to examine whether subjects elicit similar physiological patterns for the same emotion. Thanks to the vast studies in psychology such as the proposal of six basic emotions by Ekman [4,5] or the development of the International Affective Picture System (IAPS). This culture-related impact of “individual varieties” can be alleviated to some extent by using the scientific categorization of emotions and the usage of standardized emotion elicitation facilities.

On the other hand however, the actual variation of physiological signals across individuals can be quite tricky to deal with. The basic intention of using physiological signals for emotion recognition is to discover the inner trend of signal variation during human's emotional variation. So in fact, it yields a fundamental knowledge that the physical properties of all signals should at least follow the same pattern. In other words, if signals from different subjects differ too much, the exact pattern of signal changes during emotional variation could be buried by the other distinct patterns brought by the "individual varieties". That is why some studies limited the experimental subject to a single person [2, 48], to potentially remove the variability. This compromise of single-subject approach might be considered in the early stages of affective recognition as it is valuable to develop subject-dependent methods. However the specific features and recognition results obtained from one single person may not be the same for other subjects [2].

Hence, in order to develop a general emotion recognition system, subject-independent (meaning all signals mixed together for different subjects) approach is inevitable [10, 49, 50]. Unfortunately, compared to subject-dependent (single subject) approach, the results of subject-independent approach tends to be less than satisfactory, especially when more subjects are involved. Upon such case, researchers start to think whether there is any method to reduce the heavy influence of the "individual varieties" so to increase the recognition accuracies.

In this chapter, a novel framework called the biometric signatures based (BSB) emotion recognition system is proposed aiming to resolve the problem of "individual varieties" when multiple subjects are involved. The basic idea is to transform a traditional subject-independent problem into several subject-dependent cases (separate the mixed data samples into subject groups based on their biometric characteristics) by conducting an additional subject classification process prior to the usual emotion recognition procedure. During this process, data from each sub-

ject are trained into separated statistical subject models. A new incoming sample will be classified into one subject model that best suits its inner structure. In other words, this data sample shows the most similar characteristics to the assigned statistical model. Then in the succeeding stage, a traditional feature selection plus classification procedure is applied to fulfill the task of emotion recognition. Compared with the conventional structure, the BSB system eliminates the influence of the “individual varieties” and therefore enhances the recognition performance of the whole system.

6.2 Conventional methods for emotion recognition

Usually when most studies proceed to the stage of classification, they already possess a bunch of features extracted from various physiological signals that are involved in the process of data collection. Since there is still no conclusive outcome on which feature extracted from which signal best annotates certain emotions, studies tend to investigate a wide range of features from different signals in order to explore the robustness towards targeted emotional states [2, 10, 50]. As a result, the high dimensionality of the original feature space can be quite troublesome in two ways. Firstly, the existence of feature redundancy is highly possible, which usually damages the performance of the applied classifier. And secondly the increment of computational complexity and expenses when more features are involved in the classification process. Hence, though some studies [48, 49, 75, 76] stick to the idea of directly inputting the whole feature set for classification, others [2, 10, 50] chose to introduce feature selection/reduction procedures into the procedure, with the purpose of reducing the amount of features therefore increasing the prediction accuracy.

Figure 6.1: Block diagram of the conventional emotion recognition system

Fig. 6.1 draw a block diagram of the classification system for emotion recognition, where SFFS is employed as the feature set searching procedure, k-nearest neighbors (kNN) rule is applied as the classifier.

The process of feature selection and classification shown in Fig. 6.1 is briefly descried below. Suppose S_t , $\{x_{\alpha}(\alpha=1,\dots,a)\}$ is the training data set, and $\{y_{\beta}(\beta=1,\dots,b)\}$ is the testing data set.

- Generate a subset of features following the rule of SFFS.
- Use the selected features for both training and testing data set.
- Apply kNN rule on the training and testing set, obtain the correct classification ratio (CCR).
- Take the CCR obtained as the criteria function for SFFS to adjust the next round of feature subset selection.

A normal feature set search algorithm performs in a “wrapper” kind of way to specifically find the optimal feature subset that best suits the corresponding classification algorithm. On one hand, it usually leads to good classification performance. However on the other hand, it could result in relatively high computational expense, which mainly depends on the dimensionality of the original feature space. This causes a dilemma whether to examine a broad range of features that may provide more opportunities improving the recognition rates, or just a few of them in

Figure 6.2: Architecture of the biometric signature based emotion recognition system

consideration of the heavy burden of computational expenses. In our study, 36 statistical time-domain features from 6 physiological signals are applied to adjust the balance. Details are elaborated in Chapter 4. Fig. 6.1 depicts a brief diagram of the conventional emotion recognition system.

6.3 Methodology of the BSB system

As illustrated in Figs. 6.2 and 6.3, the complete BSB emotion recognition system consists of two separate stages. Firstly, the system classifies the whole training data set into several Gaussian mixture models (GMM), each of which represents one subject model. A new incoming sample is classified to one subject model using the Maximum Posterior Probability (MAP) rule to capture the intrinsic characteristics of the input data. By doing so, the original subject-independent problem is able to transformed into several subject-dependent cases. So the second stage follows a conventional emotion recognition procedure.

Stage 1: subject classification

(a) Training phase

(b) Testing phase

Figure 6.3: Stage 1 of the BSB system

6.3.1 Stage 1: subject classification process

Instead of applying the entire 36 features from all six channels of the physiological signals, only the 6 features from the HR signals are chosen to implement the first stage of subject classification. First of all, the more the features involved, the more the complex the system would be. Our intention is to see whether by using such simple features we can still achieve a good performance in the proposed system. In our case, the initial objective is to separate the mixed testing data into groups that best suit the inner structure of the data sample (using a model-based soft classification). Therefore it is rather a classification than an actual identification process.

The detailed implementation procedures and formulas for building and testing the GMM based subject models is described below. Fig. 6.3 shows the illustration of both training and testing phases.

1. Training phase

Suppose the training data collected from $S_1, \dots, S_K$ subjects, the objective is to build $M_1, \dots, M_K$ models each representing one subject. The following steps explain the process of generating M_g for subject S_g by taking the maximum likelihood (ML) solutions via Expectation Maximization (EM) algorithm [77, 78].

By definition, a GMM is expressed as follows:

$$P(X) = \sum_{i=1}^N \sum_{c=1}^C \pi_c p_c(x_i) = \sum_{i=1}^N \sum_{c=1}^C \pi_c N(x_i | \mu_c, \Sigma_c), \quad (6.3.1)$$

and

$$\begin{aligned}
 p_c(x_i) &= N(x_i|\mu_c, \Sigma_c) \\
 &= \frac{1}{(2\pi)^{g/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x_i - \mu_c)^T \Sigma^{-1}(x_i - \mu_c)\right)
 \end{aligned} \tag{6.3.2}$$

where C is the number of Gaussian components, $N(x|\mu_c, \Sigma_c)$ is a normal distribution with mean μ_c and covariance matrix Σ_c , and π_c is the weight of component with the constraint of $\sum \pi_c = 1$. x_i is a g -dimensional feature vector. $i = 1, \dots, N$ is the number of observed feature vector for each subject. The parameters of GMM are estimated using the Expectation-maximization (EM) algorithm [79], which yields a Maximum Likelihood (ML) estimation. Each iteration of the EM algorithm consists of the E-step (Expectation) and the M-step (Maximization). During the E-step, the missing data are estimated given the observed data and current estimate of the model parameters. And in the M-step, the likelihood function is maximized using the estimated missing data from the E-step instead of the actual missing data. Eq. (6.3.3) depicts the estimation formulas of the model parameters.

$$\begin{aligned}
 \pi_{c|i} &= \frac{\pi_c N(x_i|\mu_c, \Sigma_c)}{\sum_{j=1}^C \pi_j N(x_i|\mu_j, \Sigma_j)} \\
 \hat{\pi}_c &= \frac{1}{n} \sum_{i=1}^N \pi_{c|i} \\
 \hat{\mu}_c &= \frac{\sum_{i=1}^N \pi_{c|i} x_i}{\sum_{i=1}^N \pi_{c|i}} \\
 \hat{\Sigma}_c &= \frac{\sum_{i=1}^N \pi_{c|i} (x_i - \mu_c)(x_i - \mu_c)^T}{\sum_{i=1}^N \pi_{c|i}}
 \end{aligned} \tag{6.3.3}$$

The posterior probabilities of the input data sample, $P = \{p_k\}_{k=1,2,\dots,K}$, are calculated based on each GMM generated for corresponding subject. According to the

MAP rule, the data sample is assigned to the subject class with the highest posterior probability, $S_k = \operatorname{argmax}(P)$. The original subject-independent problem is transformed to a normal within-subject case for classification of the affective states.

- **Initialization:** Initialize cluster centers $\{\mu_c\}$ using k-means algorithm, cluster covariance matrices $\{\Sigma_c\}$, and mixing proportions $\{\pi_c\}$. Set the maximum iteration number $iter_max$ (e.g. 1000) and the threshold for parameter change ε (e.g. 10^{-3}).
- **E-step:** For the $(r+1)th$ iteration of the sequence of EM algorithm, calculate the following expectation:

$$E^{(r)}(\pi_{c|i}|x_i) = \frac{\pi_c^{(r)} N(x_i|\mu_j^{(r)}, \Sigma_j^{(r)})}{\sum_{c=1}^C \pi_c^{(r)} N(x_i|\mu_j^{(r)}, \Sigma_j^{(r)})}, \quad (6.3.4)$$

- **M-step:** For each cluster's parameters, $\pi_c^{(r+1)}$, $\mu_c^{(r+1)}$ and $\Sigma_c^{(r+1)}$ are updated according to:

$$\pi_c^{(r+1)} = \frac{1}{n} \sum_{i=1}^N E^{(r)}(\pi_{c|i}|x_i) \quad (6.3.5)$$

$$\mu_c^{(r+1)} = \frac{\sum_{i=1}^N E^{(r)}(\pi_{c|i}|x_i)x_i}{\sum_{i=1}^N E^{(r)}(\pi_{c|i}|x_i)} \quad (6.3.6)$$

$$\Sigma_c^{(r+1)} = \frac{\sum_{i=1}^N E^{(r)}(\pi_{c|i}|x_i)(x_i - \mu_c^{(r+1)})(x_i - \mu_c^{(r+1)})^T}{\sum_{i=1}^N E^{(r)}(\pi_{c|i}|x_i)} \quad (6.3.7)$$

- **Convergence test:** Iterate between step 2 and 3 until the percentage difference of complete-data log likelihood is less than ε .

2. Testing phase

The generated Gaussian models for each training subject $\{M_g\}$ ($g = 1, \dots, c$) are stored in the system for evaluating the new coming data samples. Each data sample x_l is put through the following steps.

- Calculate the posterior probabilities $\{P_{lg}\}$ based on each subject model M_g using 6.3.1.
- Locate the subject model M_s that generates $\max_g(P_{lg})$
- According to the MAP rule, assign x_l to the corresponding data pool of subject s .

6.3.2 Stage 2: emotion recognition process

After assigning each testing data to one subject data pool, the original subject-dependent task has been transformed into several subject-independent cases. In the 2nd stage, it becomes a conventional within-subject emotion recognition procedure. The process is actually the same as that described in Section 6.2. The training data set now is the assigned subject data pool, and the testing data is the new incoming data sample that is ready for recognition. In our experiment, SFFS is used to select the subset of features that provides the optimal classification performance. The main reason to choose kNN as the classifier lies in the fact that it is able to achieve simultaneous multi-class classification.

6.4 Experimental results

To better analyze the performance of the BSB system, three tasks are designed and implemented. Firstly, use the BSB system with data from Experiment 1 and

Figure 6.4: Categories of affective states on 2D emotion model: EQ1 = $val_rating > 5$ and $aro_rating > 5$, EQ2 = $val_rating \leq 5$ and $aro_rating > 5$, EQ3 = $val_rating \leq 5$ and $aro_rating \leq 5$, and EQ4 = $val_rating > 5$ and $aro_rating \leq 5$.

2 respectively, so as to investigate the effectiveness of the system on two separate data sets. Secondly, examine the BSB system using a combined data set from Experiment 1 and 2 together, so as to examine the robustness of the system. Results of both task one and two are compared with the ones using conventional method of SFFS-kNN ($k = 7$). Lastly for task three, in order to test the general usage of the system, the data from both experiments are used to mutually test each other's subject models. By doing so we prove that it is not required for the new incoming data to be known to the system, which is a practical concern mentioned in [10].

For each emotion recognition task, three goals are prepared: recognition for (1) valence ($val_rating > 5$) vs. non-valence ($val_rating \leq 5$), (2) arousal ($aro_rating > 5$) vs. non-arousal ($aro_rating \leq 5$), and (3) 4 quarters of the valence-arousal plane (4EQs) as shown in Fig. 6.4. Each task uses 10-cross validation, meaning the whole data set is divided into 10 subsets (no overlap upon the validation subsets), for each round, one out of the ten subsets acts as the test set while the other nine subsets are combined together as the training set. The details of the data sets used in experiment 1 and 2 are shown in Table 6.1 and 6.2. All the data entries are original with 36 feature dimensions, the same as in Chapter 5.

6.4.1 Effectiveness and robustness

Fig. 6.5 and 6.6 illustrated the results of valence, arousal and 4EQs recognition of both experiment 1 and 2. These results were displayed using 3D bar graph. The X-axis (the long axis) represents the number of subject models. There are 28 and 31 subjects in experiment 1 and 2 respectively, therefore the label of X-axis is from 1 to 28 in Fig. 6.5, from 1 to 31 in Fig. 6.6. The Y-axis (the short axis) represents the number of cross validation. All the 10-cross validations are also labeled using different color. The Z-value indicates the emotion recognition

Table 6.1: Data distribution of the 1st experiment

		Arousal		
		Rating 1-5	Rating 6-9	Total
Valence	Rating 1-5	187	294	481
	Rating 6-9	93	210	303
	Total	280	504	784

Table 6.2: Data distribution of the 2nd experiment

		Arousal		
		Rating 1-5	Rating 6-9	Total
Valence	Rating 1-5	283	261	544
	Rating 6-9	134	190	324
	Total	417	451	868

(a) Experiment 1: valence

(b) Experiment 1: arousal

(c) Experiment 1: 4EQs

Figure 6.5: Results of experiment 1

(a) Experiment 2: valence

(b) Experiment 2: arousal

(c) Experiment 2: 4EQs

Figure 6.6: Results of experiment 2

accuracy when there are test subjects falling into that particular model under the current validation subsets. For example, for the 3rd validation subset, there are 3 test subjects grouped into model 27 and the Z-value is 0.9, which indicates that the recognition accuracy of using the 3 test subjects and all the training subjects in the 2nd stage is 90%.

To better analyze the recognition results, average results across all models were created and shown in Table 6.3 and 6.4 (the column of “Avg.”). From these two tables, the recognition accuracy obtained via our BSB system was much higher than that of conventional SFFS-kNN method, under all subcategories.

To further test the robustness of our system, another experiment was performed. This time, all the subjects from Experiment 1 and Experiment 2 were put together and the whole process is repeated. The result was shown in Table 6.5. The mixture process introduces varieties into the experiment, also expands the whole data pool to 1652 data entries (784 data entries of experiment 1, 868 data entries of experiment 2). However, our BSB system can still successfully achieve consistent and high recognition accuracy as compared to the SFFS-kNN system.

6.4.2 General usage of the system

This subsection investigates the issue that whether the proposed BSB system can be used on data samples from the “unknown” subjects. In other words, whether the system requires cumulative collection of the data from each subject in the learning phase (to let the subject “known” by the system), in order to recognize any emotional data from these subjects. According to our findings, it is actually not required for the system to learn the subjects beforehand, as long as there are certain subject models that are representative enough for the “unknowns”. In order to prove this, the following experiments are designed.

Table 6.3: Results comparison of BSB and SFFS-kNN: experiment 1

	BSB											SFFS-kNN
	cross1	cross2	cross3	cross4	cross5	cross6	cross7	cross8	cross9	cross10	Avg.	
valence	0.95	0.98	0.93	0.79	0.95	0.95	0.99	0.98	0.99	0.9	0.94	0.78
arousal	0.89	0.93	0.9	0.94	0.96	0.93	0.92	0.97	0.83	0.88	0.92	0.79
4EQs	0.77	0.93	0.69	0.73	0.83	0.75	0.75	0.92	0.71	0.79	0.79	0.53

Table 6.4: Results comparison of BSB and SFFS-kNN: experiment 2

	BSB										SFFS-kNN	
	cross1	cross2	cross3	cross4	cross5	cross6	cross7	cross8	cross9	cross10		Avg.
valence	0.94	0.99	0.96	0.71	0.88	0.94	0.98	0.97	1	0.93	0.93	0.76
arousal	0.85	0.9	0.83	0.89	0.89	0.83	0.86	0.9	0.83	0.84	0.86	0.79
4EQs	0.75	0.81	0.7	0.68	0.72	0.8	0.77	0.82	0.81	0.76	0.76	0.55

Table 6.5: Results comparison of BSB and SFFS-kNN: experiment 1 and 2

	BSB										SFFS-kNN	
	cross1	cross2	cross3	cross4	cross5	cross6	cross7	cross8	cross9	cross10		Avg.
valence	0.95	1	0.94	0.86	0.92	0.96	0.96	1	0.99	0.94	0.95	0.73
arousal	0.87	0.95	0.83	0.9	0.9	0.88	0.89	0.84	0.85	0.8	0.87	0.73
4EQs	0.81	0.87	0.74	0.73	0.74	0.76	0.79	0.8	0.79	0.75	0.78	0.5

In the first example, only the data from experiment 1 is used to train the subject models, so there are 28 subject models generated and stored in the system. Then data from experiment 2 is fed into the system to examine the results. For the second example, the process is conducted the other way around by using data of experiment 1 to test 31 subject models built from the data of experiment 2. Results from these two examples are presented in Fig. 6.7 and 6.8. The average results of 10 crosses are listed in Table 6.6, with comparison of results of experiment 1 and experiment 2 from Table 6.3 and 6.4.

As can be seen from Table 6.6, compared to the previous experiments, the classification accuracy from these two mutual test examples are even better. For example, using the subject models of experiment 1 to test data from experiment 2 increases 0.04(valence), 0.1(arousal) and 0.12(4EQs) correctness, comparing with the ones using subject models of experiment 2. It experimentally proves that the system does not require the subjects to be explicitly learned by the system, provided there are some essential subject models exist in the input data sets.

It is interesting to note that the subject models created for the two mutual test examples are more representative than the ones from previous experiments. When doing the subject classification, what is expected is the testing data to find the best matching subject model by means of their inner data structure. Although there exists variations among subjects, the emotional characteristics are universal, which provide the basic foundation of the procedure of emotion recognition. Therefore, as long as the system provides some essential subject models (or patterns) for the testing data to be categorized into, it does not really matter whether the testing data actually belongs to the subject data pool or not.

In conclusion, the BSB system is proven to be effective and robust. And it can be generally used for multiple subjects without the necessity of user training (as long as there are enough representative models stored in the system).

Figure 6.7: Results of the mutual tests: exp 2 test models of exp 1

(a) valence

(b) arousal

(c) 4EQs

Figure 6.8: Results of the mutual tests: exp 1 test models of exp 2

Table 6.6: Comparison of Results

	exp 2 tests exp 1 models	experiment 2	exp 1 tests exp 2 models	experiment 1
valence	0.97	0.93	0.97	0.94
arousal	0.96	0.86	0.86	0.92
4EQs	0.88	0.76	0.81	0.79

6.5 Robust BSB system based on multivariate *t*-Distribution

6.5.1 Introduction

In previous sections of this chapter, a biometric signature based system has been introduced for emotion recognition. This system utilizes a two-stage framework, a step of statistical subject classification, following the usual emotion recognition procedure. In our original design, Gaussian mixture model is used for the subject classification process. This model is also widely used in many other disciplines like data mining, machine learning and others. As demonstrated in the last section, the GMM model is quite helpful in the BSB system. Nevertheless, it still has space to improve in regards to robustness. The maximum likelihood solutions for the GMM model is equivalent to finding the least-square solutions [80], which makes it sensitive to the outliers problem.

One of the solutions is to use an alternative wider-tailed distribution model to replace the Gaussian model. Multivariate *t*-Distribution is chosen in this particular study, as the multivariate *t*-Distribution offers many robust enhancements as compared to the multivariate Gaussian distribution. First of all, the Multivariate *t*-Distribution has wider tail so that it can address some input samples that lie far

way from the center. Secondly, it contains a wider class of elliptically symmetric distributions including Gaussian distribution [77], therefore it can take care of the variabilities inside the real data sets using in the emotion recognition system. Thirdly, there is an additional robust parameter ν (the Degree of Freedom) in the t-Distribution which can reduce the impact of the outliers in the calculation of its parameters and provide better robustness.

In this section, another implementation on the BSB system is introduced which utilizes the multivariate t-Distribution model in the first stage. Compared to the general Multivariate Gaussian model, this model has advantages in providing better recognition rates and simplifying the model complexities of the system.

6.5.2 Stage 1: robust subject classification processing using multivariate t-Distribution

Replacing the GMM model with the mixture model of multivariate t-Distribution (mmTD), the overall process of the subject classification is still the same as the one introduced in Section 6.3. The stage 1 - subject classification process is rewritten below.

Suppose that the collected training data comes from $S_1, \dots, S_c$ subjects, and our objective is still to build $M_1, \dots, M_c$ models each representing one subject. By replacing the gaussian distribution with t-Distribution, a multivariate t-Distribution mixture model can then expressed as follows:

$$T(X) = \sum_{j=1}^n f(y; \Psi) = \sum_{j=1}^n \sum_{i=1}^g w_i f(y; m, V, \nu) \quad (6.5.8)$$

where $f(y; m, V, \nu)$ is the probability density function of multivariate t-Distribution [78] ($i = 1, \dots, g$ and $j = 1, \dots, n$, where g represents the number of mixing distributions and n is the number of data samples).

The g -dimensional multivariate t-Distribution density function is given by

$$f(y; m, V, \nu) = \frac{\Gamma(\frac{\nu+d}{2})|V|^{-1/2}}{(\pi\nu)^{\frac{1}{2}d}\Gamma(\frac{\nu}{2})(1 + \delta(y, m; V)/\nu)^{\frac{1}{2}(\nu+d)}}, \quad (6.5.9)$$

where $\delta(y, m; V)$ denotes the Mahalanobis squared distance between y and the center m , with V as the covariance matrix.

The difference between the Gaussian distribution and the t-Distribution is the introduction of a new random variable $U(u_1, \dots, u_n)$ with Gamma distribution to scale the normal distribution as mentioned in [77], i.e.

$$U \sim \text{Gamma}(\frac{1}{2}\nu, \frac{1}{2}\nu) \quad (6.5.10)$$

where ν is the degree of freedom.

The hidden robust estimator u_j is influenced by the Mahalanobis squared distance $\delta(y, m; V)$ and the degree of freedom ν . By adjusting the degree of freedom ν , one can appropriately assign different weights to the outliers via the mean and covariance of the data. Smaller degrees of freedom will emphasize to reject outliers, while relatively larger values of the degree of freedom tends to fit a GMMⁱ to the data [81]. Hence, one can adjust the models' robustness to a particular data set or application by tuning the parameter ν [78].

Similar to the GMM case, the detailed implementation steps of the robust subject classification processing can be described as:

- **Initialization:** Initialize cluster centers $\{m_i\}$ using k-means algorithm, cluster covariance matrices $\{V_i\}$, mixing proportions $\{w_i\}$ and degrees-of-freedom parameter $\{\nu_i\}$. Set the maximum iteration number $iter_max$ (e.g. 1000) and the threshold for parameter change ε (e.g. 10^{-3}).

ⁱFor the case $\nu \rightarrow \infty$, the multivariate t-Distribution will reduce to a multivariate Gaussian distribution with m and V .

- **E-step:** For the $(k+1)th$ iteration of the sequence of EM algorithm, calculate the three expectations $E_{\Psi^{(k)}}(z_{cj}|Y_j)$ ($E^{(k)}$ in 6.5.11), $E_{\Psi^{(k)}}(U_j|y_j, z_j)$ ($E_u^{(k)}$ in 6.5.12) and $E_{\Psi^{(k)}}(\log U_j|y_j, z_j)$ ($E_{\log u}^{(k)}$ in 6.5.13).

$$E^{(k)} = \frac{w_i^{(k)} f(y_j; m_i^{(k)}, V_i^{(k)}, \nu_i^{(k)})}{\sum_{i=1}^g w_i^{(k)} f(y_j; m_i^{(k)}, V_i^{(k)}, \nu_i^{(k)})}, \quad (6.5.11)$$

$$E_u^{(k)} = \frac{\nu_i^{(k)} + d}{\nu_i^{(k)} + \delta(y_j, m_i^{(k)}; V_i^{(k)})}, \quad (6.5.12)$$

$$E_{\log u}^{(k)} = \log(E_u) + \psi\left(\frac{\nu_i^{(k)} + d}{2}\right) - \log\left(\frac{\nu_i^{(k)} + d}{2}\right). \quad (6.5.13)$$

where

$$\psi(s) = \frac{\Gamma'(s)}{\Gamma(s)} \quad (6.5.14)$$

is the Digamma function [77].

- **M-step:** For each cluster's parameters, $w_i^{(k+1)}$, $m_i^{(k+1)}$ and $V_i^{(k+1)}$ are updated according to:

$$w_i^{(k+1)} = \sum_{j=1}^n E^{(k)} / n \quad (6.5.15)$$

$$m_i^{(k+1)} = \frac{\sum_{j=1}^n E^{(k)} E_u^{(k)} y_j}{\sum_{j=1}^n E^{(k)} E_u^{(k)}} \quad (6.5.16)$$

$$V_i^{(k+1)} = \frac{\sum_{j=1}^n E^{(k)} E_u^{(k)} (y_j - m_i^{(k+1)})(y_j - m_i^{(k+1)})^T}{\sum_{j=1}^n E^{(k)}}. \quad (6.5.17)$$

- **Convergence test:** Iterate between step 2 and 3 until the percentage difference of complete-data log likelihood is less than ε .

6.5.3 Testing phase and emotion recognition process

The testing phase and emotion recognition process are exactly the same as what has been described in the previous sections. Therefore, they will not be repeat in

this Robust BSB system case. For more information, please refer to the respective subsections in Section 6.3.

6.5.4 Experimental results

In this section, the results are from three methods including the conventional architecture of SFFS-kNN are presented, the BSB system implemented by GMM, and the BSB system implemented by mixture model of multivariate *t*-Distribution, for a multi-subject emotion recognition task. The experimental data are gathered from two rounds of experiments, same as in Section 6.4. In order to evaluate the consistency of the results, the data was applied separately from each experiment. The classification goal is set to simultaneously differentiate 4EQs.

Figure 6.9 showed the comparison of results using the general emotion recognition system implemented by SFFS hybrid with kNN ($k = 5$)(GenSys), the BSB system with GMM (BSB-GM), and the BSB system with multivariate *t*-Distribution (BSB-TD). Figure 6.9(a) and (b) respectively depicted the results from Experiment I and II. The left graph of each figure drawn the curves of recognition accuracies based on each and every fold of cross validation (10-fold therefore 10 units for the X-axis). And the right graph showed the final accuracies corresponding to each recognition method by averaging the results of the 10 cross validations. As expected, both the BSB-GM and BSB-TD perform better than the SFFS-kNN structure by an average 30% accuracies for Experiment I, slightly less but still consistent for Experiment II. The results strongly support the basic idea of the BSB system by separating the mixed multi-subject data samples into individual groups based on their biometric characteristics (6 features from the ECG signal specifically). Also, using the multivariate *t*-Distribution model performs better than the GMM models. The reason is due to the fact that Student's *t*-Distribution [82] has been proven to be more robust to real data sets since it has longer tail than

traditional Gaussian distribution [77, 83].

6.6 Conclusion

This chapter presented a novel biometric signature based system (BSB), which introduces an additional subject classification process prior to the conventional emotion recognition procedure. The basic idea of this additional step is to transform a subject-independent emotion recognition problem into several subject-dependent tasks by separating the mixed data samples into subject groups based on their biometric characteristics. In the 1st stage, data from each subject are trained into separated statistical subject models. A new incoming sample will be classified to one subject model that best suits its inner structure. In other words, this data sample shows the similar characteristics to the assigned statistical model. Then in the 2nd stage, a general feature selection plus classification procedure is applied to achieve the task of emotion recognition. Compared with the conventional subject-independent procedure, the BSB system eliminated the influence from the “individual varieties” therefore enhanced the recognition performances of the whole system.

The experimental results showed that, the BSB system performs more effectively than using a conventional method of SFFS-kNN: roughly 20% increase of correctness for valence, arousal and 4EQs recognition based on the two data sets from our experiment. Also it shows good robustness for an extended database (59 subjects of 1652 data entries). In addition, through a mutual test experiment, the system shows a general usage in that it is not required for the users to be trained by the system.

In order to add robustness into our BSB system, another study is also performed on the BSB system that utilizes the multivariate t-Distribution model. Compared to

(a) Experiment I

(b) Experiment II

Figure 6.9: Comparison of results using three kinds of implementations for the 4EQs recognition

Figure 6.10: Results of subject classification using data of exp 1 and 2 for 4EQ recognition

the GMM version, the new robust BSB system has advantages in providing better recognition rates and is able to simplify the model complexities of the original one.

These two BSB systems, i.e. BSB-GM, BSB-TD, require each and every subject to train one subject model. In another word, this structure could increase the complexity of the system when more subjects are involved. Also, another finding is that some of the subject models can be “useless” or “marginal” ones during the subject classification processes. Fig. 6.10 presents the subject classification results of using data from experiment 1 and 2 for 4EQs recognition. It clearly shows that there are certain subject models into which not a single testing data has been categorized (subject model 2, 10, 13, 23, 28, 36, 42, 54, 55, 58). These “useless” models can be removed from the “essential” groups, or maybe combined into other groups. Also some other models e.g. subject model 8, 9, ... 40, etc., have only one or two samples identified into out of 10 rounds of cross validations. These “margin” models may also be removed or combined.

The next chapter will introduce another improved structure using a self-adaptive model generating method. It will be different from the original BSB system in that, instead of directly creating several models based on individual subject, the self adaptive procedure first assumes the whole data pool as one single statistical model, and successively splits the model into two new ones based on their statistical structures, until the model number reaches a pre-defined value. By doing so, the

models in the first stage are built in a flexible way rather than limited to one subject. This is discussed in the next chapter.

Chapter 7

The Self-adaptive BSB System

7.1 Introduction

When developing a general system for the multi-subject emotion recognition, one of the major issues is to solve the problem introduced by the intricate variety of nonemotional individual contexts among the subjects. For that, the Biometric Signature Based (BSB) system is proposed. The basic idea is to transform a multi-subject problem into several single subject cases (separate the mixed data samples into subject groups based on their biometric characteristics), then proceed to the usual emotion recognition process described before. The initial implementation of the BSB system requires to train data from each subject into one statistical subject model. So the new incoming sample would be classified into one of the subject models that best suits its inner structure. Chapter 6 implemented the BSB system using Gaussian Mixture Model (GMM) as well as mixture model of multivariate t-Distribution (mmTD). Both implementations considerably improved the accuracy of emotion recognition compared to the conventional SFFS-kNN method.

One limitation of the BSB system lies in the fact that it requires each and every subject to train one subject model. On one hand, this structure could increase the

complexity of the system when more subjects are involved. On the other hand, as discussed in Section 6.6 of the previous chapter, some of the subject models can be “useless” or “marginal” ones during the subject classification process. In other words, it may not be necessary for the system to train all the subject models.

In this chapter, the first stage of the original BSB system is improved by a self-adaptive model generating method. The new BSB system is named as the Self-adaptive BSB (SaBSB) system. The SaBSB system is different from the original BSB system in that, instead of directly creating several models based on individual subjects, the self adaptive procedure first assumes the whole data pool as one single statistical model, and successively splits into two new ones based on the statistical structures, until the number of models reaches a pre-defined value. By doing so, the models in the first stage can be built in a more flexible way rather than limit to each individual subject.

The following section will further demonstrate the detailed implementation of the SaBSB system using the mmTD model.

7.2 The self-adaptive BSB (SaBSB) system

7.2.1 Bayesian information criterion (BIC)

Before going into details of the SaBSB system, let us first introduce the term, BIC [84].

In statistics, BIC is a criterion for model selection among a class of parametric models with different numbers of parameters. Generally, it is used for determining the number of components in a model and for deciding which among two or more partitions most closely matches the data for a given model. Mathematically BIC

[84] can be expressed as:

$$2\log p(x|M) + \text{const.} \approx 2l_M(x, \hat{\theta}) - m_M \log(n) \equiv BIC, \quad (7.2.1)$$

where $p(x|M)$ is the (integrated) likelihood of the data for the model M , $l_M(x, \hat{\theta})$ is the maximized mixture log-likelihood for the model, m_M is the number of independent parameters to be estimated, and n is the sample size. According to Equation (7.2.1), the larger the value of BIC, the stronger the evidence for the model is.

7.2.2 Self-adaptive procedure

By using BIC and mixture model of multivariate t-Distribution (mmTD), the self-adaptive procedure can be described as follows:

Assume $\phi_s = \{\mu_s, \sigma_s, v_s\}$ are the parameters of the sth component.

- **Initialization:**

1. Build an mTD1 model (one single multivariate t-Distribution in the mixture model) based on the whole data. Calculate the parameters ϕ_1 : $\mu_1 = \text{samplemean}$; $\sigma_1 = \text{samplecovariancematrix}$; $v_1 = \text{const.}$ (set to 25 in our experiment). Let $\omega_1 = p(\phi_1) = 1$.
2. Perform EM learning on the data for modeling as mmTD2 using ϕ_1 as the initial parameters. Return: $\tilde{\phi}_1, \tilde{\phi}_2, p(\tilde{\phi}_1) = \tilde{\omega}_1, p(\tilde{\phi}_2) = \tilde{\omega}_2$.

- **Loop:** Take mmTDk for example, meaning there are k clusters now ($\text{clst}N = k$).

1. Decide which cluster (clst) to split (splitC):

- Calculate Δ_{BIC} for each cluster:

$$\Delta_{BIC_i} = BIC(mmTD2, clst_i) - BIC(mmTD1, clst_i),$$
 where $i \in [1, k]$.
 - $splitC = \text{argmax}_i \Delta_{BIC_i}$.
2. Perform EM learning on $splitC$ for modeling as mmTD2 using ϕ_{splitC} .
 3. Update parameters for $splitC$ as $\tilde{\phi}_1, \tilde{\phi}_2$,

$$\tilde{\omega}_1 = p(\tilde{\phi}_1) * \omega_{splitC}, \tilde{\omega}_2 = p(\tilde{\phi}_2) * \omega_{splitC}.$$
 4. If $clstN < maxLoop$, $clstN = clstN + 1$;
Else end loop.

In order to have a better understanding of the self-adaptive procedure, let us consider the following example.

A group of randomly generated data is shown in Figure 7.1. The simulated data set consists of four clusters of data, shown in Figure 7.1(a). Assume the whole data set can be modeled as one single mmTD (mmTD1) using the mean and covariance matrix from the original data set (Figure 7.1(b)). Split mmTD1 into mmTD2 ($clstN = 2$) using EM learning (Figure 7.1(c)). Decide the $splitC$ (the cluster to split) by calculating the Δ_{BIC} , and perform EM learning on it (now $clstN = 3$ as shown in Figure 7.1(d)). Repeat this process until reaching a satisfactory result (eventually $clstN = 4$, Figure 7.1(e)).

7.3 Experimental Results

In this section, the results of using four methods are presented, which consists of including 1) the traditional architecture, 2) the BSB system implemented by Gaussian Mixture Models (GMM), 3) the BSB system implemented by mmTD,

Figure 7.1: Demonstration of the self-adaptive process

Table 7.1: SaBSB: Number of clusters for each round of validation using a fix $nu = 25$.

Validation	1	2	3	4	5	6	7	8	9	10
Experiment I	7	6	6	16	10	16	5	11	6	12
Experiment II	7	5	10	10	8	10	10	13	11	12

(a) Experiment I

(b) Experiment II

Figure 7.2: Comparison of results using four kinds of implementations for the 4ES recognition

(a) Experiment I

(b) Experiment II

Figure 7.3: Comparison of results with different number of clusters generated by the SaBSB

and 4) the SaBSB implemented by mmTD, for a multi-subject emotion recognition task. The experimental data are gathered from two rounds of our experiments, same as in Section 6.4. In order to evaluate the consistency of the results, the data is applied separately from each experiment. The classification goal is set to simultaneously differentiate 4EQs.

Figure 7.3 shows the comparison of results using the general emotion recognition system implemented by SFFS hybrid with kNN ($k = 5$)(GenSys), the BSB system with GMM (BSB-GM), the BSB system with mmTD (BSB-TD), and also the SaBSB implemented by mmTD (SaBSB). Figure 7.3(a) and (b) respectively depict the results from Experiment I and II. The left graph of each figure draws the curves of recognition accuracies based on each and every fold of cross validation (10-fold therefore 10 units for the X-axis). And the right graph shows the final accuracies corresponding to each recognition method by averaging the results of the 10 cross validations. As expected, both the BSB and the SaBSB systems perform better than the SFFS-kNN structure by an average of 30% accuracies for Experiment I, slightly less but still consistent for Experiment II. The results strongly support the basic idea of the BSB system by separating the mixed multi-subject data samples into individual groups based on their biometric characteristics (6 features from the ECG signal specifically). Also, using the mmTD model performs better than the GMM models. The reason is due to the fact that Student's t-distribution [82] has been proven to be more robust to real data sets since it has longer tail than traditional Gaussian distributions [77, 83].

Compare the results of BSB-TD and the SaBSB, both use the mmTD model with the parameter $\nu = 25$ for all 10 cross validations. Although the SaBSB yields slightly lower performance compared to the BSB-TD, Table 7.1 shows that each cross needs the number of subject clusters no more than 16 (Experiment I) and 13 (Experiment II). As to what has been explained, the BSB system requires

exclusively that each subject to learn his own statistic model for the stage of subject learning, therefore 28 (subjects) for Experiment I and 31 (subjects) for Experiment II. However, the SaBSB is able to achieve a relatively comparable results with less constraint in the number of subject models to be built. In other words, the SaBSB method can potentially reduce the complexity of the whole system without losing much of the recognition performance.

Figure 7.3 illustrates the performance of the SaBSB system in a different way by fixing the number of subject clusters generated. Notice that the parameter ν is different for every cross so to ensure the performance of each round of validation. The trend of accuracies shows a positive proportion to the number of clusters.

When in the extreme case, i.e. each data entry could form its own model, the learning process would be purely exclusive to every individual and increases the recognition accuracy. However by doing so, it will also increase the computational complexity. Hence, a better solution is to set up a reasonable threshold that will not only assure the recognition accuracy, but also reduce the computational expenses.

7.4 Conclusion

This chapter introduced the SaBSB system as an improved version of the BSB system. Instead of directly creating several models based on individual subject, the self adaptive procedure first assumes the whole data pool as one single statistical model, and then successively splits into two new ones, until the number of models reaches a pre-defined value. In this way, the total number of statistical models in the first stage can be reduced, also the models can be generated in a more adaptive and flexible manner. By comparing SaBSB with conventional SFFS-kNN, BSB-GM and BSB-TD, the results show that SaBSB is able to achieve a relatively

comparable results while reduce the number of subject models to be built.

Chapter 8

Conclusions and Recommendations for Future Research

8.1 Conclusions

For years, research groups have put their efforts to develop a generally applicable emotion recognition system based on the use of physiological signals. Basically, the complete system can be divided into three major parts: (1) data collection from the designed experiment, (2) signal preprocessing and feature extraction, and (3) feature selection/transformation and classification. Each step has its own “parameters” that vary according to different studies. The “parameters” can be various

Figure 8.1: Block diagram of the general emotion recognition system

Table 8.1: Primary parameters in an emotion recognition study

Experimental Settings	Feature Extraction	Feature Selection/Redduction	Classification
(1) Single vs. Multiple subjects - Size of the subject pool (2) Basic emotions vs. Continuous dimensions (3) Material for emotion elicitation - Pictures/Photos, Videos, Music, etc. (4) Physiological signals: ECG, SC, Rsp, etc.	(1) Various domains - Time, frequency, etc. (2) Number of features	(1) Filter model: ANOVA, FP, etc. (2) Wrapper model: SFFS, SBS, etc. (3) Hybrid model: SFFS-FP, etc.	kNN, LDF, SVM, MLP etc.

Figure 8.2: Block diagram of the improved emotion recognition system

experimental settings for data collection, the selection of targeted physiological signals, the different features being extracted, or the various usage of classification techniques. Table 8.1 listed some of the primary “parameters” that differ among studies.

This study uses the datasets from our own experiments to create a single platform for several feature selection and classification methods. From the results, a traditional yet general design of the system is concluded, as shown in Fig. 6.1. The system adopts certain feature set search algorithm for the step of feature selection. The merit of using such algorithm is that it performs in a “wrapper” kind of way to specifically find the optimal feature subset that best suits the corresponding classification algorithm.

The results also show that subject-independent approach tend to be worse than subject-dependent approach. In fact, many researchers have been aware of this problem brought by “the intricate variety of nonemotional individual contexts among the subjects, rather than an individual ANS (autonomic nervous system) specificity in emotion” [10]. On one hand, the problem shows the concern of different interpretation of emotions across individuals within the same culture [2]. Therefore it may complicate the signal processing and classification procedures

when the goal is to examine whether subjects elicit similar physiological patterns for the same emotion. Thanks to the vast studies in psychology such as the proposal of six basic emotions by Ekman [4,5] or the development of the International Affective Picture System (IAPS), this aspect of “individual differences” has been some extent alleviated by the employment of scientific categorization of emotions and the usage of standardized emotion elicitation facilities. On the other hand however, the possibility of significant physiological signal variation across individuals, can be quite difficult to solve. The basic intention of using physiological signals for emotion recognition is to discover the inner trend of signal variation during human’s emotional variation. So in fact, it yields a fundamental knowledge that the physical properties of all signals should at least follow the same pattern. In other words, if signals from different subjects differ too much, the exact pattern of signal changes during emotional variation that one is looking for could be buried by the other distinct patterns brought by the nonemotional varieties. To avoid such problem, some studies limit the experimental subject to one single person [2, 48]. Although person-dependent methods which are worthy to consider in the early stages of affective recognition [2], the specific features and recognition results obtained from one single person may not be the same as that of other subjects. Hence, single-subject approach is always argued to exhibit a weakness that any developed recognition methods is not generally applicable. Whereas the subject-independent approach tends to perform worse than subject-dependent approach due to non-emotion-related varieties. To solve the problem, it is possible to identify each individual prior to the recognition phase. In another word, the system will require a learning phase to cumulatively collect the corresponding data of each subject - let the subjects to be “known” to the system [10].

Hence in Chapter 6, an improved structure for multi-subject emotion recognition using physiological responses is introduced, the biometric signature based (BSB) system (Fig. 8.2). The basic idea is to transform a traditional subject-independent

problem into several subject-dependent cases (separate the mixed data samples into subject groups based on their biometric characteristics), then proceed with the usual emotion recognition process. In the 1st stage, data from each subject are trained into separated statistical subject models. A new incoming sample will be classified to one subject model that best suits its inner structure. In other words, this data sample shows similar characteristics to the assigned statistical model. Then in the 2nd stage, a general feature selection plus classification procedure is applied to achieve the task of emotion recognition. Compared with the traditional subject-independent procedure, the BSB system eliminated the influence from the “individual varieties” therefore enhanced the recognition performance of the whole system. Also by conducting a group of close-set and open-set experiments, the results drew a preliminary conclusion that it was not requirable for the BSB system to learn all the subjects beforehand, as long as there were enough representative models known by the system.

By further extending the idea of BSB system into a more robust way, a new robust version of BSB system is proposed by replacing the GMM model with the mixture model of multivariate t-Distribution. Compared to the original GMM version, the new BSB system has advantages in providing better recognition rates and is able to simplify the model complexities of the original one.

Chapter 7 discusses one possible limitation of our original BSB system, i.e. the needs for every subject to train one subject model. A few experiments are performed and the results show that it is not requirable for the system to prepare statistical models for each individual subject, as long as there are enough representative models known (trained) by the system. With this in mind, the first stage of the BSB system has been improved by using a self-adaptive model generating method and propose a modified BSB system, named the Self-adaptive BSB (SaBSB) system. Bayesian information criterion is used to control the splitting

process. This new architecture is able to reduce the complexity of the system (less number of subject models to be built) which maintaining relatively comparable results as compared to the original BSB system.

8.2 Recommendations for Future Research

The research results and methodologies for physiological signals based emotion recognition offer many opportunities for further investigations. In here, some future research directions are recommended which can further enhance the performance of our proposed system:

Other factors: gender, age, time of the day, etc.

This research proposed the BSB system for multi-subject emotion recognition. This system eliminated the influence from the “individual varieties” therefore enhanced the recognition performances of the whole system by introducing an additional “subject classification” prior to the emotion recognition process. Instead of grouping the data pool into different subject models, future research can be carried out on the analysis of other factors, such as subject gender, age, morning/afternoon and so on, in order to eliminate the “individual varieties” in different angles.

“Representative” subject models

The models used for this “subject classification” are not necessarily created based on each and every individual subject as long as there are enough representative subject models stored beforehand. Therefore, it is rather important to define the number of “enough” models, and what is the proper criterion for “representative”. In order to create the representative models, SaBSB system is developed by grouping the data into individual models automati-

cally. The results are promising in the sense that they are rather compatible with the BSB system but with fewer models generated. The system can still be improved further, if a better way can be found to determine the optimal cluster number, either from splitting or merging.

From continuous dimensions to discrete emotions

For a complete emotion recognition system, from data collection, data processing, to feature extraction, feature selection and classification, each process can extend to many research opportunities. Our study focuses on the last classification step. Hence, the prior procedures were fixed and the same data set was used for all the classification methods. Instead of following the continuous dimensions of valence-arousal plane, the same procedure can be replicated on discrete emotions with properly gathered emotional data, or using a mapping model from continuous dimensions to discrete emotions.

Appendix A

Table A.1: Results of valence recognition without dimensional reduction using kNN, dw-kNN and fkNN ($1 \leq k \leq 20$)

k	kNN	dw-kNN(i)	dw-kNN(ii)	dw-kNN(iii)	fkNN
1	0.5548	0.5548	0.5548	0.5548	0.5548
2	0.5548	0.5548	0.5548	0.5548	0.5548
3	0.5919	0.5548	0.5868	0.5408	0.5792
4	0.5868	0.5868	0.588	0.5792	0.5829
5	0.6085	0.5778	0.5995	0.6071	0.5931
6	0.5957	0.5868	0.5919	0.6071	0.5842
7	0.6072	0.5944	0.6034	0.6059	0.6071
8	0.6022	0.5983	0.6022	0.6034	0.5945
9	0.634	0.5931	0.6188	0.6021	0.6009
10	0.6289	0.5957	0.6315	0.6123	0.6162
11	0.6365	0.5995	0.6327	0.6175	0.6148
12	0.6237	0.6008	0.6276	0.6187	0.6123
13	0.634	0.6033	0.6365	0.6289	0.6135
14	0.6429	0.602	0.6327	0.6315	0.6263
15	0.6404	0.6058	0.6378	0.6276	0.6199
16	0.6187	0.6071	0.6148	0.6327	0.6161
17	0.6314	0.6046	0.6288	0.6327	0.6237
18	0.6251	0.6071	0.6314	0.6302	0.6301
19	0.6302	0.602	0.6429	0.6302	0.6352
20	0.6404	0.6072	0.6365	0.6302	0.6339

Table A.2: Results of arousal recognition without dimensional reduction using kNN, dw-kNN and fkNN ($1 \leq k \leq 20$)

k	kNN	dw-kNN(i)	dw-kNN(ii)	dw-kNN(iii)	fkNN
1	0.5792	0.5792	0.5792	0.5792	0.5792
2	0.5792	0.5792	0.5792	0.5792	0.5792
3	0.5816	0.5792	0.5804	0.602	0.5689
4	0.6008	0.6008	0.5906	0.6007	0.5881
5	0.6159	0.597	0.6147	0.5969	0.5994
6	0.6096	0.5982	0.602	0.602	0.5944
7	0.616	0.5906	0.6045	0.6198	0.5994
8	0.6122	0.597	0.6135	0.6249	0.6071
9	0.6148	0.5919	0.6084	0.6211	0.5918
10	0.6033	0.5893	0.6033	0.6211	0.5969
11	0.6249	0.5918	0.6172	0.6224	0.5931
12	0.6224	0.5982	0.6237	0.6186	0.5943
13	0.6275	0.5969	0.6211	0.6198	0.6058
14	0.6288	0.5994	0.6313	0.6224	0.6058
15	0.6339	0.6007	0.63	0.6236	0.616
16	0.6288	0.602	0.6338	0.6313	0.616
17	0.6314	0.6084	0.6377	0.6326	0.6198
18	0.625	0.6058	0.6249	0.6351	0.6109
19	0.6288	0.6084	0.6288	0.6389	0.6185
20	0.6301	0.6148	0.6275	0.6377	0.6185

Table A.3: Summary of each feature chosen in SFFS-kNN for valence recognition (H : HR, B : BVP, S : SC, E_c : EMGc, E_z : EMGz and R : Resp, $Frq = \frac{\sum}{\#rd}$, $\#rd = 200$)

	$k = 1, 2, \dots, 20.$																				Frq (%)
μ_H	1	1	3	0	2	1	0	0	2	4	1	3	2	4	4	5	6	5	5	2	25.5
σ_H	3	3	2	2	0	1	0	1	1	2	0	1	2	3	1	2	4	1	1	0	15
δ_H	2	2	4	2	2	6	4	4	5	7	7	2	5	1	5	2	1	2	4	3	35
$\tilde{\delta}_H$	2	2	5	3	6	4	9	4	3	2	3	4	3	1	1	2	1	2	6	4	33.5
γ_H	2	2	2	4	4	2	4	4	3	1	1	0	2	2	3	4	4	2	1	2	24.5
$\tilde{\gamma}_H$	4	4	4	3	1	4	4	4	2	2	6	4	3	3	4	1	2	2	6	2	32.5
μ_B	6	6	6	6	3	6	6	4	3	2	4	6	4	5	4	3	3	3	4	6	45
σ_B	1	1	2	2	2	2	2	1	2	4	3	3	4	4	5	3	2	3	5	2	26.5
δ_B	0	0	1	0	0	0	0	1	2	0	0	1	0	0	2	0	0	1	0	1	4.5
$\tilde{\delta}_B$	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
γ_B	1	1	1	1	1	0	0	0	1	1	0	0	0	1	1	1	1	2	0	0	6.5
$\tilde{\gamma}_B$	0	0	0	1	2	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	3
μ_S	3	3	2	1	2	2	3	2	0	5	2	1	4	5	3	4	3	1	2	1	24.5
σ_S	4	4	2	3	4	4	4	4	4	1	1	5	5	3	2	1	3	4	4	5	33.5
δ_S	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5
$\tilde{\delta}_S$	0	0	0	1	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	2.5
γ_S	0	0	0	1	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	2.5
$\tilde{\gamma}_S$	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	2	0	1	3
μ_{E_c}	2	2	7	3	5	5	5	5	2	7	3	2	6	4	2	5	5	3	6	2	40.5
σ_{E_c}	4	4	3	4	4	4	4	4	4	3	5	4	4	5	8	5	6	6	4	7	46
δ_{E_c}	5	5	1	3	2	1	3	2	4	2	3	4	2	1	3	1	1	2	2	6	26.5
$\tilde{\delta}_{E_c}$	2	2	2	4	2	3	6	3	0	1	4	5	1	3	2	2	1	2	1	2	24
γ_{E_c}	3	3	3	3	2	1	2	4	3	0	4	4	2	2	1	1	1	1	4	6	25
$\tilde{\gamma}_{E_c}$	3	3	1	4	2	2	4	5	2	0	4	3	2	3	2	1	1	3	3	5	26.5
μ_{E_z}	2	2	2	1	2	3	4	2	2	4	4	2	2	1	3	1	2	2	5	2	24
σ_{E_z}	3	3	4	4	3	5	1	4	2	3	2	2	3	2	4	1	1	3	1	1	26
δ_{E_z}	3	3	3	2	2	0	2	2	2	0	2	2	2	3	2	3	4	3	1	4	22.5
$\tilde{\delta}_{E_z}$	0	0	2	4	3	2	0	2	0	0	3	2	1	2	0	2	1	1	2	4	15.5
γ_{E_z}	2	2	2	5	3	0	2	5	2	2	3	4	4	3	2	5	3	4	2	2	28.5
$\tilde{\gamma}_{E_z}$	1	1	3	3	3	2	1	1	0	1	4	1	2	3	1	3	3	1	4	4	21
μ_R	2	2	2	4	1	3	3	3	0	2	1	1	0	0	1	1	0	2	0	1	14.5
σ_R	3	3	3	2	3	3	3	2	3	1	3	4	4	2	1	4	4	4	1	0	26.5
δ_R	0	0	0	0	1	0	0	1	2	1	0	0	0	0	0	0	0	0	0	0	2.5
$\tilde{\delta}_R$	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	1.5
γ_R	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	1.5
$\tilde{\gamma}_R$	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1

Table A.4: Results of valence and arousal recognition using SFFS-kNN and SFFS-fkNN

k	Valence		Arousal	
	SFFS-kNN	SFFS-fkNN	SFFS-kNN	SFFS-fkNN
1	0.755	0.755	0.7692	0.7692
2	0.757	0.757	0.764	0.764
3	0.7668	0.7697	0.7711	0.7753
4	0.7767	0.7767	0.7851	0.7739
5	0.7682	0.7612	0.7837	0.7641
6	0.7696	0.7809	0.7837	0.7711
7	0.7795	0.7697	0.7852	0.7683
8	0.757	0.7668	0.7767	0.7612
9	0.7499	0.7795	0.7809	0.7683
10	0.7584	0.7781	0.788	0.7753
11	0.7626	0.7683	0.7753	0.7711
12	0.7626	0.7654	0.7781	0.7725
13	0.7682	0.764	0.7823	0.7781
14	0.7612	0.7668	0.7753	0.7655
15	0.7598	0.7655	0.757	0.7669
16	0.7542	0.764	0.7711	0.7838
17	0.7528	0.757	0.7612	0.7754
18	0.75	0.7626	0.7683	0.7753
19	0.75	0.7584	0.7683	0.7711
20	0.7584	0.7626	0.7598	0.7697

Table A.5: Summary of each feature chosen in SFFS-kNN for arousal recognition (H : HR, B : BVP, S : SC, E_c : EMGc, E_z : EMGz and R : Resp, $Frq = \frac{\sum}{\#rd}$, $\#rd = 200$)

	$k = 1, 2, \dots, 20.$																				Frq (%)
μ_H	0	0	2	1	1	1	1	1	1	1	1	1	2	0	1	1	0	2	0	2	9.5
σ_H	2	2	3	0	2	1	2	1	2	2	3	2	2	1	2	2	3	1	0	2	17.5
δ_H	4	4	3	3	3	4	1	2	4	3	3	2	3	3	2	3	2	3	2	5	29.5
$\tilde{\delta}_H$	2	3	5	4	3	3	3	6	2	4	4	3	1	3	5	2	2	2	2	4	31.5
γ_H	2	2	5	2	2	3	2	2	2	3	3	4	4	3	2	3	2	2	3	4	27.5
$\tilde{\gamma}_H$	2	2	7	4	2	4	5	3	2	3	7	4	3	2	3	5	4	3	5	3	36.5
μ_B	4	3	4	3	3	2	2	2	4	3	4	3	2	4	3	1	3	2	4	4	30
σ_B	3	3	4	5	3	2	3	3	5	3	3	2	2	1	2	4	4	2	3	1	29
δ_B	0	0	0	1	0	1	2	1	0	0	1	0	0	2	0	0	0	3	1	1	6.5
$\tilde{\delta}_B$	1	1	0	1	0	1	1	2	1	0	0	0	0	0	0	0	1	1	0	0	5
γ_B	0	0	0	0	0	0	1	0	3	0	1	0	1	2	3	1	1	1	0	1	7.5
$\tilde{\gamma}_B$	0	0	0	1	1	0	0	1	1	0	0	0	0	1	1	1	0	0	0	0	3.5
μ_S	1	2	2	2	1	1	0	4	2	2	1	3	3	3	4	3	1	2	2	4	21.5
σ_S	3	2	5	4	3	3	5	1	1	4	5	6	5	3	3	3	4	4	4	4	36
δ_S	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	2
$\tilde{\delta}_S$	1	1	0	0	0	0	1	1	1	1	0	0	0	1	0	0	1	1	2	0	5.5
γ_S	1	1	0	0	1	0	0	1	1	0	0	0	1	2	0	0	1	1	1	0	5.5
$\tilde{\gamma}_S$	0	0	0	0	0	0	0	2	2	1	0	0	0	2	1	2	2	2	1	1	8
μ_{E_c}	5	5	3	4	4	3	2	2	5	5	4	2	2	2	5	1	3	0	1	1	29.5
σ_{E_c}	3	3	4	5	3	2	2	3	2	2	4	4	3	2	3	3	2	2	3	3	29
δ_{E_c}	5	5	3	1	3	2	5	2	3	3	2	2	1	2	3	4	2	3	3	2	28
$\tilde{\delta}_{E_c}$	5	5	5	3	1	2	2	3	4	4	3	4	4	3	5	2	1	3	4	2	32.5
γ_{E_c}	4	4	4	3	1	1	3	2	3	3	3	3	5	1	4	1	1	1	3	3	26.5
$\tilde{\gamma}_{E_c}$	5	4	3	6	2	2	2	4	2	4	4	4	4	2	3	2	1	3	3	2	31
μ_{E_z}	1	2	2	0	1	2	2	2	4	3	4	1	1	1	2	2	2	1	0	1	17
σ_{E_z}	3	3	3	1	1	2	1	1	5	4	1	3	3	2	1	3	3	2	3	1	23
δ_{E_z}	5	4	5	5	5	0	4	1	2	3	3	1	4	3	6	3	5	2	3	3	33.5
$\tilde{\delta}_{E_z}$	4	3	2	0	2	0	2	2	1	3	2	5	5	2	4	0	0	4	2	2	22.5
γ_{E_z}	3	3	4	4	2	2	2	1	3	3	3	2	5	2	5	5	1	4	4	4	31
$\tilde{\gamma}_{E_z}$	1	1	5	2	2	3	4	4	2	1	3	5	4	2	4	4	0	3	2	3	27.5
μ_R	2	2	3	5	4	3	1	2	1	3	4	1	2	4	3	2	4	3	2	3	27
σ_R	2	1	4	5	4	3	4	2	5	3	5	4	4	3	3	4	3	1	2	3	32.5
δ_R	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	2	0	0	2
$\tilde{\delta}_R$	0	0	0	0	0	0	2	3	1	2	0	0	0	2	0	1	2	1	0	0	7
γ_R	1	1	0	1	0	0	0	1	1	1	0	0	0	2	0	1	1	0	1	0	5.5
$\tilde{\gamma}_R$	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	2	0	0	3.5

Table A.6: Summary of each feature chosen in SFFS-LDF (H : HR, B : BVP, S : SC, E_c : EMGc, E_z : EMGz and R : Resp, $Frq = \frac{\sum}{\#rd}$, $\#rd = 10$)

	Valence		Arousal			Valence		Arousal	
	$\frac{\sum}{\#rd}$	Frq (%)	$\frac{\sum}{\#rd}$	Frq (%)		$\frac{\sum}{\#rd}$	Frq (%)	$\frac{\sum}{\#rd}$	Frq (%)
μ_H	2/10	20	3/10	30	μ_B	2/10	20	4/10	40
σ_H	1/10	10	2/10	20	σ_B	4/10	40	3/10	30
δ_H	2/10	20	1/10	10	δ_B	5/10	50	1/10	10
$\tilde{\delta}_H$	2/10	20	2/10	20	$\tilde{\delta}_B$	2/10	20	2/10	20
γ_H	1/10	10	2/10	20	γ_B	3/10	30	1/10	10
$\tilde{\gamma}_H$	2/10	20	4/10	40	$\tilde{\gamma}_B$	2/10	20	3/10	30
μ_S	6/10	60	5/10	50	μ_{E_c}	4/10	40	2/10	20
σ_S	2/10	20	1/10	10	σ_{E_c}	3/10	30	2/10	20
δ_S	4/10	40	1/10	10	δ_{E_c}	4/10	40	0/10	0
$\tilde{\delta}_S$	6/10	60	2/10	20	$\tilde{\delta}_{E_c}$	1/10	10	6/10	60
γ_S	3/10	30	0/10	0	γ_{E_c}	5/10	50	1/10	10
$\tilde{\gamma}_S$	2/10	20	2/10	20	$\tilde{\gamma}_{E_c}$	2/10	20	0/10	0
μ_{E_z}	5/10	50	3/10	30	μ_R	2/10	20	2/10	20
σ_{E_z}	3/10	30	3/10	30	σ_R	1/10	10	4/10	40
δ_{E_z}	6/10	60	2/10	20	δ_R	3/10	30	1/10	10
$\tilde{\delta}_{E_z}$	3/10	30	1/10	10	$\tilde{\delta}_R$	3/10	30	3/10	30
γ_{E_z}	3/10	30	2/10	20	γ_R	3/10	30	0/10	0
$\tilde{\gamma}_{E_z}$	2/10	20	3/10	30	$\tilde{\gamma}_R$	2/10	20	1/10	10

References

- [1] R. W. Picard, *Affective computing*. Cambridge, Mass: The MIT Press, 1997.
- [2] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.
- [3] R. W. Picard, "Affective computing," MIT Media Laboratory Perceptual Computing Section, Tech. Rep. No. 321, 1995.
- [4] P. Ekman, "Are there basic emotions?" *Psychological Review*, vol. 99, no. 3, pp. 550–553, 1992.
- [5] ———, "An argument for basic emotions," *Cognition and Emotion*, vol. 6, no. 3/4, pp. 169–200, 1992.
- [6] Y. L. Lin and G. Wei, "Speech emotion recognition based on HMM and SVM," in *Proceedings of the 4th International conference on machine Learning and Cybernetics*, 2005, pp. 4898–4901.
- [7] J. Nicholson, K. Takahashi, and R. Nakatsu, "Emotion recognition in speech using neural networks," in *Proceedings of the 6th International Conference on Neural Information Processing*, 1999, pp. 495–501.
- [8] E. H. Kim, K. H. Hyun, and Y. K. Kwak, "Robust emotion recognition feature, frequency range of meaningful signal," in *2005 IEEE International*

- Workshop on Robots and Human Interactive Communication*, 2005, pp. 667–671.
- [9] Z. J. Chuang and C. H. Wu, “Emotion recognition using acoustic features and textual content,” in *2004 IEEE International Conference on Multimedia and Expo*, 2004, pp. 53–56.
- [10] J. Kim and E. André, “Emotion recognition based on physiological changes in music listening,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 12, pp. 2067–2083, 2008.
- [11] P. Ekman, “Strong evidence for universals in facial expressions: a reply to russell’s mistaken critique,” *Psychological Bulletin*, vol. 115, no. 2, pp. 268–287, 1994.
- [12] R. Plutchik, *Emotions and life: perspectives from psychology, biology and evolution*. American Psychological Association (APA), 2002.
- [13] W. D. TenHouten, *A general theory of emotions and social life*. London and New York: Routledge, 2007.
- [14] S. S. Tomkins, *Affect, imagery, consciousness*. New York: Springer Publishing Company, Inc, 1962, vol. I, The positive affects.
- [15] —, *Affect, imagery, consciousness*. New York: Springer Publishing Company, Inc, 1963, vol. II, The negative affects.
- [16] J. J. Gross and R. W. Levenson, “Emotion elicitation using films,” *Cognition and Emotion*, vol. 9, no. 1, pp. 87–108, 1995.
- [17] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, “International affective picture system (IAPS): Affective ratings of pictures and instruction manual,” University of Florida, Gainesville, FL, Tech. Rep. A-6, 2005.

-
- [18] C. Darwin, *The expression of the emotions in man and animals*. Chicago: University of Chicago Press, 1965, (Original work published 1872).
- [19] P. Ekman, "All emotions are basic," in *The nature of emotion: Fundamental questions*, P. Ekman and R. J. Davidson, Eds. New York: Oxford University Press, 1994, pp. 7–19.
- [20] I. Eibl-Eibesfeldt, *Human ethology*. Aldine Transaction, 1989.
- [21] C. E. Izard, *The face of emotion*. New York: Appleton-Century-Crofts, 1971.
- [22] —, "Basic emotions, relations among emotions, and emotion-cognition relations," *Psychology Review*, vol. 99, pp. 561–565, 1992.
- [23] R. Plutchik, "A general psychoevolutionary theory of emotion," in *Emotion Theory, Research, and Experience*, R. Plutchik and H. Kellerman, Eds. Academic Press, 1980, vol. 1.
- [24] A. Ortony and T. J. Turner, "hat's basic about basic emotions?" *Psychological Review*, vol. 97, pp. 315–331, 1990.
- [25] N. L. Stein and K. Oatley, Eds., *Basic emotions*. Hove, UK: Lawrence Erlbaum Associates, 1992.
- [26] J. Panksepp, "A critical role for 'affective neuroscience' in resolving what is basic about emotions," *Psychological Review*, vol. 99, no. 3, pp. 554–560, 1992.
- [27] R. A. Schweder and J. Hoidt, "The cultural psychology of the emotions: ancient and new," in *Handbook of emotions*, M. Lewis and J. M. Haviland-Jones, Eds. New York: Guilford Press, 2002, pp. 397–414.
- [28] R. Plutchik, *Emotions: a psychoevolutionary synthesis*. New York: Harper & Row, 1980.

-
- [29] A. Ortony, G. L. Clore, and A. Collins, *The cognitive structure of emotions*. Cambridge, MA: Cambridge University Press, 1988.
- [30] H. Schlosberg, "The description of facial expressions in terms of two dimensions," *Journal of Experimental Psychology*, vol. 44, pp. 229–237, 1952.
- [31] L. E. U. Bush, "Individual differences in multidimensional scaling of adjectives denoting feelings," *Journal of Personality and Social Psychology*, vol. 25, pp. 50–57, 1973.
- [32] R. S. Green and N. Cliff, "Multidimensional comparisons of structures of vocally and facially expressed emotions," *Perception Psychophysics*, vol. 17, pp. 429–438, 1975.
- [33] J. A. Russell and M. Bullock, "Multidimensional scaling of emotional facial expressions: similarity from preschoolers to adults," *Journal of Personality and Social Psychology*, vol. 48, pp. 1290–1298, 1985.
- [34] P. J. Lang, "Cognition in emotion: concept and action," in *Emotions, Cognition, and Behavior*, C. E. Izard, J. Kagan, and R. B. Zajonc, Eds. Cambridge: Cambridge University Press, 1984, pp. 192–226.
- [35] H. Schlosberg, "Three dimensions of emotion," *Psychological Review*, vol. 61, no. 2, pp. 84–88, 1954.
- [36] A. Mehrabian and J. Russell, "Evidence for a three-factor theory of emotions," *Journal of Research in Personality*, vol. 11, pp. 273–294, 1977.
- [37] P. J. Lang, "The emotion probe: studies of motivation and attention," *American Psychologist*, vol. 50, no. 5, pp. 372–385, 1995.
- [38] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, pp. 1161–1178, 1980.

- [39] T. T. Rolls, “A theory of emotion, and its application to understanding the neural basis of emotion,” *Cognition Emotion*, vol. 4, pp. 161–190, 1990.
- [40] E. T. Rolls, *The brain and emotion*. Oxford, UK and New York: Oxford University Press, 2001.
- [41] J. A. M. abd B. L. Fredrickson, G. R. Larkin, C. M. Lindberg, S. J. Maglio, and P. A. Reuter-Lorenz, “Emotional category data on images from the international affective picture system,” *Behavior Research Methods*, vol. 37, no. 4, pp. 626–630, 2005.
- [42] W. S. N. Reilly, “Believable social and emotional agents,” Ph.D. dissertation, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1996.
- [43] I. J. Roseman, A. A. Antoniou, and P. E. Jose, “Appraisal determinants of emotions: constructing a more accurate and comprehensive theory,” *Cognition and Emotion*, vol. 10, no. 3, pp. 241–277, 1996.
- [44] R. E. Thayer, *The biopsychology of mood and arousal*. New York: Oxford University Press, 1989.
- [45] A. Gerrards-Hesse, K. Spies, and F. W. Hesse, “Experimental inductions of emotional states and their effectiveness: A review,” *British Journal of Psychology*, vol. 85, pp. 55–78, 1994.
- [46] P. J. Lang, “Behavioral treatment and bio-behavioral assessment: Computer applications,” in *Technology in Mental Health Care Delivery Systems*, J. B. sidowski, J. H. Johnson, and T. A. Williams, Eds. NY: Ablex Pub, 1980, pp. 119–167.
- [47] D. M. Clynes, *Sentics: The touch of emotions*. Anchor Press/Doubleday, 1977.

- [48] A. Haag, S. Goronzy, P. Schaich, and J. Williams, “Emotion recognition using biosensors: first step towards an automatic system,” in *Affective Dialogue Systems, Tutorial and Research Workshop*, Kloster Irsee, Germany, June 2004, pp. 36–48.
- [49] F. Nasoz, K. Alvarez, C. L. Lisetti, and N. Finkelstein, “Emotion recognition from physiological signals for presence technologies,” *International Journal of Cognition, Technology and Work, Special Issue on Presence*, vol. 6, no. 1, 2003.
- [50] J. Wagner, J. Kim, and E. André, “From physiological signals to emotions: implementing and comparing selected methods for feature extraction and classification,” in *Proceedings Of IEEE ICME International Conference on Multimedia and Expo*, 2005, pp. 940–943.
- [51] E. L. van den Broek, V. Lisý, J. H. Janssen, J. H. D. M. Westerink, M. H. Schut, and K. Tuinenbreijer, “Affective man-machine interface: unveiling human emotions through biosignals,” in *Biomedical Engineering Systems and Technologies, Communications in Computer and Information Science*, J. F. A. Fred and H. Gamboa, Eds., 2010, pp. 21–47.
- [52] M. M. Bradley and P. J. Lang, “The international affective picture system (IAPS) in the study of emotion and attention,” in *Handbook of Emotion Elicitation and Assessment*, J. A. Coan and J. J. B. Allen, Eds. USA: Oxford University Press, 2007, pp. 29–46.
- [53] D. Watson, L. A. Clark, and A. Tellegen, “Development and validation of brief measures of positive and negative affect: The panas scales,” *Journal of Personality and Social Psychology*, vol. 54, pp. 1063–1070, 1988.

- [54] W. Q. E. Perunovic, D. Heller, and E. Rafaeli, "Within-person changes in the structure of emotion: The role of cultural identification and language," *Psychological Science*, vol. 18, pp. 607–613, 2007.
- [55] O. P. John and S. Srivastava, "The big five trait taxonomy: History, measurement, and theoretical perspectives," in *Handbook of personality: Theory and research*, 2nd ed., L. A. Pervin and O. P. John, Eds. New York: Guilford, 1999, pp. 102–138.
- [56] I. Tankova, A. Adan, and G. Buela-Casal, "Circadian typology and individual differences: A review," *Personality and Individual Differences*, vol. 16, p. 671C684, 1994.
- [57] G. D. Wilson, "Personality, time of day and arousal," *Personality and Individual Differences*, vol. 11, no. 2, pp. 153–168, 1990.
- [58] J. A. Horne and O. Östberg, "A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms," *International Journal of Chronobiology*, vol. 4, pp. 97–110, 1976.
- [59] M. A. Carskadon, C. Vieira, and C. Acebo, "Association between puberty and delayed phase preference," *Sleep*, vol. 16, no. 3, pp. 258–262, 1993.
- [60] B. U. Kohler, C. Hennig, and R. Orglmeister, "The principles of software qrs detection," *IEEE Engineering in Medicine and Biology*, vol. 21, no. 1, pp. 42–57, 2002.
- [61] R. M. Rangayyan, *Biomedical signal analysis: A case-study approach (IEEE Press Series on Biomedical Engineering)*. Wiley-IEEE Press, 2001.
- [62] K. Mera and T. Ichimura, "Emotion analyzing method using physiological state," in *Knowledge-Based Intelligent Information and Engineering Systems*. Springer Berlin / Heidelberg, 2004, pp. 195–201.

-
- [63] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
- [64] R. G. Miller and B. W. Byron, *Beyond ANOVA: Basics of Applied Statistics*. New York: Chapman & Hall, 1997.
- [65] I. T. Jolliffe, *Principle Component Analysis*. New York: Springer Publishing Company, Inc, 2002.
- [66] H. Liu and L. Yu, “Toward integrating feature selection algorithms for classification and clustering,” *IEEE Trans. on Knowledge and Data Engineering*, vol. 17, no. 4, 2005.
- [67] P. Pudil, J. novovičová, and J. Kittler, “Floating search methods in feature selection,” *Pattern Recognition Letters*, vol. 15, pp. 1119–1125, 1994.
- [68] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. MA: Addison-Wesley, Reading, 1989.
- [69] T. M. Cover and P. E. Hart, “Nearest neighbor pattern classification,” *IEEE Transactions on Information Theory*, vol. 13, pp. 21–27, 1967.
- [70] W. Siedlecki and J. Sklansky, “A note on genetic algorithms for large-scale feature selection,” *Pattern Recognition Letters*, vol. 10, pp. 335–347, 1988.
- [71] S. A. Dudani, “The distance-weighted k-nearest-neighbor rule,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 6, pp. 325–327, 1976.
- [72] J. M. Keller, M. R. Gray, and J. J. A. Givens, “A fuzzy k-nearest neighbor algorithm,” *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 15, no. 4, pp. 580–585, 1985.

- [73] B. E. Boser, I. Guyon, and V. Vapnik, “A training algorithm for optimal margin classifiers,” in *Proceedings of the 5th Annual Workshop on Computational Learning Theory*. ACM Press, 1992, pp. 144–152.
- [74] J. Zhai and A. Barreto, “Stress detection in computer users based on digital signal processing of noninvasive physiological variables,” in *Proceedings of the 28th IEEE EMBS Annual International Conference*, 2006.
- [75] C. K. Lee, S. K. Yoo, Y. J. Park, N. H. Kim, K. S. Jeong, and B. C. Lee, “Using neural network to recognize human emotions from heart rate variability and skin resistance,” in *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2005.
- [76] L. Li and J. H. Chen, “Emotion recognition using physiological signals from multiple subjects,” in *Proceedings of the 2006 International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2006.
- [77] D. Peel and G. J. McLachlan, “Robust mixture modelling using the t-distribution,” *Statistics and Computing*, vol. 10, pp. 339–348, 2000.
- [78] S. Shoham, “Robust clustering by deterministic agglomeration em of mixtures of multivariate t-distributions,” *Pattern Recognition*, vol. 35, pp. 1127–1142, 2002.
- [79] J. A. Bilmes, “A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models,” International Computer Science Institute, U.C. Berkeley, Tech. Rep., 1998.
- [80] M. Svensén and C. M. Bishop, “Robust bayesian mixture modelling,” in *Trends in Neurocomputing: 12th European Symposium on Artificial Neural Networks 2004*, vol. 64, 2004, pp. 235–252.

-
- [81] D. Gerogiannis, C. Nikou, and A. Likas, “The mixtures of students t-distribution as a robust framework for rigid registration,” *Image and Vision Computing*, vol. 27, pp. 1285–1294, 2008.
- [82] STUDENT, “The probable error of a mean,” *Biometrika*, vol. 6, no. 1, pp. 1–25, 1908.
- [83] C. Liu and D. B. Rubin, “Ml estimate of the t distribution using em and its extensions, ecm and ecme,” *Statistica Sinica*, vol. 5, pp. 19–39, 1995.
- [84] S.-S. Cheng, H.-M. Wang, and H.-C. Fu, “A model-selection-based self-splitting gaussian mixture learning with application to speaker identification,” *Journal on Applied Signal Processing*, vol. 17, pp. 2626–2639, 2004.